Parameter estimation of geometrically sampled fractional Brownian traffic

A. Vidács

High Speed Networks Labs
Dept. of Telecommunications and Telematics
Technical University of Budapest
Pázmány Péter sétány 1/D,
H-1117 Budapest, Hungary

J.T. Virtamo

Laboratory of Telecommunications Technology
Helsinki University of Technology
P.O.B. 3000, FIN-02015 HUT, Finland

Abstract—**The parameter estimation of a traffic model based on the fractional Brownian motion (fBm) is studied. The model has three parameters: mean rate m, variance parameter a and the Hurst parameter H. Explicit expressions for the maximum likelihood (ML) estimates $\hat{m}$ and $\hat{a}$ in terms of H are given, as well as the expression for the log-likelihood function from which the estimate $\hat{H}$ is obtained as the maximizing argument. A geometric sequence of sampling points, $t_i = \alpha^i$, is introduced, which fits neatly to the self-similar property of the process and also reduces the number of samples needed to cover several time scales. It is shown that by a proper 'descaling' the traffic process is stationary on this grid leading to a Toeplitz-type covariance matrix. Approximations for the inverted covariance matrix and its determinant are introduced. The accuracy of the estimations is studied by simulations. Comparisons with estimates obtained with linear sampling and with the wavelet-based A-V estimator show that the geometrical sampling indeed improves the accuracy of the estimate $\hat{H}$ with a given number of samples.**

I. INTRODUCTION

One of the simplest and most studied models for aggregated data traffic is the fractional Brownian motion (fBm) model [1], which is a model for truly self-similar Gaussian traffic. Though the model has its limitations and, in particular, breaks down at small time scales, it has gained popularity because of its simplicity. In its basic form the model contains only three parameters, the mean rate m, the variance parameter a and the Hurst parameter H describing the scaling behaviour of the traffic. A small number of traffic parameters is a very desirable feature from the point of view of the applicability of the model for traffic engineering purposes. The estimation of even a small number of parameters poses a problem for long range dependent traffic. Some early work [1] suggested that to obtain a reasonable accuracy a very large number of sample points may be required. The problem arises e.g. in the estimation of the Hurst parameter H. As H describes the scaling behaviour of the traffic variability, the sample points have to cover several time scales, i.e., the total time range must be several orders of magnitude greater than the finest time resolution in the measurement.

In this paper we show that by an appropriate choice of the sampling instants, the number of sampling points can be considerably reduced. In particular, we will introduce a grid of geometrically distributed sampling points $t_i = \alpha^{i-1}$, $i = 1, \ldots, n$ where α is some constant (smaller than one).

The idea here is that such a sampling grid covers several time scales with fewer points. The second point is that the geometrical grid, being "self-similar", fits well with the traffic process and gives rise to a simple structure in the covariance matrix.

Throughout this work we apply the maximum likelihood estimation (MLE) method [2]. MLE method has previously been applied to this problem by Deriche and Tewfik [3] and Ninness [4] using ordinary linear sampling. Explicit formulas for the estimators of m and a are given along with the log-likelihood function for determining the estimator for H. A major difficulty in this method is the calculation of the inverse and determinant of the covariance matrix appearing in the likelihood function. An approximate calculation is facilitated if the process is stationary whence the matrix is of Toeplitz type. For the original fBm process the increment process is stationary. We show that another stationary process is obtained from the fBm process by 'descaling' and changing the process index to logarithmic time, i.e., on the geometrical sampling grid the descaled process is stationary.

We compare the effectiveness of the ML estimator based on ordinary evenly spaced sampling grid with that obtained with a geometrical grid, and also with the wavelet based parameter estimation method proposed by Abry and Veitch in [5], [6], by simulations. The simulations indicate that in the estimation of H the geometrical grid gives an advantage. (In this study we will concentrate on the estimates of the Hurst parameter H only. For the parameters a and m, see [7].)

The rest of this paper is organized as follows. In section II we review the fractional Brownian motion traffic model with its three parameters. The general problem of the estimation of these parameters by the maximum likelihood method is considered in section III. The idea of geometrical sampling and the descaled process, along with an approximate form of the MLE, are introduced in section IV. For comparison, in section V we present the MLE method for the case of ordinary linear sampling and a brief overview of the wavelet based Abry-Veitch estimator. In section VI, we present results for estimating the Hurst parameter with the described methods from simulated realizations of the process. Section VII concludes the paper.

II. FRACTIONAL BROWNIAN TRAFFIC

A normalized *fractional Brownian motion* with Hurst parameter $H \in [0.5, 1)$, denoted by $Z(t)$, ($t \in \mathbb{R}$), is characterized by the following properties [8]:
1. $Z(t)$ has stationary increments;
2. $Z(0) = 0$, and $\mathrm{E}[Z(t)] = 0$ for all t;
3. $\mathrm{Var}[Z(t)] = \mathrm{E}[Z(t)^2] = |t|^{2H}$ for all t;
4. $Z(t)$ has continuous paths;
5. $Z(t)$ is a Gaussian process, i.e., all its finite-dimensional marginal distributions are Gaussian.

