



Fractal' versus 'Markov' models of traffic (conférence invitée)

Sandrine Vaton

► To cite this version:

Sandrine Vaton. Fractal' versus 'Markov' models of traffic (conférence invitée). Eighth IFIP Workshop on Performance Modelling and Evaluation of ATM and IP Networks, Jul 2000, Bradford, United Kingdom. hal-02338788

HAL Id: hal-02338788

<https://hal.science/hal-02338788>

Submitted on 30 Oct 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

'Fractal' versus 'Markov' models of traffic.

Sandrine Vaton

ENST Bretagne, BP 832, 29285 Brest Cedex, France.

e-mail: Sandrine.Vaton@enst-bretagne.fr

June 26, 2000

Summary :

This tutorial paper provides an overview of the state of the art in teletraffic modelling, with regards to the debate about the possible 'fractal' nature of the traffic on modern networks.

1 Introduction

In recent years numerous statistical studies on high quality traffic data have concluded that a major part of the traffic that is likely to circulate on future B-ISDN networks exhibits long term correlations and a form of self similarity over several time scales [Leland et al.94] [Duffield et al.] [Paxson et al.94] [Willinger et al.95b] [Willinger et al.95a]. As a result of these findings, most contributors reject traditional "Poisson-like" models of traffic such as the Poisson process [Erlang09] [Erlang17] or the Markov Modulated Poisson Process [Fischer et al.93]. This is an important bend in teletraffic modelling because, at first glance, that means that traditional performance evaluation is not valuable; most authors think that traditional performance evaluation underestimate the ressources that are necessary to provide Quality of Service because of the 'fractal' nature of the modern traffic.

On the other hand, many classical indexes that are used to decide for the 'fractal' nature of the traffic are not valid if the traffic is not stationary. The hypothesis of second order (or "weak") stationarity is essential to speak of long range dependence. This remark is not cumbersome; some non stationarities -e.g. shifting levels- are likely to occur and they can be misinterpreted as a strong degree of long range dependence if they are not detected. The question of deciding whether the traffic has some 'fractal' nature or presents some non stationarities is in fact very difficult.

What is more, any stochastic process can be approximated by a Markov model, provided that the number of states of the Markov model is sufficiently high. In particular there exists some Markov models that approximate the correlation structure of long range dependent processes on a number of timescales; and when the number of states of the Markov model tends to infinity the process is asymptotically long range dependent. These Markov models present some near complete decomposability. The advantages of modelling the traffic with a near completely decomposable Markov model is twofold : (i) it is possible to propose some near completely decomposable Markov models that are parcimonious, that is to say that it has very few parameters even if the number of states is infinite (ii) the second reason is that it is possible to re-use well known traditional performance analysis to dimension and manage the network under this modelling.

As one can see the question of the validity of 'Poisson' and of 'fractal' like models of traffic is still a very open question.

2 Long range dependence and self-similarity : a definition

2.1 Long range dependence

Denote by $(X_t)_t$ a **second order stationary process**; we denote by $\rho(k)$ the correlations of $(X_t)_t$ and by $S(\lambda)$ its power density spectrum.

Definition 1 (X_t) is long range dependent if and only if the following (equivalent) conditions are satisfied :

1. the correlations $\rho(k)$ decrease as a non summable power law when k tends to infinity :

$$\rho(k) \stackrel{k \rightarrow \infty}{\asymp} O(k^{2H-2}), 0.5 < H < 1.0.$$

2. or, equivalently, the power density spectrum $S(\lambda)$ diverges as an integrable power law near the origin :

$$S(\lambda) \stackrel{\lambda \rightarrow 0}{\asymp} O(\lambda^{1-2H}), 0.5 < H < 1.0.$$

The parameter H ($0.5 < H < 1.0$) is the Hurst parameter of the process.

2.2 Self-similarity

Denote now by $(Y_t)_t$ a continuous time process ($t \in \mathbb{R}$); the definition of self similarity is following :

Definition 2 $(Y_t)_t$ is self similar with self similarity parameter H if and only if $c^{-H}Y_{ct}$ is equal in distribution to Y_t .

It is noteworthy that, unless $H = 0$, $(Y_t)_t$ is not stationary. For the purpose of modelling time series that look stationary we will consider self similar processes with stationary increments; moreover we reduce our study to the case $H > 0$ ¹.

2.3 Stationary increments of a self similar process

Denote by $X_t = Y_t - Y_{t-1}$ the increment process of the self similar process $(Y_t)_t$. The variance of the increment process is

$$\text{var}(Y_t - Y_s) = \text{var}(X(t-s)) = \sigma^2|t-s|^{2H}$$

and the correlations of X_t are given by

$$\rho(k) = \text{corr}(X_t, X_{t+k}) = \frac{1}{2}[(k+1)^{2H} - 2k^{2H} + (k-1)^{2H}]$$

Now if we let k tend to ∞ in this equation it comes that :

¹the case $H < 0$ is pathological : if Y_t were self similar with stationary increments and $H < 0$ then the Y_t would not be measurable; we do not consider this case

$$\lim_{k \rightarrow \infty} \rho(k) = H(2H - 1)k^{2H-2}$$

or equivalently

$$S_X(\lambda) = O(|\lambda|^{1-2H})$$

and consequently $(X_t)_t$ is long range dependent if H is in the range $0.5 < H < 1.0$.

Theorem 1 *If Y_t is self similar with self similarity parameter H in the range $0.5 < H < 1.0$ then the increment process $X_t = Y_t - Y_{t-1}$ is a stationary long range dependent process with Hurst parameter H .*

3 Fractional Brownian motion, fractional Gaussian noise and fractional ARIMA models

The Brownian motion (H=0.5)

The standard Brownian motion is a continuous time stochastic process $B(t)$ that verifies the following properties :

(i) the sample paths are continuous (ii) $B(t)$ is Gaussian (iii) $B(t)$ has indepent increments (iv) $\mathbb{E}(B(t) - B(s)) = 0$ (v) $\text{var}(B(t) - B(s)) = \sigma^2|t - s|$

If $B(t)$ verifies these properties then $B(t)$ is a Brownian motion. It is easy to prove that the 'standard' Brownian motion $B(t)$ is self similar with self similarity parameter $H = 0.5$.

3.1 The fractional Brownian motion

The fractional Brownian motion can be defined as a 'weighted average' of the standard Brownian motion, or more precisely, as a stochastic integral, the convergence of the integral being understood in the sense of the L^2 -norm with respect to the Lebesgue measure on \mathbb{R} :

Definition 3 *Note $w_H(t, u)$ the 'weight' function*

$$\begin{aligned} w_H(t, u) &= 0 & t &\leq u \\ w_H(t, u) &= (t - u)^{H - \frac{1}{2}} & 0 \leq u < t \\ w_H(t, u) &= (t - u)^{H - \frac{1}{2}} - (-u)^{H - \frac{1}{2}} \end{aligned}$$

The fractional Brownian motion with self similarity parameter H is defined as the stochastic integral

$$B_H(t) = \sigma \int w_H(t, u) dB(u)$$

in which $B(t)$ is the standard Brownian motion as defined above and normalized such that $\text{var}(B(1)) = 1$.

Theorem 2 *The fractional Brownian motion $B_H(t)$ is self similar with self similarity parameter H*

$$B_H(t) \stackrel{(d)}{=} B(t)$$

where $\stackrel{(d)}{=}$ represents the equality in distribution.

3.2 The fractional Gaussian noise

Definition 4 *The fractional Gaussian noise is the increment process associated to the fractional Brownian motion*

$$X_H(t) = B_H(t) - B_H(t-1)$$

Theorem 3 *The fractional Gaussian noise $X_H(t)$ is long range dependent with Hurst parameter H :*

$$\begin{aligned} \rho_X(k) &\stackrel{k \rightarrow \infty}{\sim} O(k^{2H-2}) \\ S_X(\lambda) &\stackrel{\lambda \rightarrow 0}{\sim} 0(\lambda^{1-2H}) \end{aligned}$$

3.3 The fractional ARIMA models

ARMA processes

Denote by B the Backshift operator defined by $(BX)(t) = X(t-1)$ and let p and q be integers.

