APPLIED STOCHASTIC MODELS IN BUSINESS AND INDUSTRY
Appl. Stochastic Models Bus. Ind., (in press) (DOI: 10.1002/asmb.542)

Long-range dependence and performance in telecom networks

Pier Luigi Conti[1,‡], Antonio Lijoi[2,*†] and Fabrizio Ruggeri[3,§]

[1] *Dipartimento di Statistica, Probabilità e Statistica Applicate, Università di Roma "La Sapienza", Italy*
[2] *Dipartimento di Economia Politica e Methodi Quantitativi, Università di Pavia, Italy*
[3] *Istituto di Matematica Applicata e Tecnologie Informatiche, C.N.R., Milano, Italy*

SUMMARY

Telecommunications systems have recently undergone significant innovations. These call for suitable statistical models that can properly describe the behaviour of the input traffic in a network. Here we use fractional Brownian motion (FBM) to model cumulative traffic network, thus taking into account the possible presence of long-range dependence in the data. A Bayesian approach is devised in such a way that we are able to: (a) estimate the Hurst parameter H of the FBM; (b) estimate the overflow probability which is a parameter measuring the quality of service of a network: (c) develop a test for comparing the null hypothesis of long-range dependence in the data versus the alternative of short-range dependence. In order to achieve these inferential results, we elaborate an MCMC sampling scheme whose output enables us to obtain an approximation of the quantities of interest. An application to three real datasets, corresponding to three different levels of traffic, is finally considered. Copyright © 2004 John Wiley & Sons, Ltd.

KEY WORDS: fractional Brownian motion; long-range dependence; overflow probability; teletraffic data

1. INTRODUCTORY ASPECTS AND MOTIVATIONS

The introduction of new technologies for telecommunications, based on packet switched networks, has led to new teletraffic problems. Due to the exponential growth of the Internet, the study and analysis of telecommunications is of considerable and increasing importance. In Internet communications, the transmission of data files, e-mail, video signals, etc. generates an information stream. The user generating such a stream is commonly referred to as a traffic source. The information stream produced by a traffic source is segmented into variable size packets called *datagrams*, according to the Internet Protocol (IP, for short); see Reference [1]. A datagram is composed by a header area and a data area. The header area essentially contains routing information, i.e. the source and destination IP addresses, as well as the information to interpret the data area. IP datagrams are routed through IP routers, which interconnect input links with output links. Output links are equipped with buffers to store and schedule IP datagrams for transmission.

*Correspondence to: Antonio Lijoi, Dipartimento di Economia Politica e Methodi Quantitativi, Università di Pavia, Italy.
† E-mail: lijoi@unipv.it
‡ E-mail: pierluigi.conti@uniroma1.it
§ E-mail: fabrizio@mi.imati.cnr.it

Received 8 November 2002
Revised 22 October 2003
Accepted 15 March 2004

Apart from the case of dedicated telephone lines, IP packets are transmitted using standard commercial telephone lines. The common transport tool in telephone lines is the asynchronous transfer mode (ATM) technique; see, e.g. Reference [2]. In ATM, the information stream produced by a user (traffic source) is split into fixed-length packets. To identify both the source and the destination, a fixed-length label is added to each information packet, to form an ATM *cell*. According to their labels, cells are routed through ATM nodes connecting input links with output links. The typical behaviour of a traffic source consists of alternating activity (ON) and silence (OFF) periods. The transmission rate is usually constant during each ON period. Because of the presence of several traffic sources simultaneously connected, it is far from being unusual that different cells simultaneously require the same output link. To overcome the competition among these cells, a buffer is used, and cells that cannot be immediately transmitted are stored in it. This means that an ATM element is characterized by a *queue* of cells at the output link. Since the buffer size is finite, a cell entering the system when the buffer is already full cannot be either transmitted or stored in the buffer and, then, it is *lost*.

New standards have been recently defined in order to transmit IP datagrams by ATM technique *via* 'Cell switch routers' (see References [1,3,4]). The basic idea consists in interconnecting IP routers by ATM links. IP datagrams are first fragmented into ATM cells, then transmitted by an ATM link and finally reassembled. If one ATM cell originated by the fragmentation of an IP datagram is lost in the output buffer, then all the other ATM cells belonging to the same IP datagram are lost, and the IP datagram must be retransmitted.

Packet switching networks are essentially networks of queues. The evaluation of the performance of a single server queue, the ATM multiplexer, composed by an ATM link with buffer, is a fundamental step in assessing the performance of ATM networks and, because of the use of ATM switching fabrics in IP routers, of IP networks. This last point is particularly relevant, because of the recent growth of applications using Internet protocol. IP technology does not eliminate the need to deploy ATM networks, because ATM offers a standard set of traffic management mechanisms that can inter-operate among different providers to allow efficient support for different types of services and effectively guarantee a good quality of service (QoS) to the connections.

Telecommunication networks are characterized by QoS requirements. A fundamental QoS parameter is the *cell loss probability*, as suggested in Reference [5] and the ITU-T Recommendation I.356 Reference [6]. The cell loss probability is defined as the 'long term fraction of lost cells'. As a convenient approximation of it, the *overflow probability* is commonly used; see, e.g. Reference [5]. It is defined as the probability that the number of cells in an infinite queue exceeds the buffer size. Cell loss probability is the most important QoS parameter in ATM and/or IP networks. As already stressed, in applications using the IP protocol, if an ATM cell is lost, then the corresponding IP datagram is completely lost, and must be retransmitted. This may clearly cause the congestion of communications networks and delay in the transmission. The higher the cell loss probability, the stronger the phenomena of congestion and transmission delay.

In order to guarantee an acceptable QoS, it is of primary importance to have at least an idea of the corresponding cell loss probability (or better, of the overflow probability). In ATM networks, the traffic is controlled by the connection admission control (CAC) function. Before establishing a new connection, the corresponding source is asked to declare some standard intensity traffic parameters. On the basis of such parameters, the CAC function computes the cell loss probability, and then checks if the system has enough resources to accept the new connection without infringing QoS requirements. On the other hand, in IP networks there is no preventive traffic control. Sources do not declare any intensity traffic parameters, so that the cell loss probability cannot be computed. This motivated the need of estimating the cells loss probability on the basis of observed data. In fact, the evaluation of the cell loss probability allows one to answer some basic problems, such as determining the utilization level of a link such that the QoS requirements are met. Therefore the statistical estimation of the buffer overflow probability is unavoidable if one wants to assess the performance of ATM and/or IP networks.

The main contribution of the present paper is a Bayesian approach to the estimation of the overflow probability. In detail, Section 2 contains a description of the model and the related and related statistical problems. In Section 3, a Bayesian technique to estimate the overflow probability is developed. Finally, in Section 4 an application to real data is considered. Data come from measurements made in Italy by Telecom Italia, in the framework of the European ATM Pilot Project. The applications considered are videoconference, teleteaching, and transport of routing information between IP network routers. All applications use the Internet Protocol over ATM, as described above.