In the special case $H = 0.5$, $Z(t)$ is the standard Brownian motion. It follows from the above properties that $Z(t)$ is a self-similar process whose scaling behaviour is defined by the Hurst parameter H as follows

$$Z(\alpha t) \sim \alpha^H Z(t). \tag{1}$$

The covariance structure of the process is given by

$$\mathrm{Cov}\left[Z(t_1), Z(t_2)\right] = \frac{1}{2}\left\{t_1^{2H} + t_2^{2H} - |t_2 - t_1|^{2H}\right\}. \tag{2}$$

Furthermore, in the case $H > 0.5$ the strongly correlated stationary sequence $Z(n+1) - Z(n)$, the increment process of $Z(t)$ (often called *fractional Gaussian noise*), is ergodic [8].

Fractional Brownian motion is a popular model for long-range dependent traffic. Norros [8] has suggested the following model

$$X(t) = mt + \sqrt{a}Z(t), \tag{3}$$

where $X(t)$ represents the amount of traffic arrived in $(0, t)$. The model has three parameters, m, a and H with the following interpretations and intervals for allowed values: $m > 0$ is the mean input rate, $a > 0$ is a variance parameter, and $H \in [0.5, 1)$ is the self-similarity parameter of $Z(t)$.

III. EXACT GAUSSIAN MLE

Assume the traffic has been observed at n time instants forming the vector $\mathbf{t} = (t_1, \ldots, t_n)^{\mathrm{t}}$ where $(\cdot)^{\mathrm{t}}$ denotes the transpose, and let $\mathbf{X} = (X(t_1), \ldots, X(t_n))^{\mathrm{t}}$ be the vector of observed traffic values at these instants. Since $X(t)$ is Gaussian, the joint probability density function of $\mathbf{X}$ is

$$h(\mathbf{x}) = (2\pi)^{-\frac{n}{2}} |\mathbf{\Gamma}|^{-\frac{1}{2}} e^{-\frac{1}{2}(\mathbf{x}-\mathbf{m})^{\mathrm{t}} \mathbf{\Gamma}^{-1}(\mathbf{x}-\mathbf{m})}, \tag{4}$$

where $\mathbf{x} = (x_1, \ldots, x_n)^{\mathrm{t}} \in \mathbb{R}^n$, $\mathbf{m} = m\mathbf{t}$, and $|\mathbf{\Gamma}|$ is the determinant of the covariance matrix

$$\mathbf{\Gamma} = \mathrm{Cov}\left[\mathbf{X}, \mathbf{X}^{\mathrm{t}}\right] = \mathrm{E}\left[\mathbf{X}\mathbf{X}^{\mathrm{t}}\right] - \mathrm{E}[\mathbf{X}]\,\mathrm{E}\left[\mathbf{X}^{\mathrm{t}}\right]. \tag{5}$$

Since $\mathbf{\Gamma}$ is a simple linear function of a, i.e., $\mathbf{\Gamma}(a) = a\mathbf{\Gamma}_H$, where

$$\mathbf{\Gamma}_H = \mathrm{E}\left[\mathbf{Z}\mathbf{Z}^{\mathrm{t}}\right] = \left[\mathrm{Cov}\left[Z(t_i), Z(t_j)\right]\right]_{i,j=1,\ldots,n}, \tag{6}$$

the log-likelihood function is given by

$$\log h(\mathbf{X}; m, a, H) = -\frac{n}{2}\log 2\pi \tag{7}$$
$$-\frac{1}{2}\log a^n|\mathbf{\Gamma}_H| - \frac{1}{2a}(\mathbf{X} - m\mathbf{t})^{\mathrm{t}}\mathbf{\Gamma}_H^{-1}(\mathbf{X} - m\mathbf{t}).$$

The ML estimates $\hat{m}$ and $\hat{a}$ are given by taking the derivative of the log-likelihood function with respect to m and a respectively, and setting it to zero. By doing this, we get

$$\hat{m} = \frac{\mathbf{t}^{\mathrm{t}}\,\mathbf{\Gamma}^{-1}\mathbf{X}}{\mathbf{t}^{\mathrm{t}}\,\mathbf{\Gamma}^{-1}\,\mathbf{t}}, \tag{8}$$

$$\hat{a} = \frac{1}{n}\left[(\mathbf{X}^{\mathrm{t}}\,\mathbf{\Gamma}_H^{-1}\,\mathbf{X}) - \frac{(\mathbf{t}^{\mathrm{t}}\,\mathbf{\Gamma}_H^{-1}\,\mathbf{X})^2}{\mathbf{t}^{\mathrm{t}}\,\mathbf{\Gamma}_H^{-1}\,\mathbf{t}}\right].$$