Definition 5 *An ARMA(p, q) is the stationary solution to*

$$\Phi(B)X_t = \Psi(B)\epsilon_t$$

where ϵ_t is a white noise and $\Phi(\bullet)$ and $\Psi(\bullet)$ are polynomials of order p and q , $\Phi(x) = 1 - \sum_{i=1}^p \phi_i x^i$ and $\Psi(x) = 1 + \sum_{j=1}^q \psi_j x^j$.

ARIMA processes

Now let d be an integer and consider the polynomial $(1-B)^d = \sum_{k=0}^d C_d^k (-1)^k B^k$ where $C_d^k = \frac{d!}{k!(d-k)!} = \frac{\Gamma(d+1)}{\Gamma(k+1)\Gamma(d-k+1)}$ with $\Gamma(x) = \int_0^\infty t^{x-1} \exp(-t) dt$.

Definition 6 *If $(X_t)_t$ satisfies to the equation*

$$\Phi(B)(1-B)^d X_t = \Psi(B)\epsilon_t$$

then $(X_t)_t$ is an ARIMA(p, d, q).

The operator $(1-B)^d$ can be generalized to d non integer as follows :

$$(1-B)^d = \sum_{k=0}^{\infty} \frac{\Gamma(d+1)}{\Gamma(k+1)\Gamma(d-k+1)} (-1)^k B^k$$

and the definition of the ARIMA(p, d, q) can be extended to non integer values of d .

Definition 7 *Let $(X_t)_t$ be a stationary process such that*

$$\Phi(B)(1-B)^d X_t = \Psi(B)\epsilon_t$$

with d in the range $-0.5 < d < 0.5$ then $(X_t)_t$ is a fractional ARIMA(p, d, q) process.

The spectral density of the fractional ARIMA process is given by

$$S(\lambda) = \frac{\sigma_\epsilon^2}{2\pi} |1 - e^{-i\lambda}|^{-2d} \left| \frac{\Psi(e^{i\lambda})}{\Phi(e^{i\lambda})} \right|^2 = \frac{\sigma_\epsilon^2}{2\pi} |2 \sim (\lambda/2)|^{-2d} \left| \frac{\Psi(e^{i\lambda})}{\Phi(e^{i\lambda})} \right|^2$$

In this expression if λ tends to 0 it comes that

$$S(\lambda) \xrightarrow{\lambda \rightarrow 0} \frac{\sigma_\epsilon^2}{2\pi} |\lambda|^{-2d} \left| \frac{\Psi(1)}{\Phi(1)} \right|^2 = \frac{\sigma_\epsilon^2}{2\pi} |\lambda|^{1-2H} \left| \frac{\Psi(1)}{\Phi(1)} \right|^2 \quad \text{with} \quad H = d + 0.5$$

Consequently $(X_t)_t$ is long range dependence if d is in the range $0 < d < 0.5$.

Theorem 4 *The fractional ARIMA(p,d,q) is long range dependent with Hurst parameter $H = d + 0.5$ if d is in the range $0 < d < 1.0$.*

The case $d = 0$ corresponds to the ARMA(p,q) process. For $d \geq 0.5$ there is no stationary solution for the equation $\Phi(B)(1 - B)^d X_t = \Psi(B)\epsilon_t$; but the case $d \geq 0.5$ can reduce to the case $-0.5 < d \leq 0.5$ by taking appropriate differences. For the purpose of modelling time series with long range dependence the case of interest is the case $0 < d < 0.5$.

3.4 Generation of an ARIMA(p,d,q)

Direct computation of the covariances of the ARIMA(p,d,q) is a difficult matter. Gradshteyn and Ryzhik have established the exact expression of the covariances of the ARIMA(0,d,0) X_t^*

$$(1 - B)^d X_t^* = \epsilon_t$$

The covariances of the ARIMA(0,d,0) process are

$$\gamma(k) = \sigma_\epsilon^2 \frac{(-1)^k \Gamma(1 - 2d)}{\Gamma(k - d + 1) \Gamma(1 - k - d)}$$

The Durbin Levinson algorithm

The Durbin Levinson algorithm is a method for determining recursively the best linear predictors of increasing order for a stationary process $(X_t)_t$ with known covariances.

Denote by \hat{X}_{n+1} the best linear predictor of X_{n+1} given X_1, X_2, \dots, X_n

$$\hat{X}_{n+1} = \phi_{n,1} X_n + \phi_{n,2} X_{n-1} + \dots + \phi_{n,n} X_1$$

and denote by v_n the mean squared error of prediction

$$v_n = \mathbb{E}((X_{n+1} - \hat{X}_{n+1})^2)$$

Proposition 1 *If X_t is a stationary zero mean process with covariances $\gamma(\bullet)$ such that $\gamma(0) > 0$ and $\lim_{k \rightarrow \infty} \gamma(k) = 0$ then the coefficients $\phi_{n,j}$ and the mean squared errors v_n satisfy $\phi_{1,1} = \gamma(1)/\gamma(0)$, $v_0 = \gamma(0)$,*

$$\begin{aligned} j = 1, \dots, n-1 \quad & \phi_{n,n} = [\gamma(n) - \sum_{j=1}^{n-1} \phi_{n-1,j} \gamma(n-j)] v_{n-1}^{-1} \\ & \phi_{n,j} = \phi_{n-1,j} - \phi_{n,n} \phi_{n-1,n-j} \\ & v_n = v_{n-1} (1 - \phi_{n,n}^2) \end{aligned}$$

Generation of the ARIMA(0,d,0)

The following algorithm produces a process that is distributed as the ARIMA(0,d,0) $(X_t^*)_t$.

1. Initialization :

- (a) $n = 1$.
- (b) Generate $X_1^* \sim \mathcal{N}(0, \sqrt{\gamma(0)})$.
- (c) $\phi_{1,1} = \gamma(1)/\gamma(0)$, $v_0 = \gamma(0)$, $v_1 = v_0(1 - \phi_{1,1}^2)$.

2. Recursion :

- (a)

$$X_{n+1} = \sum_{j=1,n} \phi_{n,j} X_{n+1-j}^* + \sqrt{v_n} \epsilon_n \quad \text{where } (\epsilon_n)_n \sim WGN(0, 1)$$

- (b) Compute $\phi_{n+1,\bullet}$ and v_{n+1} as described in the Durbin Levinson algorithm.
- (c) $n=n+1$.

Generation of the ARIMA(p,d,q)

The ARIMA(p,d,q) $(X_t)_t$ can be obtained as the output of an ARMA(p,q) filter with input the ARIMA(0,d,0) $(X_t^*)_t$

$$\begin{aligned} \Phi(B)X_t &= \Psi(B)X_t^* \\ X_t &= \sum_{j=1,p} \phi_j X_{t-j}^* + X_t^* + \sum_{k=1,q} \psi_k X_{t-k}^* \end{aligned}$$

4 Estimation of the Hurst parameter: heuristic approaches

4.1 The R/S statistic or pox diagram

Suppose that one wants to detect long memory in a series $(X_t)_t$. Note Y_t the cumulated series $Y_t = \sum_{i=1}^t X_i$. Note $R(t, k)$ the adjusted range

$$R(t, k) = \max_{0 \leq i \leq k} [Y_{t+i} - Y_t - \frac{i}{k} (Y_{t+k} - Y_t)] - \min_{0 \leq i \leq k} [Y_{t+i} - Y_t - \frac{i}{k} (Y_{t+k} - Y_t)]$$

and $S(t, k)$ the sample variance of $X_{t+1}, X_{t+2}, \dots, X_{t+k}$

$$S(t, k) = \sqrt{k^{-1} \sum_{i=1, k} (X_{t+i} - \bar{X}_{t, k})^2}$$

where $X_{t, k}$ is the sample mean $X_{t, k} = k^{-1} \sum_{i=1, k} X_{t+i}$. The R/S statistics or rescaled adjusted range is the ratio

$$Q(t, k) = R(t, k)/S(t, k)$$

Theorem 5 *Let X_t be such that X_t^2 is ergodic and $t^{-0.5} \sum_{s=1, t} X_s$ converges weakly to Brownian motion as t tends to ∞ . Then, as $k \rightarrow \infty$*

$$k^{-0.5} Q \xrightarrow{(d)} \zeta$$

where ζ is a non degenerate random variable.