2. THE MODEL AND ITS MOTIVATIONS

Stochastic models for packed switched traffic traditionally fall into one of two categories: burst scale models, and cell scale models; see Reference [5, pp. 309, 310, 389] for a good description. Burst scale models are based on the fluid flow approximation for the packet stream produced by a source. Cell level models are primarily useful for ATM traffic data; cf. e.g. Reference [7]. They are essentially based on the idea that all transmission systems work in discrete time. In fact, there exists an elementary time unit, the *time slot*, such that no more than one cell per time slot can be transmitted. The relative merits of these two different approaches to modelling teletraffic phenomena is briefly discussed, for instance, in Reference [5, Chapters 16, 17]. We adopt here the burst scale approach, which proves useful especially when studying the characteristic of the aggregated traffic produced by several users simultaneously connected.

Suppose that N sources are simultaneously connected to a traffic node, and let $A_i(t)$ be the amount of traffic generated by the ith source during the time interval $(0, t]$, $t > 0$, $i = 1, \ldots, N$. Furthermore, let

$$X(t) = \sum_{i=1}^{N} A_i(t) \tag{1}$$

be the global amount of traffic generated by the N sources up to time t. In the sequel, the stochastic process $(X(t);\ t > 0)$ will be referred to as the 'cumulative arrival process'. Suppose that the service time (i.e. the channel capacity) is constant, and equal to c, and let V be the unfinished work of the system. It is not difficult to show (cf. Reference [8]) that the following equality in distribution holds true:

$$V \stackrel{\mathrm{d}}{=} \sup_{t \geqslant 0} (X(t) - ct) \tag{2}$$

Relationship (2) is of basic importance in studying the performance of telecommunication systems. Let u be the buffer size. Then, the *overflow probability*, closely related to the loss probability, is equal to

$$Q(u) = P(V > u) = P\left(\sup_{t \geqslant 0} (X(t) - ct) > u \right) \tag{3}$$

As mentioned before, the overflow probability is the most important measure of performance for telecommunication systems, since it is a good approximation of the loss probability for buffered systems. Formula (3) shows that the problem of evaluating the overflow probability essentially consists of studying the distribution of the supremum of a stochastic process.

Hence, some assumptions on the process $(X(t);\ t>0)$ are in order. A common hypothesis is that it is a Gaussian process with stationary increments. Such an assumption essentially rests on (1) and the functional central limit theorem. The assumptions on the covariance function of $(X(t);\ t\geqslant 0)$ are more delicate. In the sequel, we will suppose that the process can be expressed in the form

$$X(t) = \mu t + \sigma Z(t) \tag{4}$$

where $(Z(t);\ t\geqslant 0)$ is a fractional Brownian motion (FBM, for short). It is characterized by the following properties:

(i) $Z(t)$ is a Gaussian process with stationary increments.
(ii) $Z(0)\ = 0$ a.s.
(iii) $E[Z(t)] = 0$ and $E[Z(t)^2] = t^{2H}$ for all positive t.

The parameter H is the *Hurst parameter*: it takes values in the interval $(\frac{1}{2}, 1)$ and, if $H = \frac{1}{2}$, $Z(t)$ reduces to the standard Brownian motion.

Model (4), with $Z(t)$ FBM, was first proposed as a realistic model for aggregated traffic by Norros [9,10]. Its most important feature is that it is a *self-similar* process: $Z(\alpha t) \stackrel{d}{=} \alpha^H Z(t)$ for every positive α. Clearly, the process $(X(t)-\mu t)$ possesses the same property. The self-similar nature of Ethernet traffic was first shown by statistical (frequentist) analysis of Bellcore traffic data. See References [11,12]. A good bibliographical guide for the subject is in Reference [13].

As a consequence, the increments of $Z(t)$ (and those of $X(t)$, as well) are stationary with *long-range dependence* whenever $H > \frac{1}{2}$. To be precise, let $Y_i\ = X(i)-X(i-1)$, $i = 1, \ldots, N$. From the well-known formula (see Reference [14], pp. 52, 56)

$$E[Z(t)Z(s)] = \tfrac{1}{2}(t^{2H} + s^{2H} - |t-s|^{2H}) \qquad \forall t,\ s\geqslant 0$$

it is seen that the correlation coefficient between the increments Y_i and Y_{i+k} is equal to

$$\rho(k) = \tfrac{1}{2}((k+1)^{2H} - 2k^{2H} + (k-1)^{2H}) \qquad \forall k\geqslant 1 \tag{5}$$

The most important property of (5) is that $\rho(k)$ tends to zero very slowly as k tends to infinity. In fact, by a Taylor expansion it is easy to see that $\rho(k)\ \sim\ H(2H-1)k^{2(H-1)}$ as $k\to\infty$, for every $\frac{1}{2}<H<1$.

Long-range dependence is essentially generated by ON and/or OFF periods with infinite variance; see Reference [5]. From a practical point of view, this means that when the traffic sources generate traffic with high variability, where the ON periods can be very long (cf. Reference [15]), one should expect that the aggregated traffic stream entering the transmission system, $X(t)$, is characterized by the presence of long-range dependence. This is important not only from a theoretical point of view. In fact, long-range dependence could potentially have a great influence on the performance of telecommunication systems, since it could considerably increase the overflow probability; see Reference [16]. Furthermore, stochastic models that do not allow for long-range dependence could severely underestimate the loss probability. These two facts provide the most important justifications to the use of model (4).

Unfortunately, even for model (4) the loss probability (3) cannot be written in a closed form. However, using a result by Hüsler and Piterbarg [17] it can be shown that, under the stability condition $\mu < c$, the following holds:

$$q_u(H,\mu,\sigma^2) := Q(u) \sim \frac{\sqrt{\pi}(c-\mu)^{1-H}(1-H)^{3/2-H-1/H}H_{2H}}{H^{3/2-H}2^{(H-1)/(2H)}\sigma^{1/H-1}}u^{(1-H)^2/H}\Psi(A\sigma^{-1}u^{1-H}) \quad (6)$$

as u tends to infinity, where $\Psi(\cdot)$ is the survival function of normal standard distribution,

$$A = \frac{(c-\mu)^H}{H^H(1-H)^{1-H}}$$

and

$$H_{2H} = \lim_{t\to\infty}(1/t)E\left[\exp\left(\max_{0\leqslant s\leqslant t}(-s^{2H}+\sqrt{2}Z(s))\right)\right] \quad (7)$$

$Z(t)$ being a FBM. Since the usual buffer size is $u = 500$, or $u = 1000$, the asymptotic approximation (6) is satisfactory.

The exact value of constant (7) is not known. Luckily enough, in Reference [18] it is shown that

$$0.12 \leqslant H_{2H} \leqslant 3.1 \quad (8)$$

Estimate (8) will be used in the sequel.

Remark 1

The most important part of relationships (8) is the upper bound 3.1, at least from a practical point of view. In fact, as already outlined in the Introduction, communication providers should guarantee a loss probability smaller than a given threshold. Hence, the upper bound in (8) is much more important than the lower bound.

Remark 2

The constant H_{2H} is given an equivalent definition in Reference [19]. It suggests, as the author himself points out, the possibility of evaluating H_{2H} numerically for some values of $2H$.

Observe that the unfinished work V is infinite a.s. whenever $\mu \geqslant c$, so that in our setting the loss probability turns out to be equal to

$$Q(u) \sim \frac{\sqrt{\pi}(c-\mu)^{1-H}(1-H)^{3/2-H-1/H}H_{2H}}{H^{3/2-H}2^{(H-1)/(2H)}\sigma^{1/H-1}}u^{(1-H)^2/H}\,\Psi(A\sigma^{-1}u^{1-H})I_{\mu<c} + I_{\mu\geqslant c} \quad (9)$$

where $I_{\mu<c}$ is 1 if $\mu < c$ and is zero otherwise (similarly one defines $I_{\mu\geqslant c}$).