Inserting the estimates from (8) into (7), essentially we have to minimize

$$L(\mathbf{X}; H) = \log|\mathbf{\Gamma}_H| \tag{9}$$
$$+ n\log\left[(\mathbf{X}^{\mathrm{t}}\,\mathbf{\Gamma}_H^{-1}\,\mathbf{X}) - \frac{(\mathbf{t}^{\mathrm{t}}\,\mathbf{\Gamma}_H^{-1}\,\mathbf{X})^2}{\mathbf{t}^{\mathrm{t}}\,\mathbf{\Gamma}_H^{-1}\,\mathbf{t}}\right].$$

The first term is a decreasing function of H, and the second term is an increasing function of H. The minimum is obtained for some value $\hat{H}$ which is the ML estimate. The minimization can be done by setting the derivative of $L(\mathbf{X}; H)$ with respect to H to zero. Using the notations $\mathbf{u} = (\mathbf{t}^{\mathrm{t}}\,\mathbf{\Gamma}_H^{-1}\,\mathbf{t})\mathbf{X}$ and $\mathbf{v} = (\mathbf{X}^{\mathrm{t}}\,\mathbf{\Gamma}_H^{-1}\,\mathbf{X})\mathbf{t}$, and the relationships

$$\frac{\partial}{\partial\theta}\mathbf{A}^{-1} = -\mathbf{A}^{-1}\left(\frac{\partial}{\partial\theta}\mathbf{A}\right)\mathbf{A}^{-1}, \tag{10}$$

$$\frac{\partial}{\partial\theta}\log|\mathbf{A}| = \mathrm{Tr}\left(\mathbf{A}^{-1}\frac{\partial}{\partial\theta}\mathbf{A}\right) \tag{11}$$

valid for any matrix $\mathbf{A}$ depending on a parameter θ [9], $\hat{H}$ can be calculated as the solution of (with $\mathbf{\Gamma}_H' = \frac{\partial}{\partial H}\mathbf{\Gamma}_H$)

$$\frac{\partial}{\partial H}L(\mathbf{X}, H) = \mathrm{Tr}(\mathbf{\Gamma}_H^{-1}\,\mathbf{\Gamma}_H') \tag{12}$$
$$+ n\frac{(\mathbf{v} + \mathbf{u})^{\mathrm{t}}\left(\mathbf{\Gamma}_H^{-1}\,\mathbf{\Gamma}_H'\,\mathbf{\Gamma}_H^{-1}\right)(\mathbf{v} + \mathbf{u})}{(\mathbf{v} + \mathbf{u})^{\mathrm{t}}\,\mathbf{\Gamma}_H^{-1}\,(\mathbf{v} - \mathbf{u})} = 0.$$

Note, that solving (12) we do not need to calculate the determinant of $\mathbf{\Gamma}_H$.

IV. GEOMETRICAL SAMPLING

The Hurst parameter H describes the scaling behaviour of the traffic. Therefore, in order to determine its value from measured traffic, the sample points have to cover several time scales, i.e. the total time of the measurements has to be many orders of magnitude greater than the smallest interval between the sampling points. With the ordinary linear sampling, i.e. sampling points at constant intervals, this leads to the requirement of very large number of sampling points. Obviously, because of the correlations, there is a lot of redundancy in measured traffic values at these points. In order to use the measurements more efficiently we introduce a geometric sequence of sampling points, $t_i = \alpha^i$, $i = 1, \ldots, n$, with some α between zero and one.

In addition to distributing the sampling points in a better way on different time scales, geometric sampling fits neatly
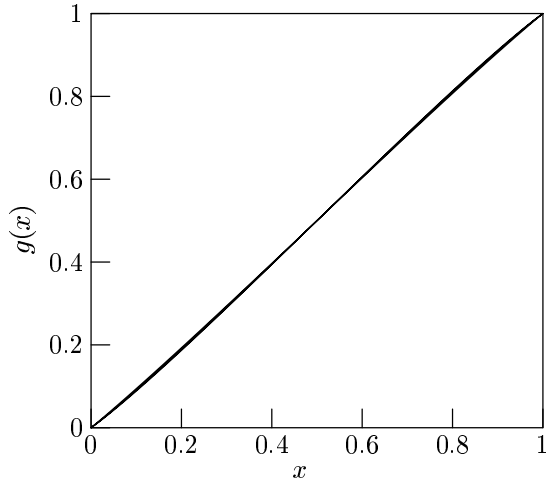
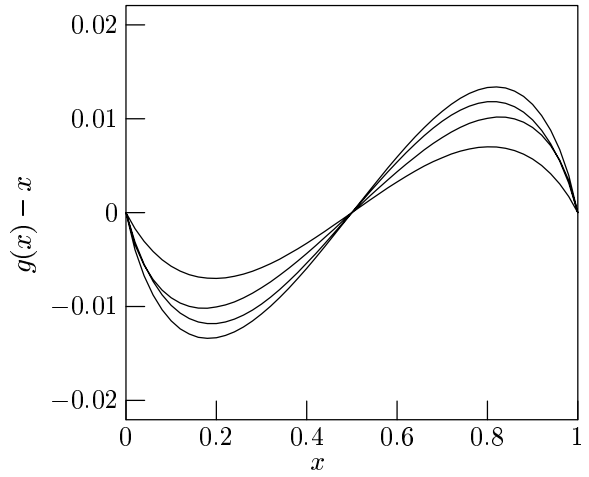Fig. 1. The approximation $g(x) \approx x$ for $H = 0.6, 0.7, 0.8$ and 0.9.