Theorem 6 *Let X_t be such that X_t^2 is ergodic and $t^{-H} \sum_{s=1, t} X_s$ converges weakly to fractional Brownian motion as t tends to ∞ . Then, as $k \rightarrow \infty$*

$$k^{-H} Q \xrightarrow{(d)} \zeta$$

where ζ is a non degenerate random variable.

The R/S statistic should be evaluated for many k and one should plot $\log(R/S)$ versus $\log k$. For most common short memory process the points should be scattered around a straight line with slope 0.5. For long memory processes the points should be scattered around a straight line with slope H in the range $0.5 < H < 1.0$. The slope of this straight line is an heuristic estimate of the Hurst parameter of the series.

4.2 The variance time plot

This method is based on the slower decrease of the sample mean of long range dependent processes.

Theorem 7 *Let X_t be a stationary process with long range dependence. Then*

$$\text{var}(n^{-1} \sum_{i=1, n} X_i) \stackrel{n \rightarrow \infty}{\sim} O(n^{2H-2})$$

with H the Hurst parameter of the process.

Note $(X_t^{(m)})_t$ the series obtained by taking sample means over non overlapping windows of size m

$$X_t^{(m)} = m^{-1} \sum_{j=1, m} X_{(t-1)m+j}$$

and note s_m the sample variance of $(X_t^{(m)})_t$. One should calculate s_m for many m and plot $\log(s_m)$ againsts $\log(m)$. If there is some long range dependence with Hurst parameter H then the plot should fit to a straight line with slope $2H - 2$ for big values of m ($0.5 < H < 1.0$). If not then it should fit to a straight line with slope -1 .

4.3 The Index of Dispersion for Counts (IDC)

The Index of Dispersion for Counts is commonly used for capturing the variability of traffic over different time scales. Note A_t the number of arrivals during the interval of length t ; the IDC is the variance of A_t divided by the expected value of A_t

$$IDC(t) = \frac{var(A_t)}{\mathbb{E}(A_t)}$$

Theorem 8 *If A_t is a self similar process with self similarity parameter H then*

$$\frac{var(A_t)}{\mathbb{E}(A_t)} \stackrel{t \rightarrow \infty}{\sim} O(t^{2H-1})$$

In practice one counts the number of arrivals over non overlapping windows of length t ; the sample mean and the sample variance of the counts A_t are evaluated; this is done for many values of t . One plots $\log \frac{var(A_t)}{\mathbb{E}(A_t)}$ versus $\log t$. If the series of the counts is self similar with self similarity parameter H then the plot is approximately linear for t big with slope $2H - 1$.

4.4 Periodogram based analysis

$(X_t)_t$ is long range dependent if and only if its spectral density $f(\lambda)$ has an integrable singularity in $\lambda = 0$

$$f(\lambda) \stackrel{\lambda \rightarrow 0}{\sim} c \lambda^{1-2H} \quad 0.5 < H < 1.0$$

or in the log-log scale

$$\log f(\lambda) \stackrel{\lambda \rightarrow 0}{\sim} (1 - 2H) \log \lambda + \log c$$

The basic idea of these methods is to use a Fast Fourier Transform algorithm to compute the periodogram

$$I_n(\lambda) = \frac{1}{n} \left| \sum_{t=1}^n X_t e^{it\lambda} \right|^2$$

at the Fourier frequencies $\frac{2\pi j}{n}$ with $j = 1, 2, \dots, [\frac{n-1}{2}]$, and then to make a least square fit of $\log I(\lambda)$ versus $\log \lambda$ near the origin. It is important that the least square fit is made on frequencies $\lambda_{n,1}, \lambda_{n,2}, \dots, \lambda_{n,k}$ such that

$$\lambda_{n,j} \xrightarrow{n \rightarrow \infty} \lambda_j$$

with $\lambda_j \neq \lambda_{j'} \neq 0 \pmod{2\pi}$.

In particular the fit should not be made on the first Fourier frequencies $\frac{2\pi j}{n}$ with $j = 1, 2, \dots, k$ since $\lim_{n \rightarrow \infty} \frac{2\pi j}{n} = 0$.

The slope $1 - 2\hat{H}$ of the regression provides an estimate of the Hurst parameter H .

Theorem 9 Let $(X_t)_t$ be a stationary process with spectral density $f(\lambda)$; suppose there exists a positive continuous function $f^*(\lambda)$ such that

$$f(\lambda) = f^*(\lambda)|1 - e^{i\lambda}|^{1-2H}$$

with $0.5 < H < 1.0$. Also assume that $(X_t)_t$ has a one sided infinite moving average representation

$$Y_t = \sum_{k=0}^{\infty} a_k \epsilon_{t-k}$$

with $(\epsilon_t)_t$ a white noise with finite variance and $\sum_k a_k^2 < \infty$. Consider a finite number of frequencies $\lambda_{n,1}, \lambda_{n,2}, \dots, \lambda_{n,k}$ such that

$$\lambda_{n,j} \xrightarrow{n \rightarrow \infty} \lambda_j$$

with $\lambda_j \stackrel{+}{-} \lambda'_j \neq 0 \text{ mod } (2\pi)$ for $j \neq j'$. Then

$$[I(\lambda_{n,1}), I(\lambda_{n,2}), \dots, I(\lambda_{n,k})] \xrightarrow{n \rightarrow \infty}^{(d)} [f(\lambda_1)\zeta_1, \dots, f(\lambda_2)\zeta_2, \dots, f(\lambda_k)\zeta_k]$$

where $\zeta_1, \zeta_2, \dots, \zeta_k$ are k independent exponential random variables with mean 1.

It results from this Theorem that approximately when n is big

$$\log(I(\lambda_j)) = (1 - 2H) \log |\lambda_j| + \log c + \log \zeta_j$$

with $\zeta_1, \zeta_2, \dots, \zeta_k$ approximately independent identically distributed exponential random variables with mean 1. Note

$$\begin{aligned} y_j &= \log(I(\lambda_j)) \\ x_j &= \log |\lambda_j| \\ \beta &= \log c + \mathbb{E}(\log(\zeta_j)) \\ e_j &= \log \zeta_j - \mathbb{E}(\log(\zeta_j)) \end{aligned}$$

then

$$y_j = (1 - 2H)x_j + \beta + e_j$$

This is a least square fit with independent identically distributed errors with zero mean.

5 Non stationarities

5.1 Influence of non stationarities on the variance time estimate of the Hurst parameter

Suppose now that $(X_t)_t$ is a short memory process and define

$$Y_t = X_t + f_t \quad \text{with} \quad f_t = \sum_{i=1, N} f_i \mathbb{I}_{[T_{i-1}, T_i[}(t)$$

with $1 = T_0 < T_1 < \dots < T_N = n$ and with

$$\lim_{T \rightarrow \infty} \frac{T_i}{T} = c_i$$

Note $\mu_i = \lim_{T \rightarrow \infty} \frac{T_i - T_{i-1}}{T}$. We suppose that Y_t is a centered process so that

$$\sum_i \mu_i f_i = 0$$

Note

$$\begin{aligned} X_t^{(m)} &= \frac{1}{m} \sum_{u=(t-1)m+1}^{tm} X_u \\ f_t^{(m)} &= \frac{1}{m} \sum_{u=(t-1)m+1}^{tm} f_u \\ Y_t^{(m)} &= X_t^{(m)} + f_t^{(m)} \end{aligned}$$

The sample variance of $Y^{(m)}$ can be decomposed into four terms

$$\begin{aligned} \hat{var}(Y_t^{(m)}) &= \hat{var}(X_t^{(m)}) + \hat{var}(f_t^{(m)}) \\ &+ \frac{2}{T/m} \sum_{t=1}^{T/m} X_t^{(m)} f_t^{(m)} - \frac{2}{(T/m)^2} \sum_{t=1}^{T/m} X_t^{(m)} \sum_{t=1}^{T/m} f_t^{(m)} \end{aligned}$$

When m tends to ∞ it comes that

$$\begin{aligned} \hat{var}(X_t^{(m)}) &\xrightarrow{m \rightarrow +\infty} O(m^{-1}) \\ \frac{1}{T/m} \sum_{t=1}^{T/m} (f_t^{(m)})^2 &\xrightarrow{T/m \rightarrow +\infty} \sum f_i^2 \mu_i \\ \frac{1}{T/m} \sum_{t=1}^{T/m} f_t^{(m)} &\xrightarrow{T/m \rightarrow +\infty} \sum f_i \mu_i = 0 \\ \frac{1}{T/m} \sum_{t=1}^{T/m} X_t^{(m)} &\xrightarrow{T/m \rightarrow +\infty} 0 \\ \frac{1}{T/m} \sum_{t=1}^{T/m} X_t^{(m)} f_t^{(m)} &\xrightarrow{T/m \rightarrow \infty} \sum \mu_i f_i \mathbb{E}(X_t) = 0 \end{aligned}$$

and consequently

$$\hat{var}(Y^{(m)}) \xrightarrow{m \rightarrow \infty} \sum_i \mu_i f_i^2$$

which is a non null constant. The estimate produced by the variance time analysis is consequently very close to 1 when $Y_t = X_t + f_t$; in this expression X_t is a short memory process and f_t is used for modelling a series of shifts in the mean of the non stationary time series.