Since in applications the parameters of model (4) are unknown, they must be estimated by the observed data. The goal of the present paper is to propose a Bayesian approach to such an estimation problem. More specifically, in Section 3, the priors for the unknown parameters are introduced, and updated on the basis of the sample data. Since the posteriors cannot be expressed in a closed form, a computational scheme based on MCMC is adopted. As a by-product, a Bayesian approach to the problem of testing for independence ($H = \frac{1}{2}$) against long-range dependence ($H > \frac{1}{2}$) is obtained. Finally, in Section 4 an application to real data is provided.

3. BAYESIAN ANALYSIS

According to guidelines provided in the previous sections, we now proceed to illustrate the Bayesian set-up for our statistical analysis.

Let n be the sample size and let $t_1, \ldots, t_n$ be fixed time instants. Correspondingly, n observations of the process $\{X(t)\colon t \geqslant 0\}$ are made and they are denoted by $X(t_1) = x_1, \ldots, X(t_n) = x_n$. For the sake of brevity, in the sequel, we will use the vector notation

$$\boldsymbol{x} = (x_1, \ldots, x_n)' \quad \boldsymbol{t} = (t_1, \ldots, t_n)'$$

Moreover, Ω_H denotes (apart from the constant σ^2) the covariance matrix of $(X(t_1), \ldots, X(t_n))$, whose (i, j)th element is

$$\omega_{i,j}(H) = \tfrac{1}{2}(t_i^{2H} + t_j^{2H} - |t_i - t_j|^{2H})$$

Since the process $(X(t);\ t \geqslant 0)$ is assumed to be a FBM, the likelihood function coincides with

$$f(H, \mu, \sigma^2; \boldsymbol{x}) = \frac{|\Omega_H|^{-1/2}}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{x} - \mu\boldsymbol{t})'\Omega_H^{-1}(\boldsymbol{x} - \mu\boldsymbol{t})\right\}$$

with $|\Omega_H|$ denoting the determinant of matrix Ω_H. As far as prior specification for the vector of parameters $(H, \mu, \sigma^2) \in [\frac{1}{2}, 1) \times \mathbb{R} \times \mathbb{R}^+$ we set

$$\pi_0(\mathrm{d}H) = \varepsilon\delta_{1/2}(\mathrm{d}H) + (1 - \varepsilon)\pi_0^*(H)\boldsymbol{I}_{1/2<H<1}\,\mathrm{d}H$$

$$\pi_1(\mu|\sigma^2) = \frac{1}{\sigma\sqrt{2\pi w}} \exp\left(-\frac{\mu - \mu_c)^2}{2\sigma^2 w}\right)$$

$$\pi_2(\sigma^2) = \frac{\lambda^v}{\Gamma(v)}\left(\frac{1}{\sigma^2}\right)^{v+1} \exp\left(-\frac{\lambda}{\sigma^2}\right)$$

for some $\varepsilon \in [0,1]$, and $\delta_x(\cdot)$ is the Dirac function at x. The prior specification we are adopting deserves some further explanation. The prior π_0 for the Hurst parameter H is essentially motivated by the following important fact: the (FBM) input process does possess completely different characteristics according to the values of H. In particular, if $H = \frac{1}{2}$, it reduces to a standard Brownian motion, which is an independent increments process (with short-range dependence behaviour and strong Markov, as well). If $H \in (\frac{1}{2}, 1)$ then the (FBM) input process possesses increments with long-range dependence and it turns out to be non-Markov. The prior π_0 takes into account this basic fact, and allows to compare short- vs long-range dependence. The value of ε measures the degree of prior belief about short-range dependence of the original series. Moreover, the diffuse component π_0^* on $(\frac{1}{2},1)$ is taken to be a uniform distribution so to reflect lack of prior information concerning the strength of the long-range dependence behaviour, if present.

As far as the prior for μ and σ^2 are concerned, they depend upon the hyperparameters λ, v, w and μ_c. A discussion of their choice is postponed to Section 4, where sensitivity of posterior estimates is considered as well.

3.1. Estimation of H

Since Bayes' theorem can be applied in our case, the posterior distribution of (H, μ, σ^2), given the vector of observations $\boldsymbol{x} = (x_1, \ldots, x_n)'$, is

$$\pi(\mathrm{d}H, \mathrm{d}\mu, \mathrm{d}\sigma^2|\boldsymbol{x}) \propto \pi_0(\mathrm{d}H)\pi_1(\mu|\sigma^2)\pi_2(\sigma^2)f(H, \mu, \sigma^2; \boldsymbol{x})\,\mathrm{d}\mu\,\mathrm{d}\sigma^2$$

where $\propto$ means that equality holds true up to a normalizing constant. In order to obtain a posterior estimate of the self-similarity parameter H, we determine its posterior distribution by integrating out μ and σ^2 in the joint posterior distribution, so that one has

$$\pi(\mathrm{d}H|\boldsymbol{x}) \propto \left\{ \frac{\varepsilon|\Omega_{1/2}|^{-1/2}}{\zeta_{1/2}^{1/2}M_{1/2}^{v+n/2}}\delta_{1/2}(\mathrm{d}H) + \frac{(1-\varepsilon)|\Omega_H|^{-1/2}}{\zeta_H^{1/2}M_H^{v+n/2}}\pi_0^*(H)\,\mathrm{d}H \right\}$$

where

$$\zeta_H := \boldsymbol{t}'\Omega_H^{-1}\boldsymbol{t} + \frac{1}{w}, \qquad \xi_H := \boldsymbol{t}'\Omega_H^{-1}\boldsymbol{x} + \frac{\mu_c}{w}, \qquad M_H := \lambda + \frac{\boldsymbol{x}'\Omega_H^{-1}\boldsymbol{x}}{2} + \frac{\mu_c^2}{2w} - \frac{\xi_H^2}{2\zeta_H}$$

are, for any H in $[\frac{1}{2}, 1)$, computable. On the contrary, the normalizing constant

$$\begin{aligned} k^*(\boldsymbol{x}) &= \varepsilon\frac{|\Omega_{1/2}|^{-1/2}}{\zeta_{1/2}^{1/2}M_{1/2}^{v+n/2}} + (1-\varepsilon)\int_{(1/2,1)}\frac{|\Omega_H|^{-1/2}}{\zeta_H^{1/2}M_H^{v+n/2}}\pi_0^*(H)\,\mathrm{d}H \\ &=: \varepsilon k_d(\boldsymbol{x}) + (1-\varepsilon)k_c(\boldsymbol{x}) \end{aligned}$$

has to be approximated. With the posterior distribution $\pi(\mathrm{d}H|x)$ at hand, one can obtain the posterior mean for H

$$E(H|\boldsymbol{x}) = (k*(\boldsymbol{x}))^{-1}\left\{ \frac{\varepsilon}{2}\frac{|\Omega_{1/2}|^{-1/2}}{\zeta_{1/2}^{1/2}M_{1/2}^{v+n/2}} + (1-\varepsilon)\int_{(1/2,1)}\frac{H|\Omega_H|^{-1/2}}{\zeta_H^{1/2}M_H^{v+n/2}}\pi_0^*(H)\,\mathrm{d}H \right\} \tag{10}$$

Since the integral appearing in the right-hand side of (10) cannot be exactly evaluated, we approximated it numerically. Details on the approximation procedures for $k^*(x)$ and $E(H|x)$ are illustrated in Appendix A.1.