Fig. 2. Error of approximation $g(x) \approx x$ for $H = 0.6, 0.7, 0.8$ and 0.9.

with the self-similar behaviour of the fBm traffic. We show first that by a simple transformation we can obtain from the fBm process another process which is a stationary process of logarithmic time. As a geometric sequence corresponds to equidistant points in logarithmic time, the samples of the modified process constitute a stationary sequence. This leads to a simple Toeplitz-type structure of the covariance matrix and allows us to develop approximations to the inverse and determinant of the covariance matrix. (Similar ideas were presented in [10], [11] where the notion of scale-stationarity was used. The ML estimation technique was also investigated to some extent, but no approximations were used to make the method practically tractable.)

A. Descaled process

There is a one to one correspondence between self-similar and stationary processes: $Y = \{Y(t), t \geq 0\}$ is self-similar with parameter H if and only if $\{e^{-tH}Y(e^t), -\infty < t < \infty\}$ is stationary [12]. This transformation can be achieved by first *descaling* the process and then distorting the time axis exponentially. The descaled process $\{t^{-H}Y(t)\}$ can be obtained from the original Y process by multiplying it with the term t^{-H}, and the time distortion can be achieved by changing the process' index to logarithmic time, i.e., $u = -\log t$, where u is the new process index. (The minus sign here is appropriate for us when dealing with the interval $t \in (0, 1]$.)

If we descale the process $X(t)$ by the factor t^{-H} and use u as the process index, we finally have

$$\tilde{X}(u) \overset{d}{=} me^{(H-1)u} + \sqrt{a}\tilde{Z}(u), \tag{13}$$

where $\tilde{Z}(u) \overset{d}{=} t^{-H}Z(t)$ is the descaled fractional Brownian motion with the following property:

$$
\begin{aligned}
\tilde{Z}(u - \log \alpha) &= e^{-H(u-\log \alpha)}Z(e^{-u+\log \alpha}) \tag{14}\\
&= \alpha^H t^{-H} Z(\alpha t) \sim t^{-H}Z(t)\\
&= \tilde{Z}(u),
\end{aligned}
$$

where we have used the self-similar property of $Z(t)$, $Z(\alpha t) \sim \alpha^H Z(t)$. Thus the process $\tilde{Z}(u)$ is stationary and has the following covariance structure:

$$\mathrm{Cov}\left[\tilde{Z}(u_1), \tilde{Z}(u_2)\right] = \tag{15}$$

$$\tfrac{1}{2}e^{H(u_2-u_1)}\left\{1 + e^{-2H(u_2-u_1)} - \left(1 - e^{-(u_2-u_1)}\right)^{2H}\right\}.$$

We see that the descaled process $\tilde{Z}(u)$ is short range dependent.

Note, that our geometrical grid is now equally spaced with regard to u. The covariance matrix $\tilde{\boldsymbol{\Gamma}}$ of the descaled samples $\tilde{\mathbf{X}} = (\tilde{X}(u_1), \tilde{X}(u_2), \ldots, \tilde{X}(u_n))^{\mathrm{t}}$ with $u_i = -\log t_i = (1-i)\log \alpha$ can be written as

$$\tilde{\boldsymbol{\Gamma}} = \mathrm{E}\left[\tilde{\mathbf{X}}\tilde{\mathbf{X}}^{\mathrm{t}}\right] = a \cdot \mathrm{E}\left[\tilde{\mathbf{Z}}\tilde{\mathbf{Z}}^{\mathrm{t}}\right]. \tag{16}$$

Thus, if we use the notation $\tilde{Z}_i = \tilde{Z}(u_i)$ the process $\tilde{\mathbf{Z}} = (\tilde{Z}_1, \tilde{Z}_2, \ldots, \tilde{Z}_n)$ is a stationary process in discrete time with zero mean and unit variance and its auto-correlation function $\rho(k)$ can be defined as

$$\rho(i - j) = \frac{1}{2}\alpha^{-H|i-j|}\left\{1 + \alpha^{2H|i-j|} - \left(1 - \alpha^{|i-j|}\right)^{2H}\right\}, \tag{17}$$

and thus

$$\tilde{\boldsymbol{\Gamma}}_{ij} = a\rho(i - j), \quad i, j = 1, 2, \ldots, n. \tag{18}$$