It is noteworthy that to establish $\hat{H} \simeq 1$ we have supposed that the series has are shifting levels at all the timescales

$$\lim_{T \rightarrow \infty} \frac{T_i}{T} = c_i$$

But in practice we have checked that in the presence of shifting levels on only one timescale the variance time estimate of the Hurst parameter is still severely positively biased, even if the shifts are very small with respect to the variance of the stationary short memory process X_t .

5.2 Influence of non stationarities on the periodogram based analysis

$$\frac{1}{T} \left| \sum Y_t \exp(-j\lambda t) \right|^2 = \frac{1}{T} \left| \sum X_t \exp(-j\lambda t) \right|^2 + \frac{1}{T} \left| \sum f_t \exp(-j\lambda t) \right|^2 + \frac{2}{T} \Re \left(\sum X_t \exp(-j\lambda t) \sum f_t \exp(j\lambda t) \right)$$

Note $e(t) = \mathbb{I}_{\mathbb{R}+}(t)$ the Heaviside function and $s_i = f_{i+1} - f_i$. Then $f_t = \sum_{0 \leq i \leq N} s_i e(t - T_i) = (\sum_{0 \leq i \leq N} s_i \delta_{T_i}) * e(t)$.

$$\begin{aligned} e(t) &\xrightarrow{FT} \frac{1}{1 - \exp(-j\lambda)} = \frac{\exp(j\lambda/2)}{2j \sin(\lambda/2)} \\ \delta_{T_i}(t) * e(t) &\xrightarrow{FT} \exp(-j\lambda T_i) \frac{\exp(j\lambda/2)}{2j \sin(\lambda/2)} \\ \sum_{i=0}^N s_i \delta_{T_i}(t) * e(t) &\xrightarrow{FT} \sum_{i=0}^N s_i \exp(-j\lambda T_i) \frac{\exp(j\lambda/2)}{2j \sin(\lambda/2)} \end{aligned}$$

and consequently the periodogram of f_t as a pole of order 2 in $\lambda = 0$

$$\frac{1}{T} \left| \sum f_t \exp(-j\lambda t) \right|^2 \xrightarrow{\lambda \rightarrow 0; T \rightarrow \infty} O\left(\frac{1}{\lambda^2}\right)$$

The series X_t and f_t are centered so that

$$\frac{2}{T} \Re \left(\sum X_t \exp(-j\lambda t) \sum f_t \exp(j\lambda t) \right) \xrightarrow{\lambda \rightarrow 0} 0$$

and consequently

$$\frac{1}{T} \left| \sum Y_t \exp(-j\lambda t) \right|^2 \xrightarrow{\lambda \rightarrow 0} O\left(\frac{1}{\lambda^2}\right) \quad (1)$$

Thus the estimate produced by least square fit of $\log(I(\lambda))$ versus $\log(\lambda)$ is very close to one; in practice we have checked that this result still holds for a shifting level process with shifts on one timescale only.

5.3 An example

In practice the presence of some non stationarities in a series without long memory can mislead to the detection of a high degree of long memory if the classical visual indexes are used. For example we consider a White Gaussian Noise with mean 0 and variance 1

$$(X_t)_t \sim WGN(0, 1)$$

and $(f_t)_t$ models some shifting levels in a process $Y_t = X_t + f_t$

$$f_t = \sum_{i=1, N} f_i \mathbb{1}_{[T_{i-1}, T_i]}(t)$$

$$Y_t = X_t + f_t$$

We simulate a series $(f_t)_t$ in which (i) the shifts are rare and (ii) the amplitude of the shifts is very small by comparison with the standard deviation $\sigma = 1$ of the White Gaussian Noise $\mathcal{N}(0, 1)$. Thus the shifting levels can not be detected at first glance in the series $Y_t = X_t + f_t$ and the series Y_t looks stationary.

Figure 1 displays (i) the shifting levels and the fractiles of the $\mathcal{N}(0, 1)$ distribution and (ii) a variance time analysis of Y_t ; this analysis produces an estimate $\hat{H} \simeq 0.86$ and yet the shifts in the mean of the process are very rare and have a small amplitude with respect to the dispersion of the series without shifts.

5.4 Local stationarity

A possible framework for overcoming the question of non stationarities is the empirical notion of local stationarity. A process is locally stationary if (i) it is not stationary, for example in the sense of the tests for stationarity, over long time scales (ii) but the hypothesis of stationarity is reasonable on shorter time scales.

For example, in a parametric model, if the parameters vary slowly with time then the process is locally stationary. If the model is Markov modulated (e.g. MMPP, MAP...) then it is possible to track the variations of the parameters. The following algorithm has been extensively studied by Mevel [Mevel97]. Note θ the set of parameters of the model then

$$\hat{\theta}_{t+1} = \hat{\theta}_t + \gamma \frac{\partial}{\partial \theta} \log p(x_{t+1} \mid x_t, x_{t-1}, \dots; \theta) \mid_{\theta=\hat{\theta}_t}$$

where γ is a step. Suppose for example that $(X_t)_t$ is a Hidden Markov Chain that is to say that there exists a Markov chain $(A_t)_t$ with state space \mathcal{S} such that X_t depends on A_t only; then the probability density of the random variable X_{t+1} given that $\{X_t = x_t, X_{t-1} = x_{t-1}, \dots\}$ in the model with parameter θ is

$$p(x_{t+1} \mid x_t, x_{t-1}, \dots; \theta) = \sum_{i \in \mathcal{S}} p(x_{t+1} \mid Z_{t+1} = i) \alpha_{t+1}(i; \theta)$$

where $\alpha_{t+1}(\bullet; \theta)$ is the prediction filter

$$\alpha_{t+1}(i; \theta) = \mathbb{P}(X_{t+1} = i \mid x_t, x_{t-1}, \dots; \theta)$$

$$\frac{\partial}{\partial \theta} \log p(x_{t+1} \mid x_t, x_{t-1}, \dots; \theta) = F(\alpha_{t+1}(\bullet; \theta), \frac{\partial}{\partial \theta} \alpha_{t+1}(\bullet; \theta), x_{t+1})$$

The prediction filter and its gradient can be computed recursively in a lattice in a joint manner

$$\begin{aligned} \alpha_{t+1}(\bullet; \theta) &= G(\alpha_t(\bullet; \theta), x_t, \theta) \\ \frac{\partial}{\partial \theta} \alpha_{t+1}(\bullet; \theta) &= H(\alpha_t(\bullet; \theta), \frac{\partial}{\partial \theta} \alpha_t(\bullet; \theta), x_t, \theta) \end{aligned}$$

6 Near completely decomposable Markov models

6.1 Near complete decomposability in Markov chains

Broadly speaking, a system is near completely decomposable when its components can be grouped into subsystems, in such a manner that the interactions between the subcomponents are much weaker than the interactions within each of the subcomponents. This structure of dependence conducts to the following dynamics [Courtois77] :

- (i) short term dynamics each of the subsystems moves towards equilibrium under the influence of the strong interactions within the subsystem
- (ii) short term equilibrium an internal equilibrium is reached within each of the subsystems
- (iii) long term dynamics the complete system as a whole moves towards a global equilibrium under the influence of the small interactions between subsystems
- (iv) long term equilibrium the global equilibrium is attained

Remark that the short term equilibrium that is attained within each of the subsystems (phase (ii)) is maintained when the system, as a whole, moves towards a global equilibrium (phase (iii)). In many contexts the interactions within a discrete system can be represented by a graph that is composed of nodes and of branches; a 'length' is assigned to each branch; this length corresponds to the 'degree' of interactions between the two nodes.