3.2. Estimation of loss probability

As far as the problem of evaluating posterior loss probability, as a measure of performance of the system, is concerned, using formula (9) and the dominated convergence theorem, we have that

$$\begin{aligned} &P\left(\sup_{t\geqslant 0}(X(t) - ct) > u|\boldsymbol{x}\right) \\ &\quad\sim \int_{[1/2,1)\times\mathbb{R}\times\mathbb{R}^+} q_u(H, \mu, \sigma^2)\pi(\mathrm{d}H, \mathrm{d}\mu, \mathrm{d}\sigma^2|\boldsymbol{x})\boldsymbol{I}_{(\mu<c)} + P(\mu\geqslant c|\boldsymbol{x}) \end{aligned}$$

holds true for large values of u. Hence, one has

$$\begin{aligned}\int_{[1/2,1)\times\mathbb{R}\times\mathbb{R}^+} & q_u(H,\mu,\sigma^2)1_{(\mu<c)}\pi(\mathrm{d}H,\mathrm{d}\mu,\mathrm{d}\sigma^2|x)\\ &\leqslant \varepsilon\frac{k_d(\boldsymbol{x})}{k^*(\boldsymbol{x})}\int_{\mathbb{R}\times\mathbb{R}^+}\bar{q}_u\left(\tfrac{1}{2},\mu,\sigma^2\right)\pi^*\left(\tfrac{1}{2},\mu,\sigma^2|\boldsymbol{x}\right)\boldsymbol{I}_{(\mu<c)}\,\mathrm{d}\mu\,\mathrm{d}\sigma^2\\ &+(1-\varepsilon)\frac{k_c(\boldsymbol{x})}{k*(\boldsymbol{x})}\int_{(1/2,1)\times\mathbb{R}\times\mathbb{R}^+}\bar{q}_u(H,\mu,\sigma^2)\pi^{**}(H,\mu,\sigma^2|\boldsymbol{x})\boldsymbol{I}_{(\mu<c)}\,\mathrm{d}H,\mathrm{d}\mu,\mathrm{d}\sigma^2\end{aligned}\tag{11}$$

where $\bar{q}_u$ is an upper bound for q_u, obtained by substituting H_{2H} with 3.1. Moreover, $\pi^*(\cdot|x)$ and $\pi^{**}(\cdot|x)$ are probability distributions on $\mathbb{R}\times\mathbb{R}^+$ and on $(\frac{1}{2},1)\times\mathbb{R}\times\mathbb{R}^+$, respectively. One easily checks that the determination of an upper bound for the loss probability requires the numerical evaluation of $k_c(x)$ and of the two integrals appearing in (11) above. We implemented an MCMC algorithm whose features are fully described in the appendix.

However, the algorithm we have resorted to might be computationally cumbersome in some cases, because of the presence of the term $\Psi(A\sigma^{-1}u^{1-H})$ in (9). Considerable simplifications are obtained by virtue of the following well-known inequality for the Mills' ratio of the Gaussian distribution

$$\Psi(A\sigma^{-1}u^{1-H})<\frac{\sigma}{\sqrt{2\pi}A\sigma^{1-H}}\exp\left\{-\frac{1}{2\sigma^2}A^2u^{2-2H}\right\}\tag{12}$$

see, e.g. Reference [20, p. 49]. Using (12), we obtain

$$q_u(H,\mu,\sigma^2)<\frac{3.1(c-\mu)^{1-2H}(1-H)^{(5/2)-2H-(1/H)}u^{(H-1)(2H-1)/H}}{2^{1/(2H)}H^{3/2-2H}\sigma^{1/H-2}}\exp\left\{-\frac{A^2u^{2-2H}}{2\sigma^2}\right\}$$

and integration w.r.t. σ^2 gives, for an appropriate function $\bar{q}_u^*(H,\mu;\boldsymbol{x})$.

$$\begin{aligned}\int_{[1/2,1)\times\mathbb{R}\times\mathbb{R}^+} & q_u(H,\mu,\sigma^2)\pi(\mathrm{d}H,\mathrm{d}\mu,\mathrm{d}\sigma^2|\boldsymbol{x})\\ &<\int_{[1/2,1)\times\mathbb{R}}\bar{q}_u^*(H,\mu;\boldsymbol{x})\pi(\mathrm{d}H|\boldsymbol{x})\pi(\mu|H,\boldsymbol{x})\,\mathrm{d}\mu\end{aligned}\tag{13}$$

where

$$\pi(\mu|H,\boldsymbol{x})=\frac{\Gamma((n+2v+1)/2)}{\sqrt{\pi}\,\Gamma((n+2v)/2)}\sqrt{\frac{\zeta_H}{2M_H}}\left\{1+\frac{\zeta_H}{2M_H}\left(\mu-\frac{\xi_H}{\zeta_H}\right)^2\right\}^{-(n+2v+1)/2}$$

In order to provide an approximation of the integral in the right-hand side of (13), we implement both a Monte Carlo i.i.d. sampling and a Metropolis–Hastings scheme as described in appendix A.2.

3.3. *A model comparison problem: short- vs long-range dependence*

An important problem is to study the essential features of traffic, and in particular the presence/absence of self-similarity. Apropos of this we mention that there is a great debate in the literature in order to assess the characteristics of traffic in telecommunication systems, see the key paper by Willinger *et al.* [11] and the references in Reference [13]. See also Reference [15] for further important remarks and bibliographic references. The classical approach [11,12] consists of constructing an asymptotic confidence interval for the Hurst parameter H, and to check whether it contains 0.5, which implies short-range dependence, or not, this latter case implying long-range dependence. Two points have to be stressed. First of all, in References [11,12] the (frequentist) analysis involves Ethernet traffic packets; no analysis is made for ATM traffic, neither frequentist nor Bayesian. In our knowledge this is the first paper where a Bayesian analysis for ATM traffic data is carried out. As already mentioned in Section 2, the value of H does have a great influence on the performance of the system: the greater H, the worse the performance in terms of loss probability.

The problem of identifying the traffic data characteristics can be formally written down as an hypothesis problem

$$H_0 : H = \tfrac{1}{2}, \quad \text{vs} \quad H_1 : H > \tfrac{1}{2}$$

On the basis of results in previous section, a Bayesian test can be easily performed. In principle, a Bayesian test is based on the probability ratio

$$\frac{P(H = \frac{1}{2}\,|\,\boldsymbol{x})}{P(H \in (\frac{1}{2}, 1)|\boldsymbol{x})} \tag{14}$$

It is apparent from the exposition in Section 3 that the probabilities in (14) cannot be computed analytically. Using the same notation as in Section 3, the numerator of (14) is approximated by

$$\frac{\varepsilon|\Omega_{1/2}|^{-1/2}}{\hat{k}^*(\boldsymbol{x})\zeta_{1/2}^{1/2}M_{1/2}^{\nu+n/2}}$$

Hence ratio (14) is approximated by

$$\frac{\varepsilon|\Omega_{1/2}|^{-1/2}}{\hat{k}^*(\boldsymbol{x})\zeta_{1/2}^{1/2}M_{1/2}^{\nu+n/2} - \varepsilon|\Omega_{1/2}|^{-1/2}}$$

4. APPLICATION TO REAL DATA

We consider data from the experimental European ATM network [21], a project jointly developed by the leading telecommunications company in the European Union. Engineering aspects of the measurement problem are thoroughly described in Reference [22].