B. Descaled MLE

When doing the maximum likelihood estimation of the model parameters m, a and H, one can utilize the stationarity and short range dependent properties of the descaled process. Using the 'descaling matrix' $\mathbf{D} = \mathrm{diag}(t_1^{-H}, \ldots, t_n^{-H})$ we can easily derive $\tilde{\boldsymbol{\Gamma}} = \mathbf{D}\,\boldsymbol{\Gamma}\,\mathbf{D}$, and from this we get

$$\boldsymbol{\Gamma}_H^{-1} = \mathbf{D}\tilde{\boldsymbol{\Gamma}}_H^{-1}\mathbf{D}. \tag{19}$$

The determinant $|\boldsymbol{\Gamma}_H|$ can be also calculated as

$$|\boldsymbol{\Gamma}_H| = \alpha^{Hn(n-1)}|\tilde{\boldsymbol{\Gamma}}_H|. \tag{20}$$

C. Approximate MLE

In practice, the exact MLE poses computational problems because of the computation time needed in case of large data sets. To avoid these problems, one can use approximate methods to calculate the estimates. In [2], several possible approaches to approximating the Gaussian likelihood function are discussed, among them the well known Whittle's approximate MLE.

In our case we focus on the properties of the covariance matrix $\boldsymbol{\Gamma}_H$, trying to take advantage of its special structure and to find efficient approximations for its inverse and determinant.

Using the notations

$$g(x) = \frac{1}{2}\left(1 + x^{2H} - (1-x)^{2H}\right), \tag{21}$$

the elements of the autocorrelation matrix $\tilde{\boldsymbol{\Gamma}}_H$ can be written as

$$[\tilde{\boldsymbol{\Gamma}}_H]_{i,j} = \alpha^{-H|i-j|}g(\alpha^{|i-j|}), \quad i,j = 1,2,\ldots,n. \tag{22}$$

It is interesting to note, that $g(x)$ is very nearly linear for $x \in (0,1)$. Fig. 2 shows the function $g(x)$ and the difference of $g(x) - x$ for different values of H. It can be seen from the plot that the largest absolute difference is less than 0.02 for each value of H. This observation gives us the idea to use the approximation $g(x) \approx x$. So $\tilde{\boldsymbol{\Gamma}}_H$ can be approximated as $\tilde{\boldsymbol{\Gamma}}_H \approx \mathbf{R}$, where $\mathbf{R}$ is a Toeplitz-type matrix of the form

$$[\mathbf{R}]_{ij} = \gamma^{|i-j|}, \quad i,j = 1,2,\ldots n, \tag{23}$$

with $\gamma = \alpha^{1-H}$.

The inverse and the determinant of $\mathbf{R}$ can be easily calculated as [13]

$$\mathbf{R}^{-1} = \frac{1}{\frac{1}{\gamma}-\gamma}\begin{pmatrix} \frac{1}{\gamma} & -1 & 0 & \cdots & 0 \\ -1 & \gamma+\frac{1}{\gamma} & -1 & \ddots & \vdots \\ 0 & -1 & \gamma+\frac{1}{\gamma} & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & -1 \\ 0 & \cdots & 0 & -1 & \frac{1}{\gamma} \end{pmatrix},$$
$$\tag{24}$$

$$|\mathbf{R}| = (1-\gamma^2)^{n-1}. \tag{25}$$

Using the fact that $\mathbf{t}^t \mathbf{D}\mathbf{R}^{-1}\mathbf{D}\,\mathbf{t} = 1$ and $\mathbf{t}^t \mathbf{D}\mathbf{R}^{-1}\mathbf{D} = (1,0,\ldots,0)$, to get an estimate for H we have to minimize the function

$$L(\mathbf{X};H) = \frac{n-1}{n}\log\left(\alpha^{nH}(1-\alpha^{2-2H})\right) \tag{26}$$
$$+ \log\left(\mathbf{X}^t\,\mathbf{D}\,\mathbf{R}^{-1}\mathbf{D}\,\mathbf{X} - X_1^2\right).$$

Note that due to the relatively simple structure of $\mathbf{R}^{-1}$ the matrix product term in (26) can be calculated as

$$\mathbf{X}^t\,\mathbf{D}\,\mathbf{R}^{-1}\mathbf{D}\,\mathbf{X} = \frac{1}{1-\gamma^2}\sum_{i=1}^{n-1}\left(\tilde{X}_i - \gamma\tilde{X}_{i+1}\right)^2 + \tilde{X}_n^2. \tag{27}$$

It should be noted that though the linear approximation to $g(x)$ is rather accurate, the resulting inverse matrix $\mathbf{R}^{-1}$ of (24) is rather poor an approximation to $\tilde{\boldsymbol{\Gamma}}^{-1}$ for large n. Nevertheless, the use of $\mathbf{R}^{-1}$ in the log-likelihood function (26), as we will see, yields a good estimate for H.