This graphical representation can also represent the dependence structure of a discrete time Markov chain. In this context the nodes represent the states of the Markov chain and the 'lengths' are the probabilities of transition from one node to another. Denote by $(Y_n)_n$ this Markov chain and denote by P its stochastic matrix : $P(i, j) = \mathbb{P}(Y_{n+1} = j \mid Y_n = i)$.

$(Y_n)_n$ is near completely decomposable when its state space $\{1, 2, \dots, n\}$ can be grouped into N subspaces $\{1_I, 2_I, \dots, n(I)_I\}$ of cardinal $n(I)$ in such a way that the probabilities of transition between subspaces are very small. This means that, up to a convenient reordering of the states $\{1, 2, \dots, n\}$, the stochastic matrix P is in form of $P = P^* + \epsilon C$, where ϵ is a small positive real number, the maximum degree of coupling between the subsystems in P . P^* is completely decomposable : it is block diagonal, and each block (I, I) is a stochastic matrix, the matrix of the probabilities of transition in subsystem I ; we suppose that each submatrix P_I^* is irreducible. C is a real matrix such that $C[1, 1, 1, \dots, 1]^T = [0, 0, \dots, 0]^T$, and that the elements in block (I, J) are non positive if $I = J$ and non negative if $I \neq J$.

$$P^* = \begin{bmatrix} P_1^* & & & \\ & P_2^* & & \\ & & \ddots & \\ & & & P_I^* \end{bmatrix}$$

P^* is the stochastic matrix of a reducible Markov chain, say $(Y_n^*)_n$. Denote by $\pi(t)$ the distribution of the chain Y at time t and $\pi^*(t)$ the distribution of the chain Y^* at time t : $\pi(t) = \pi(0)P^t$ and $\pi^*(t) = \pi^*(0)(P^*)^t$. It results from the two Simon and Ando theorems [Courtois77] that if ϵ is sufficiently small then on the short time $\pi(t)$ varies like $\pi^*(t)$. Indeed consider the idempotent expansion of P and P^* :

$$\begin{aligned} P &= Z(1_I) + \sum_{I=2,N} \lambda(1_I)Z(1_I) + \sum_{I=1,N} \sum_{i=2,n(I)} \lambda(i_I)Z(i_I) \\ P^* &= \sum_{I=1,N} Z^*(1_I) + \sum_{I=1,N} \sum_{i=2,n(I)} \lambda^*(i_I)Z^*(i_I) \end{aligned} \quad (2)$$

$1 = \lambda^*(1_I) > \lambda^*(2_I) \geq \dots \geq \lambda^*(n_I)$ are the eigenvalues of the irreducible stochastic matrix $P^*(I)$. Denote by $u^*(i_I)$ a left line eigenvector and $v^*(i_I)$ a right column eigenvector, corresponding to the eigenvalue $\lambda^*(i_I)$; $u^*(i_I)$ and $v^*(i_I)$ are 'filled' with zeros at the 'good' positions to be n -dimensional. Denote by $Z^*(i_I) = v^*(i_I)u^*(i_I) < v^*(i_I), u^*(i_I) >^{-1}$, where $< v^*(i_I), u^*(i_I) >$ denotes the scalar product of $v^*(i_I)$ and $u^*(i_I)$; the family $Z^*(i_I)$ checks some properties of idempotency and orthogonality : $Z^*(i_I)Z^*(j_J) = 0$ if $i \neq j$ or $I \neq J$ and $Z^*(i_I)Z^*(i_I) = Z^*(i_I)$, so that the $Z^*(i_I)$ are orthogonal projectors.

The eigenvalues $\lambda(i_I)$ and the projectors $Z(i_I)$ are defined in the corresponding way. The eigenvalues and eigenvectors of a matrix being countinuous functions of its elements, there exists a one to one mapping between the $\lambda(i_I)$ and the $\lambda^*(i_I)$, and between the $Z(i_I)$ and the $Z^*(i_I)$.

It results from these properties of idempotency and of orthogonality that

$$\begin{aligned} P^t &= Z(1_I) + \sum_{I=2,N} \lambda^t(1_I)Z(1_I) + \sum_{I=1,N} \sum_{i=2,n(I)} \lambda^t(i_I)Z(i_I) \\ (P^*)^t &= \sum_{I=1,N} Z^*(1_I) + \sum_{I=1,N} \sum_{i=2,n(I)} (\lambda^*(i_I))^t Z^*(i_I) \end{aligned} \quad (3)$$

If ϵ is sufficiently small then $\lambda(1_I), I = 2, N$ are close to 1 and on the short term $\lambda^t(1_I)$ does not vary. The dominant term in the right member of the two equations in 3 are the third term in the expression of P^t and the corresponding second term in the expression of $(P^*)^t$, so that $\pi(t)$ and $\pi^*(t)$ vary similarly (short term dynamics). Once these terms have vanished the two systems attain a short term equilibrium. Then the second term in the expression of P^t is predominant : this is the long term dynamics. Once this term has vanished the system as a whole attains its equilibrium : $\lim_t \pi(t) = \pi(0)Z(1_I)$.

This structure of dependence can be generalized to a hierarchical near completely decomposable structure : each $P^*(i)$ can be a near completely decomposable stochastic matrix, very close to a block diagonal matrix with blocks that are themselves near completely decomposable, and so on... This is a possible explanation, within the Markov framework, of the 'burstiness' that one observes in the real traffic over several time scales.

6.2 Approximation of long memory processes by near completely decomposable Markov models

Consider now a near completely decomposable Markov chain $(Y_t)_t$ with stochastic matrix P , and consider $(X_t)_t$ a stochastic function of $(Y_t)_t$ such that X_t depends on Y_t only : this defines a near completely decomposable Hidden Markov Chain. It results from the idempotent expansion of P^k in Eq. 3, that the autocorrelation function of $(X_t)_t$ is

$$\begin{aligned} r_X(k) &= u(1_I)M \sum_{I=2}^N \lambda^k(1_I)Z(1_I)Mv(1_I) < u(1_I), v(1_I) >^{-1} \\ &\quad + u(1_I)M \sum_{I=1}^N \sum_{i=2}^{n(I)} \lambda^k(i_I)Z(i_I)Mv(1_I) < u(1_I), v(1_I) >^{-1} \end{aligned} \quad (4)$$

where $M = \text{diag}[\mathbb{E}(X_t | Y_t = 1), \dots, \mathbb{E}(X_t | Y_t = n)]$.

During the 'short term dynamics' period, the first term in this expansion of $r_X(k)$ does not vary significantly and $r_X(k)$ changes as $u(1_1)M \sum_{I=1}^N \sum_{i=2}^{n(I)} \lambda^k(i_I)Z(i_I)Mv(1_1) < u(1_1), v(1_1) >^{-1}$. During the 'long term dynamics' period, the last term in the expansion of $r_X(k)$ has vanished and $r_X(k)$ varies like $u(1_1)M \sum_{I=2}^N \lambda^k(1_I)Z(1_I)Mv(1_1) < u(1_1), v(1_1) >^{-1}$; the eigenvalues $\lambda(1_I), I = 2, N$ are close to 1 so that the correlations $r_X(k)$ of $(X_t)_t$ decrease very slowly on this medium time scale : this weighted sum of geometric sequences with rates close to 1 make it possible to approximate the power law decrease of the correlations of long range dependent processes. On the very long time scales, asymptotically when k tends to infinity, the long term equilibrium has been attained, and $r_X(k)$ decreases with a geometric rate $\lambda(2_I)$, as it is the case for irreducible Markov chains : $r_X(k) \sim u(1_1)M\lambda^k(2_I)Z(2_I)Mv(1_1) < u(1_1), v(1_1) >^{-1}$.