Data streams are produced by superimposing the traffic generated by three different kinds of applications: videoconference, teleteaching and transportation of routing information between IP network routers. All these applications use IP packets over ATM, so that the overflow probability must be estimated on the basis of the available data. The measurement process produces the number of cells arriving in a time slot (1/80 000 s) at an ATM multiplexer, composed by an ATM link and a buffer to store cells not immediately transmitted by the link.

The data we have used come from measurements taken for three different kind of applications: videoconference, teleteaching, and transportation of routing information. In order to study the effects of simultaneous transmissions from several different sources, as described in Section 2 (see (1)), the stream corresponding to every application has been split into substreams (1 s length), and only the data from the first quarter in each substream have been considered. Data from different quarters can be considered approximately independent, each of them coming from a different 'virtual' source. They can be superimposed as if they were coming from sources simultaneously connected to the same ATM multiplexer. We consider three different traffic scenarios: light, medium and heavy traffic. In the first case (light traffic) we have superimposed 10 'virtual sources' for each kind of applications. We have obtained a medium traffic situation by considering 30 teleteaching sources and 20 sources each for both videoconference and transportation of routing information. Finally, the heavy traffic scenario is obtained by superimposing 30 teleteaching sources and 34 sources each from both videoconference and transportation of routing information.

Data from the three scenarios are depicted in Figures 1(a)–(c). Arrivals follow a typical pattern in telecommunications, already observed in non-ATM cases; see References [11,12]. In fact, the patterns are far from being generated by a process with independent increments (i.e. with $H=0.5$). On the opposite, Figures 1(a)–(c) shows the presence of self-similarity in the arrival processes, i.e. long-range dependence in the corresponding increments ($H > 0.5$). As already mentioned in Section 2, FBM is a natural tool for modelling purposes.

Formulas in Section 3 require dealing with matrices whose dimension is given by the data stream length, in our case more than 20 000. Computational burdens have lead us to reduce the size by grouping the data considering the number of cells arriving in 300 consecutive time slots. Therefore, matrices from grouped data have been inverted by FORTRAN routines and their determinants have been computed, as well.

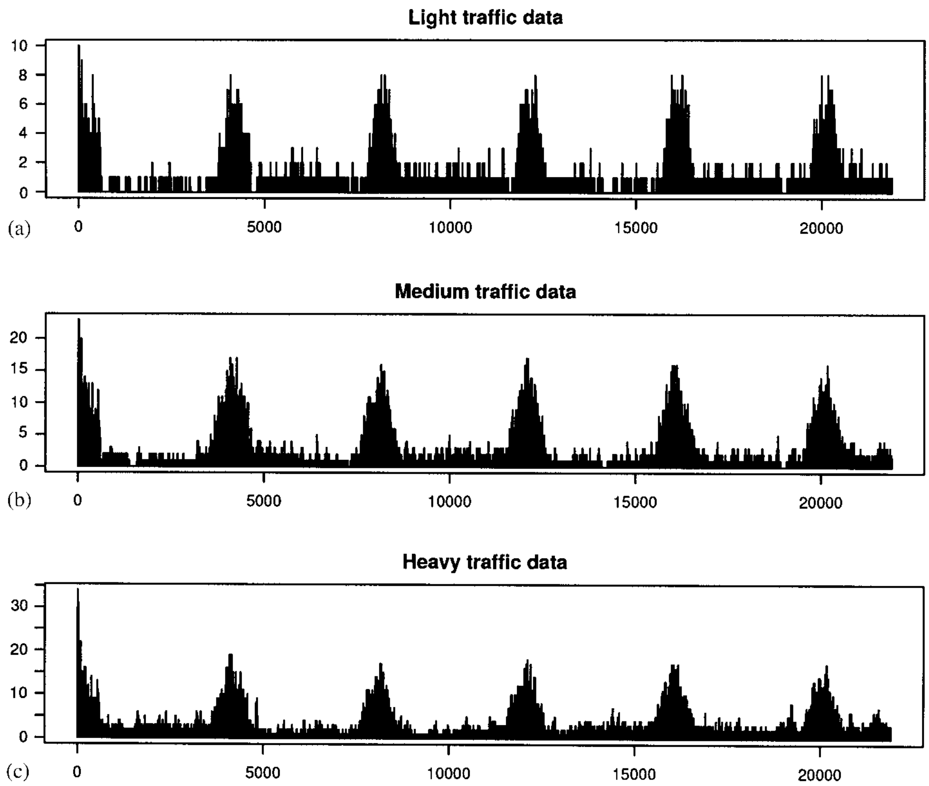

Figure 1. Number of cell arrivals under (a) light, (b) medium and (c) heavy traffic.

In relationship (3) we have taken $u = 1000$, which is a rather high buffer size, whereas the choice $c = 300$ follows from the link speed. Moreover, in the prior distribution of μ we set $w = 0.01$, in order to reduce the probability of sampling negative values for μ.

Our main goals are the detection of long-range dependence and the estimation of the overflow probability $Q(u)$. The former goal has been achieved by both estimating the parameter H and computing the posterior probability ratio for the problem $H = \frac{1}{2}$ vs $H > \frac{1}{2}$. The main finding is the detection of long-range dependence even when considering priors heavily concentrated around $H = \frac{1}{2}$. The values of ratio (14), for different values of prior hyperparameters, are shown in Tables I and II as well as the Bayes estimates of $Q(u)$. Table III shows the same quantities when the r.h.s. of inequality (12) is considered. It is worth mentioning that plots of estimated $Q(u)$'s vs number of iterations show a quick convergence.

It appears that even when the prior probability of $H = 0.5$ is very high ($\varepsilon = 0.999$), ratio (14) takes small values. By the way, the smaller ε, the smaller (14). Hence, our first conclusion is that the data show the presence of strongly correlated increments in the input processes corresponding to the three scenarios considered. Bayes' estimates of H (Table IV) are generally far from 0.5.

The value $\varepsilon = 1$ has been considered in order to study the effect of neglecting long-range dependence. Such an effect is particularly relevant in case of 'heavy' traffic (Tables I and II), which is the most important for applications. As expected, a value $\varepsilon = 1$ could produce a severe underestimation of the overflow probability. The use of the bound in (13) is less expensive from a computational point of view, but produces less accurate results. Compare Tables I and II with Table III. As far as the sensitivity of the Bayes' estimates of the overflow probabilities (with

Table I. Upper bound and probability ratio with $v = 500$, $\lambda = 10$, $\mu = 1000$.

Traffic	ε	Upper	Ratio
Heavy	1	0.0	$+\infty$
Heavy	0.999	0.267×10^{-9}	0.159×10^{-19}
Medium	1	0.0	$+\infty$
Medium	0.999	0.0	0.222×10^{-21}
Light	1	0.0	$+\infty$
Light	0.999	0.0	0.222×10^{-21}

The column labelled 'Upper' features posterior estimates of the upper bound of the overflow probability. The column labelled 'Ratio' provides estimates of the posterior probability ratio for the Bayesian test in Section 4.