V. ALTERNATIVE METHODS

Two alternative estimation methods are presented here for comparison to our proposed estimator. The first one is the same MLE technique in time domain but using linear sampling scheme, and the second is the wavelet based parameter estimation method proposed by Abry and Veitch [5], [6].

A. Linear sampling

Let $\mathbf{X} = (X(t_1), X(t_2), \ldots, X(t_n))^t$ be the vector of observed traffic values at instances

$$t_i = \frac{i}{n}, \quad i = 1,2,\ldots,n. \tag{28}$$

The increment sequence $(Y_1, Y_2, \ldots)$ with $Y_i = X(t_i) - X(t_{i-1})$ (substituting $X(t_0) \equiv X(0) = 0$) is a strongly correlated stationary sequence with

$$\mathrm{Cov}\,[Y_i, Y_j] = \frac{1}{2}an^{-2H}\left(|i-j+1|^{2H} + |i-j-1|^{2H}\right.$$
$$\left. - 2|i-j|^{2H}\right), \quad i,j = 1,2,\ldots,n. \tag{29}$$

The formulas for the exact Gaussian MLE for this increment process are nearly the same as in Section III, we only need to replace the covariance matrix $\boldsymbol{\Gamma}$ with $\boldsymbol{\Sigma} = [\mathrm{Cov}\,[Y_i, Y_j]]_{i,j=1,2,\ldots,n}$, and the vector $\mathbf{t}$ with the vector $(1/n, 1/n, \ldots, 1/n)^t$. After some minor simplifications we have to minimize

$$\tilde{L}(\mathbf{Y};H) = \log|\boldsymbol{\Sigma}_H| \tag{30}$$
$$+ n\log\left(\mathbf{Y}^t\,\boldsymbol{\Sigma}_H^{-1}\,\mathbf{Y} - \frac{\left(\mathbf{1}^t\,\boldsymbol{\Sigma}^{-1}\,\mathbf{Y}\right)^2}{\mathbf{1}^t\,\boldsymbol{\Sigma}^{-1}\,\mathbf{1}}\right).$$

The minimum is obtained for some value $\hat{H}$ which is the ML estimate.

However, to calculate the inverse and the determinant of $\boldsymbol{\Sigma}_H$ the same problems arise as in the case of geometrical sampling with the covariant matrix $\boldsymbol{\Gamma}_H$.

B. Wavelet based estimator

Here we use the notations of Veitch and Abry [6]. The continuous wavelet decomposition consists of the collection of coefficients

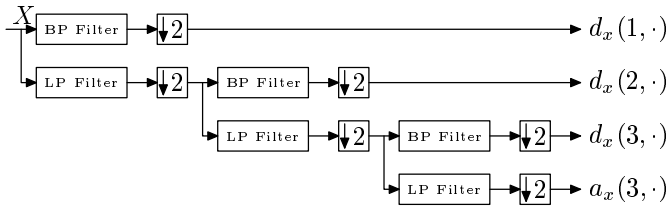$$\{T_X(a,t) = \langle X, \psi_{a,t}\rangle, \ a \in \mathbb{R}^+, t \in \mathbb{R}\} \tag{31}$$

Fig. 3. Filter bank

that compares the signal X to be analyzed with a set of analyzing functions (or *wavelets*)

$$\left\{ \psi_{a,t}(u) \equiv \frac{1}{\sqrt{a}} \psi_0 \left(\frac{u-t}{a} \right), \ a \in \mathbb{R}^+, t \in \mathbb{R} \right\}. \quad (32)$$

This set of analyzing functions is constructed from a reference pattern ψ_0, called the mother-wavelet.

Because the wavelet transform represents in a plane the information contained in a signal, it is a redundant transform. A mathematical theory, the *Multiresolution Analysis*, proves that it is possible to keep, among the $\{T_X(a,t)\}$, only a discrete set of coefficients while still retaining the total information in X. The discrete (or nonredundant) wavelet transform consists of the collection of coefficients

$$X(t) \rightarrow \quad \Big\{ \{a_X(J,k), \ k \in \mathbb{Z}\}, \quad (33)$$

$$\{d_X(j,k), j = 1, \dots, J, \ k \in \mathbb{Z}\} \Big\}.$$

The $\{d_X(j,k)\}$ constitute a subsample of the $\{T_X(a,t)\}$, located on the so-called dyadic grid $d_X(j,k) = T_X(2^j, 2^j k)$. These coefficients can be computed by a fast recursive filter-bank based pyramidal algorithm (see Fig. 3) with low computational cost [14]. (The coefficients of the low-pass (LP) and band-pass (BP) filters are derived from the chosen wavelet.)