In a similar manner, the variance time plot index for self similarity is misleading when the process is near completely decomposable. If the size of the averaging window is bigger than the time required to establish the long term equilibrium, the variance of the averaged process $m^{-1}(X_{(t-1)m+1} + \dots + X_{tm})$ is $O(m^{-1})$, as it would be the case if the X_t were independent realizations of a random variable : indeed an equilibrium is attained within each of the non overlapping windows, and the variability between windows of the mean value in the window is the same as if the windows were independent. But for 'intermediate' sizes m of the averaging window, the global equilibrium has not been attained within each of the window, and the variance of the averaged process decreases more slowly than $O(m^{-1})$.

6.3 A near completely Markov model with 3 parameters

S.Robert [Robert et al.96, Robert et al.97] has proposed a model that has a number of advantages :

- (i) it is parcimonious; indeed this model has only three parameters a , b and n .
- (ii) it is a Markov model so that standard queuing analysis (e.g. matrix geometric analysis) can be used to evaluate performances.
- (iii) asymptotically when n tends to ∞ the model is "pseudo" long range dependent that is to say that

$$\begin{aligned} \sum_{k \in \mathbb{Z}} \rho(k) &= +\infty \\ \lim_{\lambda \rightarrow 0} S(\lambda) &= +\infty \\ \text{var}(X^{(m)}) &\stackrel{m \rightarrow \infty}{\sim} O(m^{-\beta}) \quad \text{with } 0 < \beta < 1 \end{aligned}$$

where $\rho(k) = \text{corr}(X_t, X_{t+k})$, $S(\lambda)$ is the power density of $(X_t)_t$ and $\text{var}(X^{(m)})$ is the variance of the process $X_t^{(m)} = m^{-1} \sum_{u=(t-1)m+1, tm} X_u$.

The model proposed by S.Robert is following; set n a positive non null integer, and a and b two reals such that $0 < b < a$ and $a > 1$.

$$A = \begin{pmatrix} 1 - 1/a - \dots - 1/a^{n-1} & 1/a & 1/a^2 & \dots & 1/a^{n-1} \\ b/a & 1 - b/a & 0 & \dots & 0 \\ (b/a)^2 & 0 & 1 - (b/a)^2 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ (b/a)^{n-1} & 0 & \dots & 0 & 1 - (b/a)^{n-1} \end{pmatrix}$$

Denote by $(Y_t)_t$ a Markov chain with stochastic matrix A and consider the ON/OFF process $(X_t)_t$ characterized by

$$X_t = \begin{cases} 1 & \text{if } Y_t = 1 \\ 0 & \text{otherwise} \end{cases}$$

S.Robert states that asymptotically when n tends to ∞ the ON/OFF is “pseudo” long range dependent in the sense defined above.

6.4 The near completely decomposable SEHMM

We have analysed carefully a trace gathered at the Lawrence Berkeley Laboratory’s wide-area Internet gateway (LBL-PKT3, [Paxson et al.94]). Paxson and Floyd found out a high degree of self-similarity in LBL-PKT3 and conclude their seminal paper of 1994 ([Paxson et al.94], pp.12–13) by these words :

Modeling TCP traffic using Poisson or other models that do not accurately reflect the long-range dependence in actual traffic will result in simulations and analyses that significantly underestimate performance measures such as average packet delay or maximum queue size.(...) In summary: we should abandon Poisson-based modeling of wide-area traffic.

We have proposed a model for modelling the series of the interarrival times. The results of our analysis confirm the intuitions of S.Robert. We have found that a 5 states Markov model, that we call the Shifted Exponential Hidden Markov Model, fits the measured traffic extremely well.

Note $(Y_t)_t$ a Markov chain with stochastic matrix A and note $(X_t)_t$ the doubly stochastic process such that X_t depends on Y_t only and given $Y_t = i$, X_t is exponential with offset θ_i and with mean $\theta_i + 1/\lambda_i$

$$\begin{aligned} [X_t | Y_t = i] &\stackrel{(d)}{=} s_i + \lambda_i Z \quad Z \sim \mathcal{E}(1) \\ p_{[X_t|Y_t=i]}(x) &= \lambda_i \mathbb{I}_{x \geq \theta_i} \exp(-\lambda_i(x - \theta_i)) \end{aligned}$$

The parameters that fit LBL-PKT3 is following :

Number of states: $n=5$

Offsets:

$$\theta_1 = 3.60 \cdot 10^{-4} \quad \theta_2 = 3.60 \cdot 10^{-4} \quad \theta_3 = 8.0 \cdot 10^{-4} \quad \theta_4 = 3.60 \cdot 10^{-4} \quad \theta_5 = 3.60 \cdot 10^{-4}$$

Intensities of the exponentials:

$$\lambda_1 = 6000 \quad \lambda_2 = 6000 \quad \lambda_3 = 78 \quad \lambda_4 = 6000 \quad \lambda_5 = 3$$

Stochastic matrix of the modulating Markov chain :

$$A = \begin{pmatrix} \mathbf{0.95} & 0.007 & 0.007 & 0.006 & 0.03 \\ 0.008 & \mathbf{0.04} & \mathbf{0.95} & 0.001 & 0.001 \\ 0.008 & \mathbf{0.95} & \mathbf{0.04} & 0.001 & 0.001 \\ 0.025 & 0.025 & 0.025 & \mathbf{0.025} & \mathbf{0.9} \\ 0.008 & 0.001 & 0.001 & \mathbf{0.95} & \mathbf{0.04} \end{pmatrix}$$

A is the stochastic matrix of a near completely decomposable Markov chain. We have checked the the model that we propose passes a number of tests, among which :

- (i) the marginal distribution of the interarrival times
- (ii) the correlations up to lag $k = 100$
- (iii) visually the power density spectrum of the near completely decomposable Markov model displays the same low frequency component as the smoothed periodogram of the LBL-PKT3 series

(iv) the variance time analysis produces approximately the same Hurst estimate in the case (a) of LBL-PKT3 and (b) of a series simulated according to the model proposed, provided that the estimation of H is performed on the range of m for which $\log \hat{var}(X^{(m)})$ versus $\log m$ is approximately linear

(v) and the probability of loss when the series is queue into a queue with finite capacity is correctly approximated by the model proposed, and this is the best validation of the model proposed!

We refer the reader to our research paper in these proceedings for more information [Vaton00].

7 Aggregation models

Some authors propose a tentative “physical” justification of long memory. They prove that the superposition of a high number of independent short memory processes can in some cases be a long memory process. In the framework of telecommunication networks the statistical multiplexing could produce some long memory effects in the multiplexed traffic. There is nevertheless some limitation to this tentative explanation : asymptotically when the number of superposed sources tends to infinity the limit process is Gaussian and; but in practice the gaussianity is hardly verified on the real traffic.

Superposition of AR(1)

Granger and Joyeux [Granger et al.80] consider the independent superposition of some autoregressive AR(1) processes with poles a_i . Denote by X_t^i the AR(1) with power density

$$S_i(\lambda) = |1 - a_i e^{-i\lambda}|^{-2} \frac{\sigma_i^2}{2\pi} \quad (5)$$

and consider spatial averaging of K independent AR(1)

$$X_t = K^{-1} \sum_{i=1, K} X_t^i$$

The power density of X_t is

$$K^{-1} \sum_{i=1, K} S_i(\lambda) = K^{-1} \sum_{i=1, K} |1 - a_i e^{-i\lambda}|^{-2} \frac{\sigma_i^2}{2\pi}$$

This aggregation can be generalized to a continuum of processes X_t^i . We now suppose that a_i takes a continuum of values; the pole A of the AR(1) is a random variable with cumulative distribution function

$$F(a) = \mathbb{P}(A \leq a)$$

Then the spatial averaged process has power density

$$S(\lambda) = \int |1 - a e^{-i\lambda}|^{-2} \frac{\sigma^2}{2\pi} dF_A(a)$$

Granger and Joyeux prove that if $F_A(\bullet)$ is a Beta distribution with suitable parameters then $S(\lambda)$ has an integrable pole at zero and thus the process obtained by aggregation has long memory even if it results from the superposition of AR(1) processes.