Table II. Upper bound and probability ratio with $v = 300$, $\lambda = 10$, $\mu = 1000$.

Traffic	ε	Upper	Ratio
Heavy	1	0.49×10^{-16}	$+\infty$
Heavy	0.999	0.907×10^{-7}	0.687×10^{-12}
Medium	1	0.0	$+\infty$
Medium	0.999	0.996×10^{-16}	0.198×10^{-12}
Light	1	0.0	$+\infty$
Light	0.999	0.0	0.581×10^{-19}

The column labelled 'Upper' features posterior estimates of the upper bound of the overflow probability. The column labelled 'Ratio' provides estimates of the posterior probability ratio for the Bayesian test in Section 4.

Table III. Upper bound and probability ratio using Mills' ratio, with $\lambda = 10$.

Traffic	ε	v	Upper	Ratio
Heavy	0.999	500	0.556×10^{-7}	0.16×10^{-20}
Heavy	0.999	300	0.498×10^{-4}	0.66×10^{-12}
Heavy	1	300	0.0	$+\infty$
Medium	0.999	300	0.232×10^{-13}	0.193×10^{-12}
Light	0.999	300	0.423×10^{-41}	0.57×10^{-19}

The column labelled 'Upper' features posterior estimates of the upper bound of the overflow probability. The column labelled 'Ratio' provides estimates of the posterior probability ratio for the Bayesian test in Section 4.

Table IV. Posterior estimates of H.

	$v = 300$, $\lambda = 10$			$v = 3$, $\lambda = 10$		
	HT	MT	LT	HT	MT	LT
Monte Carlo	0.665	0.666	0.726	0.867	0.865	0.894
MCMC	0.694	0.688	0.756	0.821	0.818	0.832
MC error	0.00164	0.00165	0.00159	0.0015	0.0015	0.0017

The abbreviation HT stands for 'High Traffic', MT for 'Medium Traffic' and LT for 'Light Traffic'.

respect to the choice of the hyperparameters) is concerned, we may note that it is moderate although not negligible.

As far as posterior estimates of H are concerned, our results are summarized in Table IV, where all three different frameworks of heavy traffic (HT), medium traffic (MT) and light traffic (LT) are considered. Two different procedures have been employed. The first one relies upon the classical Monte Carlo procedure as illustrated in Section 3.1. The second one resorts to a Metropolis–Hastings sampling scheme for drawing from the posterior $\pi(\mathrm{d}H|\boldsymbol{x})$ and the resulting sample is used for estimating H. The estimates have been obtained after 12 000 runs and with a burn-in of 10 000 iterations. Diagnostic tests performed with the BOA package (see Reference [23]) have provided strong evidence of convergence of the estimation procedure. The reason for considering an MCMC scheme in this setting is two-fold. On one hand, it is desirable to sketch some comparison, both in terms of computational time and in terms of numerical outcomes, with the classical Monte Carlo procedure. From a computational point of view, the Monte Carlo method is much faster, since it requires on average 15 min with a HP machine (processor PA8000, 180 MHz) to be completed, whereas the MCMC sampler has been running for 95 min. From a numerical point of view, the estimates are not significantly different. On the other hand, having performed an MCMC algorithm one can use the MCMC output in order to get a kernel density estimate of the posterior distribution of H. This can give some insight on the dispersion of H around its posterior mean. In Figure 2, we provide graphs both of the histogram of the MCMC sample and the kernel density estimate of the posterior distribution of H.

The results obtained are fairly insensitive to different choices of the parameters of the prior for μ. The situation is different as far as the prior of σ^2 is concerned. From Table IV, it is argued that different values of v could have a rather strong influence on the estimation of H. However, as it appears from Tables I and II, such a negative effect is mitigated when one considers the

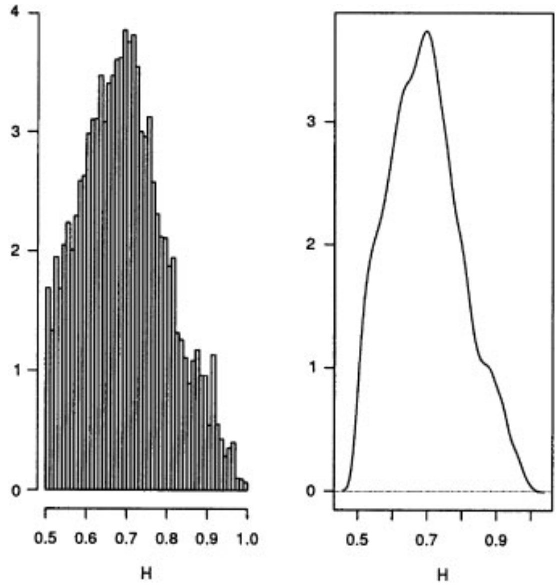

Figure 2. Histogram (on the left) and kernel density estimate of H (on the right) obtained with $v = 10$, $\lambda = 300$ and $\varepsilon = 0.999$. The posterior estimate is, in this case, $E(H|\boldsymbol{x}) = \hat{H} \approx 0.694$. The estimates are obtained basing on the 'heavy traffic' data.

Table V. Sensitivity of posterior estimates of H (with heavy traffic data).

	$\mathrm{Var}(\sigma^2) = 0.001$		$\mathrm{Var}(\sigma^2) = 1$	
$E(\sigma^2)$	E(H|x)	Acceptance ratio (%)	E(H|x)	Acceptance ratio (%)
0.5	0.6831	50.14	0.8251	58.11
1	0.6828	50.06	0.8239	57.9
10	0.6827	50.04	0.6931	58.29
100	0.6828	50.01	0.6829	57.43

overflow probability (which is the *real* goal of our analysis). In fact, large variations of v produce only moderate variations of the Bayes' estimates of the overflow probability. Here we provide some tables with estimates of H corresponding to different values of $E(\sigma^2)$ and $\mathrm{Var}(\sigma^2)$, i.e. to different choices of λ and v. See Table V. One can notice that the posterior estimates of H are more sensitive to changes in $E(\sigma^2)$ when $\mathrm{Var}(\sigma^2)$ than in the case in which $\mathrm{Var}(\sigma^2)$ is very low.

4.1. Discussion on the choice of the hyperparameters

The Bayes estimates of the overflow probability exhibit some sensitivity w.r.t. the hyperparameters w, λ, v. For this reason, it is of interest to discuss their choice. In telecommunications, prior information is frequently available in the form of *trial samples*, i.e. small samples of traffic measurements.

Assume that the available prior data consist of $k \geqslant 1$ independent samples. Each sample is obtained by observing the system under consideration for a (short) period of time. The whole observation period is split into m time intervals of length $\Delta > 0$. Let $Y_{i,j}$ be the amount of traffic entering the system in the ith time interval of the jth sample ($i = 1, \ldots, m$, $j = 1, \ldots, k$). From our previous assumptions, conditionally on μ, σ, H, the random vectors $(Y_{i,j}; i = 1, \ldots, m)$ have independent multinormal distributions, with $E[Y_{i,j}|\mu,\sigma,H] = \Delta\mu$, $\mathrm{Var}[Y_{i,j}|\mu,\sigma,H] = \Delta^2\sigma^2$, $\mathrm{Cov}[Y_{i,j},\ Y_{i+l,j}\ |\mu,\sigma,H] = \Delta^2\rho(l)$.