When X is a self-similar process, the wavelet coefficients $d_X(j,k)$ exactly reproduce the self-similarity through the scaling property [15], i.e.,

$$d_X(j,k) \overset{d}{=} 2^{j(H+0.5)} d_X(0,k) \quad (34)$$

for all j and k. If we add the requirement that X has stationary increments, we have

$$\mathrm{E}\left[d_X(j,k)^2\right] = 2^{j(2H+1)} C(H, \psi_0), \quad (35)$$

with a constant C independent of j.

The quasi-decorrelation of the $d_X(j,k)$ coefficients allows us to effectively use the simple 'time average' as an estimate of $\mathrm{E}\left[d_X(j,\cdot)^2\right]$

$$\mu_j = \frac{1}{n_j} \sum_{k=1}^{n_j} d_X(j,k)^2, \quad (36)$$

where n_j is the number of coefficients at *octave* j (i.e., essentially $n_j = 2^{-j} n$ where n is the length of the data).

This quantity is an unbiased and consistent estimator of $\mathrm{E}\left[d_X(j,\cdot)^2\right]$ [16].

The power-law form in (35) suggests that the scaling exponent $2H+1$ could be extracted simply by considering the slope in a plot of $\log_2(\mu_j)$ against j. However, care should be taken since nonlinearity is introduced by the $\log_2$, which biases the estimator. The fundamental approach underlying the Abry-Veitch estimator proposed in [6] is a weighted linear regression of $\log_2(\mu_j)$ on j.

VI. SIMULATION

A. Data sets

For small (up to 10^3) sample sizes the fBm samples were generated using the fact $\mathbf{Z} \sim \mathbf{\Gamma}_H^{1/2} \mathbf{N}$ (or, correspondingly, $\mathbf{Z} \sim \mathbf{\Sigma}_H^{1/2} \mathbf{N}$ for the linear sampling) where $\mathbf{N}$ is a vector of independent standard Gaussian variables. To generate larger data sets, the conditioned Random Midpoint Displacement (RMD) algorithm presented in [17] was used. This method provides a fast and accurate approximation of fractional Brownian motion. (The RMD-mn simulator is available over the Internet [17].)

In the simulation presented here the model parameters were set as $m = 1$, $a = 1$ and $H = 0.8$ as an example, but similar results were obtained for different values of the parameters.

B. Parameter α

Parameter α controls the number of samples and the total sample time. With a given number of samples the ratio of the total measurement time to the shortest time interval between sample points (resolution) is greater if α is smaller. To cover many time scales a small α is desirable. On the other hand, the resolution of the measurement can not be arbitrarily fine because of practical limitations (e.g., the smallest time difference our measurement equipment can record is given, or its time stamp precision is finite).

The parameter α for the geometrical grid was chosen so that the difference between the nearest two measurement time instants (the 'resolution' of the measurement) was 10^{-4}.

C. Results

Fig. 4 shows the results for the estimates of H as a function of the number of sample points using both geometrical and linear sampling. In the geometrical case (9) was minimized while for the linear sampling we minimized the formula (30). The 95% confidence interval was obtained by repeating the simulations 100 times and calculating the sample variance of the estimates.

The results show that the estimates using geometrical sampling have much smaller variance and are unbiased for sample sizes larger than 25. For the linear sampling the variance of the estimates for a given sample size is always higher than in the geometrical case. For example, the variance for 800 samples using linear sampling is nearly the same as for only 50 geometrically sampled points. This dramatic reduction in
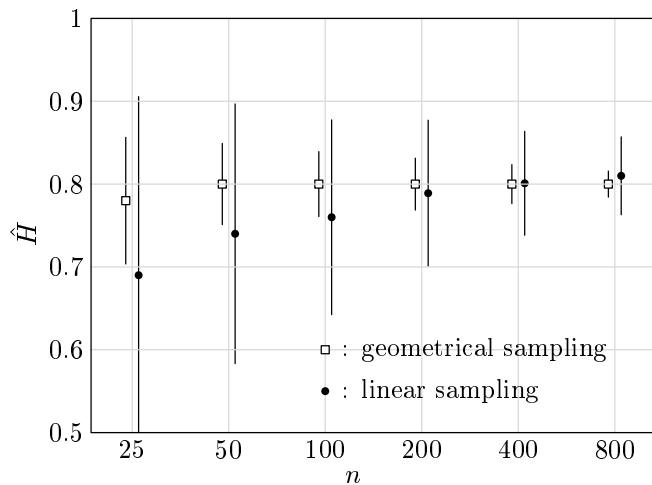
Fig. 4. Maximum likelihood estimates of H using geometrical and linear sampling.

Fig. 5. Estimates of H using MLE with geometrical sampling, and the wavelet-based Abry-Veitch estimator.

the number of samples highlights the main difference between geometrical and linear sampling: having the same total measurement interval and the same resolution, the number of samples needed is much less in the geometrical case. (Or equivalently, having the same measurement interval and number of samples, the resolution is higher in our method and thus the variance of the estimates is smaller.)