Superposition of ON/OFF sources

We now consider the superposition of some independent ON/OFF sources. The source can be in two states :

- (i) in the ON state one packet is emitted
- (ii) in the OFF state there is no emission of packet

The distribution of the duration of the ON periods is geometric with mean $1/(1 - \alpha_i)$

$$(1 - \alpha_i)\alpha_i^{k-1}, \quad k = 1, 2, \dots$$

and the distribution of the duration of the OFF periods is geometric with mean $1/(1 - \beta_i)$

$$(1 - \beta_i)\beta_i^{k-1}, \quad k = 1, 2, \dots$$

The succession of the states of the source (ON or OFF) is a Markov chain with stochastic matrix

$$A = \begin{pmatrix} \alpha & 1 - \alpha \\ 1 - \beta & \beta \end{pmatrix}$$

Geerts and Blondia [Geerts et al.98] consider a countable superposition of such ON/OFF sources

$$X_t = \sum_{i=1}^{\infty} X_t^i$$

The correlations of X_t are not absolutely summable

$$\sum_{k \in \mathbb{Z}} |corr(X_t, X_{t+k})| = \infty$$

if and only if

$$\sum_{i \in \mathbb{N}} \frac{1 - \alpha_i}{(1 - \beta_i)^2} = +\infty$$

Blondia et al. propose different series α_i and β_i that check this relation. Figure 2 displays the ON/OFF before aggregation and Figure 3 displays the power density of the aggregation of $n = 1$, $n = 3$, $n = 5$ and $n = 10$ ON/OFF sources.

8 Conclusion

In recent years traditional Markov modelling has been strongly questioned; indeed classical indexes display a strong degree of long memory in the traffic gathered on real networks. But Markov modelling still makes sense :

(i) Near completely decomposable Markov models approximate the strong correlations of long memory processes, at least on the first time scales. In practice this approximation on a finite number of time scales is enough because the queues with finite capacity in the network are likely to break up correlations on the very long time scales. What is more it is possible to develop Markov models that are completely decomposable with only a few parameters (parimonious models), and these models are long range dependent in the sense that the correlations are not summable.

(ii) It is possible to overcome the problem of long memory if one considers that the source is locally stationary. An adaptive allocation of the ressources (bandwidth...) that is based on a very simple Markov model -e.g. ON/OFF with geometric distributions for the ON and OFF periods- with parameters that are changing slowly along the time gives in practice good results [Vaton98][Vaton00], in the sense that one is able to control adaptatively the Quality of Service.

The main advantages of these approaches is that it is possible to re-use traditional performance analysis and that there are well known technical solutions for the estimation of the parameters of Markov models, in particular in a non stationary context. The combination of adaptive estimation of the parameters of the source and of adaptive allocation of ressources gives good results. The numerical complexity of the algorithms is low provided that the model proposed is simple (e.g. ON/OFF with geometric ON and OFF durations) so that very long sequences can be processed in real time.

References

- [Courtois77] Courtois (P.J.). – *Decomposability: queuing and computer system applications*. – New York, Academic Press, 1977.
- [Duffield et al.] Duffield (N.), Lewis (G.), O’Connel (N.) et Toomey (F.). – Statistical issues raised by the Bellcore data. – Preprint.
- [Erlang09] Erlang (A.K.). – The theory of probabilities and telephone conversations. *Nyt Tidsskrift Matematik*, n° B 20, 1909, pp. 33–39.
- [Erlang17] Erlang (A.K.). – Solution of some problems in the theory of probabilities of significance in automatic telephone exchanges. *Elektrotekniker*, n° 13, 1917, pp. 5–13.
- [Fischer et al.93] Fischer (W.) et Meier-Hellstern (K.S.). – The Markov Modulated Poisson Process (MMPP) Cookbook. *Performance Evaluation*, vol. 18, n° 2, 1993, pp. 149–171.
- [Geerts et al.98] Geerts (F.) et Blondia (C.). – Superposition of Markov sources and long range dependence. *Broadband Communications : the future of telecommunications*, pp. 550–561. – 1998.
- [Granger et al.80] Granger (C.W.J.) et Joyeux (R.). – An introduction to long range time series model and fractional differencing. *Journal of Time Series Analysis*, vol. 1, 1980, pp. 15–30.
- [Leland et al.94] Leland (W.E.), Taqqu (M.S.), Willinger (W.) et Wilson (D.V.). – On the Self-Similar Nature of Ethernet Traffic (Extended Version). *IEEE ACM Transactions on Networking*, vol. 2, n° 1, feb 1994, pp. 1–14.

- [Mevel97] Mevel (L.). – *Statistique asymptotique pour les modèles de Markov cachés*. – Thèse, Université de Rennes 1, 1997.
- [Paxson et al.94] Paxson (V.) et Floyd (S.). – Wide-Area Traffic: The Failure of Poisson Modeling. *SIGCOMM'94*, pp. 257–268. – London, England, aug 1994.
- [Robert et al.96] Robert (S.) et Boudec (J.Y. Le). – On a Markov Modulated Chain with Pseudo Long-Range Dependences. *Performance Evaluation*, vol. 27–28, oct. 1996, pp. 159–173.
- [Robert et al.97] Robert (S.) et Boudec (J.Y. Le). – New Models for Self-Similar Traffic. *Performance Evaluation*, vol. 30, n° 1–2, july 1997, pp. 57–68.
- [Vaton98] Vaton (S.). – *Modélisation statistique de trafic sur réseau local: application au contrôle dynamique de bande passante*. – Paris, Thèse, ENST, dec. 1998.
- [Vaton00] Vaton (S.). – Near completely decomposable markov models of traffic. *In these Proceedings*. – 2000.
- [Willinger et al.95a] Willinger (W.), Taqu (M.), Leland (W.E.) et Wilson (D.V.). – Self-Similarity in High-Speed Packet Traffic: Analysis and Modeling of Ethernet Traffic Measurements. *Statistical Science*, vol. 10, n° 1, 1995, pp. 67–85.
- [Willinger et al.95b] Willinger (W.), Taqu (M.), Sherman (R.) et Wilson (D.V.). – Self-Similarity Through High Variability: Statistical Analysis of Ethernet LAN Traffic at the Source Level. *SIGCOMM'95*. – Cambridge, USA, jan 1995.

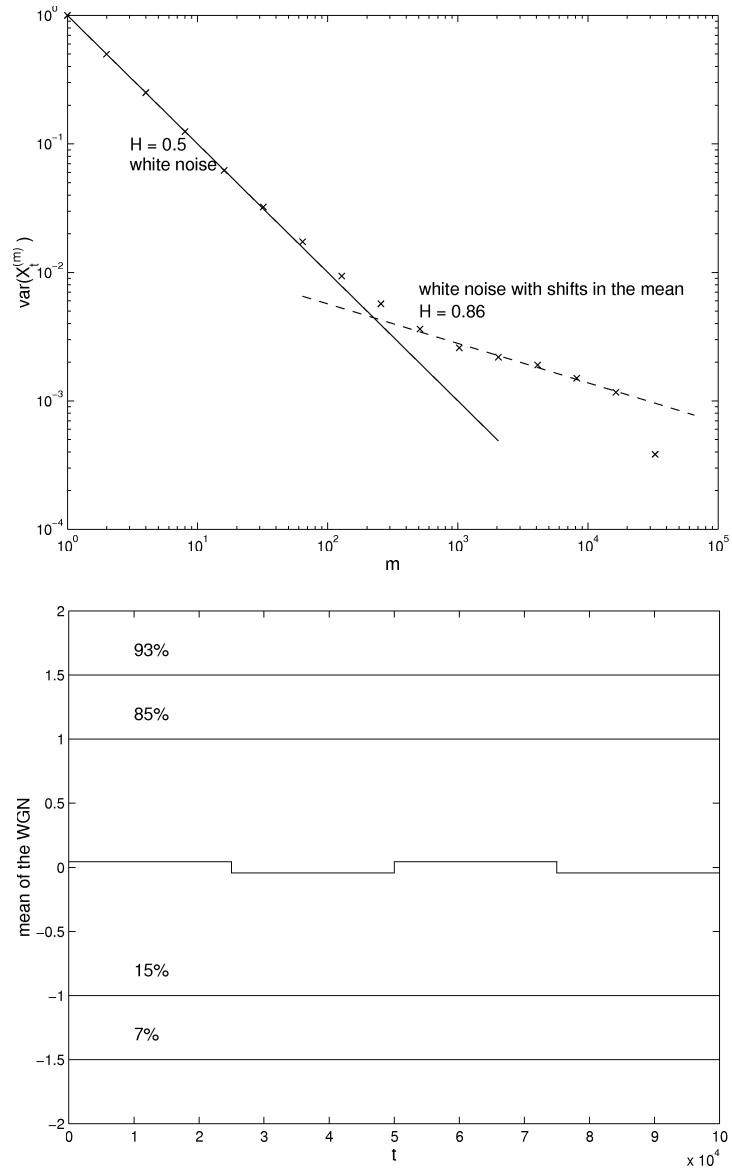


Figure 1: Variance time analysis: influence of deterministic jumps in the mean.