As a prior for H, it is reasonable to assume a mixture of a Dirac $\delta_{1/2}$ and a uniform distribution. As seen in the example, the posterior of H is robust w.r.t. the weight ε. As far as the choice of μ_c, w, λ, v is concerned, the basic idea consists in matching them with 'empirical quantities' evaluated on the basis of $Y_{i,j}$s. First of all, let $z > 0$, and

$$\tau(z) := E[z^H] = \varepsilon z^{1/2} + 2(1-\varepsilon)\frac{z - z^{1/2}}{\log z}$$

It is immediate to see that the equalities

$$E[\mu] = \mu_c, \quad E[\sigma^2] = \frac{\lambda}{v-1}, \quad E[\sigma^3] = \frac{\lambda^{3/2}}{\Gamma(v)}\Gamma\left(v - \frac{3}{2}\right)$$

$$\mathrm{Var}(\mu) = E[\mathrm{Var}(\mu|\sigma)] + \mathrm{Var}(E[\mu|\sigma]) = E[w\sigma^2] = w\frac{\lambda}{v-1}$$

$$E[\rho(l)] = \tfrac{1}{2}(\tau((l+1)^2) - 2\tau(l^2) + \tau((l-1)^2))$$

hold true, provided that $v > \frac{3}{2}$. Consider now the hth sample moments

$$\bar{Y}_{h,j} = m^{-1}\sum_{i=1}^{m} Y_{i,j}^h, \quad \bar{Y}_{h..} = k^{-1}\sum_{j=1}^{k}\bar{Y}_{h,j}, \quad h = 1, 2, 3$$

Their expected values are equal to

$$a_1 = E[\bar{Y}_{1,j}] = \Delta\mu_c$$

$$a_2 = E[\bar{Y}_{2,j}] = E|Y_{i,j}^2] = \Delta^2\left(\frac{\lambda}{v-1}(1+w) + \mu_c^2\right)$$

$$a_3 = E[\bar{Y}_{3,j}] = E|Y_{i,j}^3] = \Delta^2\left(3\frac{\lambda}{v-1}(1+w)\mu_c + \mu_c^3\right)$$

$$\begin{aligned} a_4 &= \mathrm{Var}(Y_{1,j}) = E[\mathrm{Var}(\bar{Y}_{1,j}|\mu,\sigma,H)] + \mathrm{Var}(\mu) \\ &= \Delta^2\left((m^{-1} + w)E|\sigma^2] + \frac{2}{m^2}\sum_{l=1}^{m-1}(m-l)E[\sigma^2\rho(l)]\right) \\ &= \Delta^2\frac{\lambda}{v-1}\left\{m^{-1} + w + \frac{1}{m^2}\sum_{l=1}^{m-1}(m-l)(\tau((l+1)^2) - 2\tau(l^2) + \tau((l-1)^2))\right\} \end{aligned}$$

respectively. A possible (and simple, as well) criterion to set the prior hyperparameters consists in choosing μ_c, w, λ, v in such a way that the relationships

$$a_1 = \bar{Y}_{1..}, \quad a_2 = \bar{Y}_{2..}, \quad a_3 = \bar{Y}_{3..}, \quad a_4 = \frac{1}{k-1}\sum_{j=1}^{k}(\bar{Y}_{1\cdot j} - \bar{Y}_{1..})^2$$

hold true.

5. CONCLUSIONS

In this paper, we have proposed a new, Bayesian approach to estimate the overflow probability in ATM networks, when FBM is used to model the traffic. We have analysed short- and long-range dependence using Telecom Italia data and we have discussed important issues in Bayesian analysis, like the choice of the hyperparameters and the sensitivity of the inference with respect to changes in their values. The choice of the prior distributions has been motivated by their flexibility and relative ease in their use. More general classes could have been used, but at the cost of making the computational algorithm even more cumbersome. The choice of the FBM is justified by its (relatively) tractable mathematical structure, although the request of Gaussianity of traffic data could be attenuated. For this purpose, other self-similar processes could be considered, but their use would be a very challenging task.

APPENDIX A

A description of the main computational issues associated with the estimation procedure set forth in Section 3 will be now provided.

A.1. Estimation of H

In order to provide posterior estimates of the Hurst parameter H, a simple Monte Carlo procedure is adopted. Such a choice is suggested by the expression appearing on the right-hand side of (10). A sample of N i.i.d. observations $H_1, \ldots, H_N$ from $\pi_0{}^*(H)$ can be generated. Such a sample is used to approximate the normalizing constant $k^*(\boldsymbol{x})$ by means of the empirical mean

$$\hat{k}^*(\boldsymbol{x}) = \frac{\varepsilon|\Omega_{1/2}|^{-1/2}}{\zeta_{1/2}^{1/2} M_{1/2}^{v+n/2}} + \frac{1-\varepsilon}{N}\sum_{i=1}^{N}\frac{|\Omega_{H_i}|^{-1/2}}{\zeta_{H_i}^{1/2} M_{H_i}^{v+n/2}}$$

so that $E(H|x)$ can be approximated by a ratio of empirical means, namely

$$E(H|\boldsymbol{x}) \approx (\hat{k}^*(\boldsymbol{x}))^{-1}\left\{\frac{\varepsilon}{2}\frac{|\Omega_{1/2}|^{-1/2}}{\zeta_{1/2}^{1/2} M_{1/2}^{v+n/2}} + \frac{(1-\varepsilon)}{N}\sum_{i=1}^{N}\frac{H_i|\Omega_{H_i}|^{-1/2}}{\zeta_{H_i}^{1/2} M_{H_i}^{v+n/2}}\right\}$$

A.2. Estimation of loss probability

An algorithm providing the desired approximations of the two integrals appearing in the right-hand side of (11) works as follows. If, for any H in $[\frac{1}{2}, 1)$,

$$\pi(\sigma^2|H,\boldsymbol{x}) = \frac{M_H^{v+n/2}}{\Gamma(v+n/2)}\left(\frac{1}{\sigma^2}\right)^{v+n/2+1}\exp\left\{-\frac{M_H}{\sigma^2}\right\}$$

and

$$\pi(\mu|H,\sigma^2,\boldsymbol{x}) = \frac{\zeta_H^{1/2}}{\sigma\sqrt{\pi}}\exp\left\{-\frac{\zeta_H}{2\sigma^2}\left(\mu-\frac{\xi_H}{2\zeta_H}\right)^2\right\}$$

one has $\pi^*(\frac{1}{2},\mu,\sigma^2|\boldsymbol{x}) = \pi(\mu|\frac{1}{2},\sigma^2,\boldsymbol{x})\pi(\sigma^2|\frac{1}{2},\boldsymbol{x})$. It follows that the first integral in the right-hand side of (11), corresponding to the case in which $H=\frac{1}{2}$, can be easily handled by generating an i.i.d. sample (σ_i^2,μ_i), $i=1,\ldots,N$, with $\sigma_i{}^2$ and μ_i drawn from $\pi(\sigma^2|\frac{1}{2},\boldsymbol{x})$ and from $\pi(\mu|\frac{1}{2},\sigma_i^2,\boldsymbol{x})$, respectively. Hence

$$\int_{\mathbb{R}\times\mathbb{R}^+}\bar{q}_u\left(\tfrac{1}{2},\mu,\sigma^2\right)\pi^*\left(\tfrac{1}{2},\mu,\sigma^2|\boldsymbol{x}\right)\boldsymbol{I}_{(\mu<c)}\,\mathrm{d}\mu\,\mathrm{d}\sigma^2 \approx \frac{1}{N}\sum_{i=1}^{N}\bar{q}_u\left(\tfrac{1}{2},\mu_i,\sigma_i^2\right)\boldsymbol{I}_{(\mu_i<c)}$$