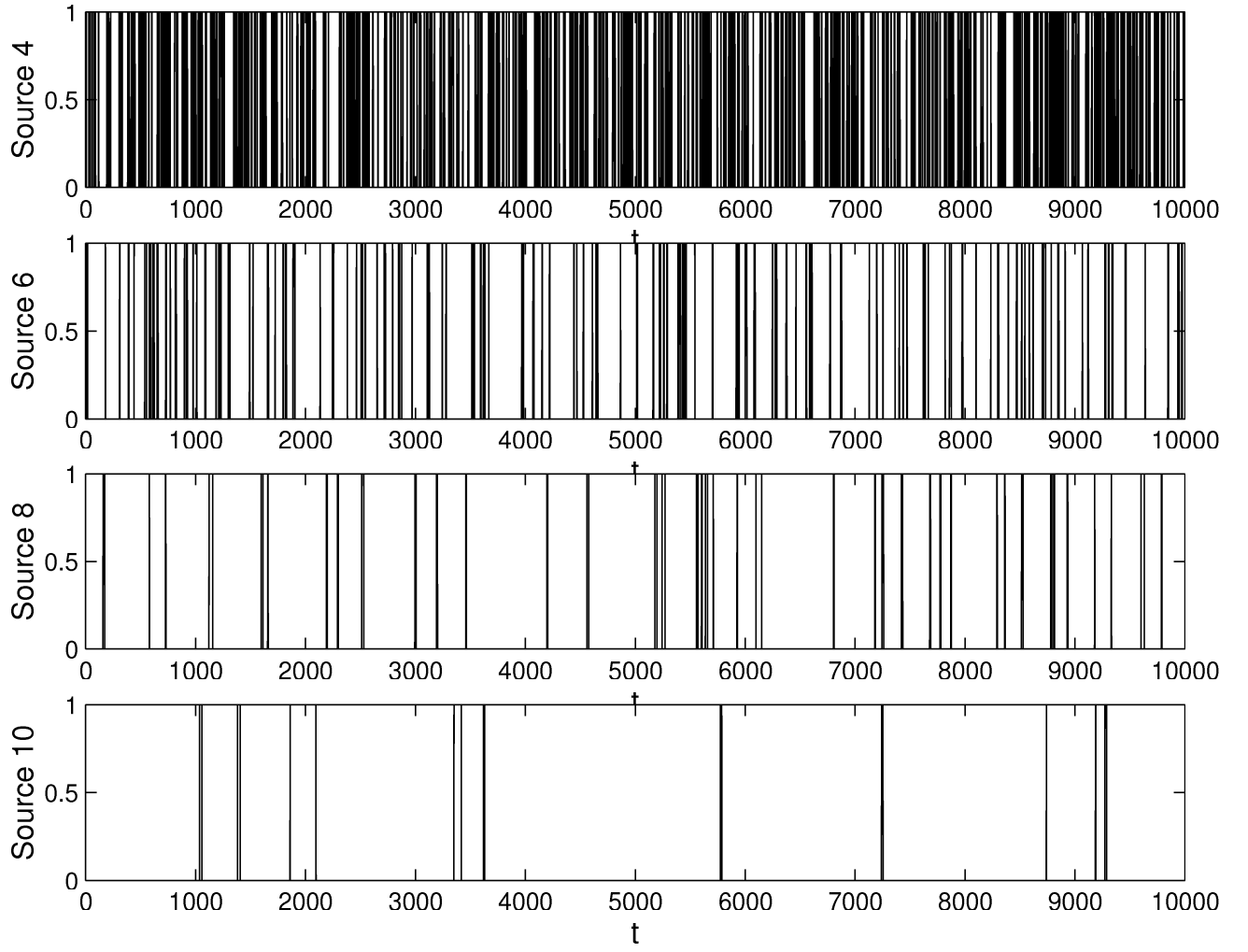


Figure 2: Sources ON/OFF before spatial aggregation.

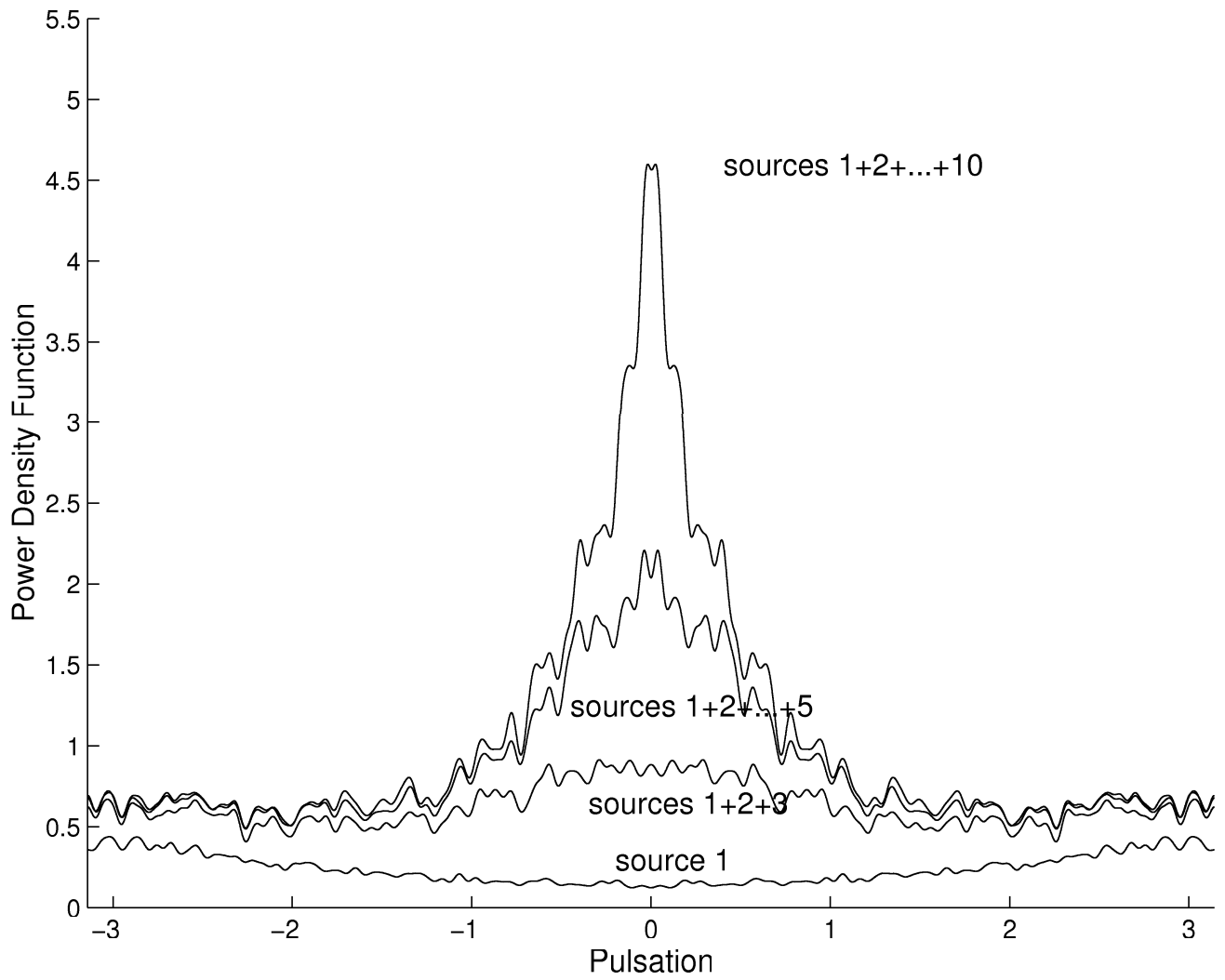


Figure 3: Power density of the spatial aggregation of ON/OFF sources.