As far as the second integral in the right-hand side of (11) is concerned, note that $\pi**(dH,\mu,\sigma^2|\boldsymbol{x}) = \pi(H|\boldsymbol{x})\pi(\sigma^2|H,\boldsymbol{x})\pi(\mu|H,\sigma^2,\boldsymbol{x})\boldsymbol{I}_{1/2<H<1}$, where

$$\pi(H|\boldsymbol{x}) \propto \frac{|\Omega_H|^{-1/2}}{\zeta_H^{1/2}M_H^{v+n/2}}\pi_0*(H)$$

A Metropolis–Hastings algorithm applies in this case, since it is not possible to sample directly from $\pi(H|x)$. The proposal we employ is

$$p(H_{i+1,}\sigma_{i+1}^2,\mu_{i+1}|H_i,\sigma_i^2,\mu_i) = \eta(H_{i+1})\pi(\sigma_{i+1}^2|H_{i+1},\boldsymbol{x})\pi(\mu_{i-1}|\sigma_{i+1}^2,H_{i+1},\boldsymbol{x})$$

$\eta(H) \propto (1-2H)^{\gamma-1}(1-H)^{\delta-1}$ being a probability density function on $(\frac{1}{2},1)$. Therefore, the adopted scheme corresponds to an independence sampler with acceptance ratio given by

$$\alpha((H_i,\mu_i,\sigma_i),(H_{i+1},\mu_{i+1},\sigma_{i+1})) = \min\left\{1,\frac{\eta(H_{i+1})\pi(H_i|x)}{\eta(H_i)\pi(H_{i+1}|\boldsymbol{x})}\right\}$$

The same MCMC output is used to estimate $k_c(\boldsymbol{x})$ and, then, $k^*(\boldsymbol{x})$.

Let us now move on to the problem of determining the upper bound in (13), which follows from inequality (12) on Mills' ratio for the Gaussian distribution. We still consider separately the cases $H=\frac{1}{2}$ and $H\in(\frac{1}{2},1)$. When $H=\frac{1}{2}$, a simple Monte Carlo integration can be done, by sampling i.i.d. μ_i's from $\pi(\mu|\frac{1}{2},\boldsymbol{x})$. If $H\in(\frac{1}{2},1)$, an independence sampler is employed again. In fact, we set

$$p(H_{i+1,}\mu_{i+1}|H_i,\mu_i) := \eta(H_{i+1})\pi(\mu_{i+1}|H_{i+1},\boldsymbol{x})$$

as the proposal distribution, with $\eta(H)$ as above. Acceptance ratio is

$$\alpha((H_i,\mu_i),(H_{i+1},\mu_{i+1})) = \min\left\{1,\frac{\eta(H_{i+1})\pi(H_i|\boldsymbol{x})}{\eta(H_i)\pi(H_{i+1}|\boldsymbol{x})}\right\}$$

ACKNOWLEDGEMENTS

Thanks are due to two anonymous referees, who carefully read this paper, and made very interesting and stimulating comments. The authors are also grateful to Giovanni Petris, Antonio Pievatolo and Mike Wiper for their valuable comments. The research was performed when the second author was affiliated with CNR-IMATI.

REFERENCES

1. Giroux N, Ganti S. *Quality of Service in ATM Networks*. Prentice-Hall: London, 1999.
2. De Pricker M. *Asynchronous Transfer Mode: Solution for Broadband ISDN*. Ellis Horwood: New York, 1993.
3. IETF RFC 2225. Classic IP and ARP over ATM. *Internet Engineering Task Force*, 1998.
4. IETF RFC 3035. MLPS using LPD and ATM VC switching. *Internet Engineering Task Force*, 2001.
5. Roberts R, Mocci U, Virtamo J (eds). *Broadband Network Teletraffic*. Lectures Notes in Computer Science, vol. 1155. Springer: Berlin, 1996.
6. ITU-T Recommendation I.356. ATM layer cell transfer performance. *International Telecommunications Union—Telecommunication*, 1996.
7. Bruneel H, Kim BG. *Discrete-Time Models for Communication Systems including ATM*. Kluwer: Boston, 1993.
8. Beneš VE. *General Stochastic Processes in the Theory of Queues*. Addison-Wesley: Reading, MA, 1963.
9. Norros I. A storage model with self-similar input. *Queueing Systems Theory and Applications* 1994; **16**: 387–396.
10. Norros I. On the use of fractional Brownian motion in the theory of connectionless networks. *IEEE Journal on Selected Areas in Communications* 1995; **13**:953–962.
11. Willinger W, Taqqu MS, Leland WE, Wilson DV. Self-similarity in high-speed packet traffic: analysis and modelling of Ethernet traffic measurements. *Statistical Science* 1995; **10**:67–85.
12. Willinger W, Taqqu MS, Sherman R, Wilson DV. Self-similarity through high-variability: statistical analysis of Ethernet *LAN* traffic at the source level. In *Proceedings of ACM SIGCOMM 95*. Cambridge, ACM: 1995; 100–113.
13. Willinger W, Taqqu MS, Erramilli A. A bibliographic guide to self-similar traffic and performance modeling for modern high-speed networks. In *Stochastic Network*, Kelly FP, Zachery S, Ziedins I (eds). Oxford University Press: Oxford, 1997.
14. Beran J. *Statistics for Long-Memory Processes*. Chapman & Hall: London, 1994.
15. Resnick SI. Heavy tail modeling, teletraffic data. *The Annals of Statistics* 1997; **25**:1805–1896.
16. Tsybakov B, Georganas ND. On self-similar traffic in ATM queues: definitions, overflow probability bound and cell delay distribution. *IEEE Transactions on Networking* 1997; **5**:397–409.
17. Hüsler J, Piterbarg V. Extremes of a certain class of Gaussian processes. *Stochastic Processes and their Applications* 1999; **83**: 257–271.
18. Shao QM. Bounds and estimators of a basic constant in extreme value theory of Gaussian processes. *Statistica Sinica* 1996; **6**: 245–257.
19. Hüsler J. Extremes of a Gaussian process and the constant H_α. *Extremes* 1999; **1**:59–70.
20. Chow YS, Teicher H. *Probability Theory*. Springer: New York, 1995.
21. Parker, M. The European ATM Pilot. In *Proceedings of XV International Switching Symposium ISS'95*, Berlin, 1995.
22. Gnetti AS. Characterizations and measurements of teletraffic generated by IP protocol on ATM network. *M.S. Thesis*, Università di Roma 'La Sapienza' Facoltà di Ingegneria (in Italian).
23. Smith B. *Bayesian Output Analysis program: version 1.0.0*. Department of Biostatistics, University of Iowa, 2001, Available at http://www.public-health.uiowa/edu/boa.