Next, our method was compared to the wavelet-based Abry-Veitch estimator. Fig. 5 shows similar results as previously, namely, than the estimates using geometrical sampling have much smaller variance and bias than the wavelet based method. For a given variance of $\hat{H}$ the reduction in the number of samples is about one order of magnitude in favor of MLE using geometrical sampling.

VII. CONCLUSION

In this paper we have introduced the idea of using geometrical sampling for the ML estimation of the parameters of fractional Brownian traffic. Intuitively, the geometrical sampling distributes the sampling points advantageously at different time scales, whereas linear sampling stresses the finest time scale and contains redundant information.

We have derived expressions for the estimators of m and a and the log-likelihood function from which the estimator of H can be derived for the geometrical sampling. Approximations were developed for the inverse and the determinant of the covariance matrix, needed for the calculation of the estimates. With these approximations the evaluation of the log-likelihood function is fast and the maximization with respect to H can easily be made.

The experiments with simulated traffic showed that the geometrical sampling does indeed give a better estimate for H leading to a reduction of sample points. In one example the number of required points was reduced from 800 to 50.
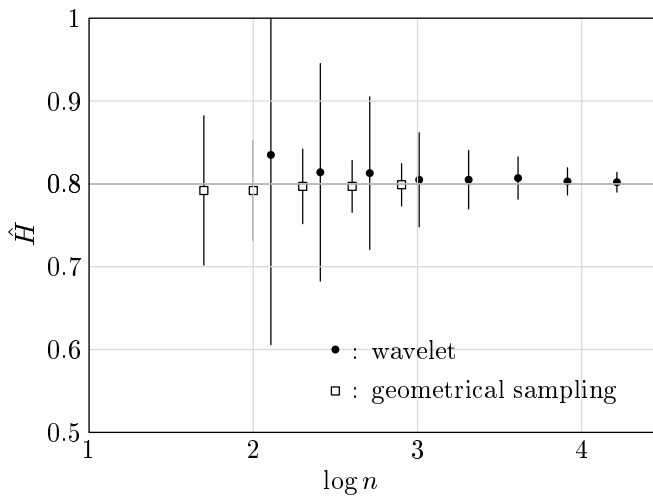
It should, however, be noted that the experiments were

made only with simulated traffic with exact 'measurements'. If the measured values are noisy then the descaling factor may amplify the noise of the points near the origin. The wavelet based Abry-Veitch estimator, for example, shows good performance in terms of robustness and computational cost. Similar investigations are necessary for our method and are the topic of future study.

Though the geometrical sampling has been shown to give better results than the linear sampling, it is not claimed that it constitutes the optimal sampling scheme. So there remains the theoretical question what is the best way of locating a given number of sampling points in the interval $[0, 1]$ with the constraint that the smallest distance between any pair of points is greater than or equal to a given minimum resolution.

REFERENCES

[1] J. Roberts, U. Mocci, and J. Virtamo, Eds., *Broadband Network Teletraffic—Performance Evaluation and Design of Broadband Multiservice Networks*, Springer, 1996, Final Report of COST 242.
[2] Jan Beran, *Statistics for Long-Memory Processes*, Chapman & Hall, One Penn Plaza, New York, NY 10119, 1995.
[3] M. Deriche and A.A. Tewfik, "Maximum likelihood estimation of the parameters of discrete fractionally differenced Gaussian noise process," *IEEE Trans. Signal Proc.*, vol. 41, no. 10, pp. 2977–2989, Oct. 1993.
[4] B. Ninness, "Maximum likelihood estimation of the parameters of fractional Brownian motions," in *Proc., 34th Conference on Decision & Control*, New Orleans, LA, Dec. 1995, pp. 4018–4023.
[5] P. Abry, P. Flandrin, M.S. Taqqu, and D. Veitch, "Wavelets for the analysis, estimation and synthesis of scaling data," *Self-Similar Network Traffic Analysis and Performance Evaluation*, K. Park and W. Willinger, Eds., in press, 1999.
[6] D. Veitch and P. Abry, "A wavelet based joint estimator of the parameters of long-range dependence," *Special Issue of the IEEE Trans. on Information Theory, Multiscale Statistical Signal Analysis and its Applications*, in press, 1998.

[7] A. Vidács and J.T. Virtamo, "Maximum likelihood estimation of the parameters of fractional Brownian traffic with geometrical sampling," in *Proc., Broadband Communications (BC'99)*, Hong Kong, Nov. 1999, pp. 51–62.

[8] I. Norros, "A storage model with self-similar input," *Queueing Systems*, vol. 16, pp. 387–396, 1994.

[9] C.R. Dietrich and M.R. Osborne, "Efficient computation of the restricted maximum likelihood function and its gradient for variance estimation of a stationary gaussian random field sampled over a regular grid," in *Proc., Computational Techniques and Applications Conference (CTAC93)*. July 1993, pp. 184–192, World Scientific.

[10] B. Yazıcı and R.L. Kashyap, "Signal modeling and parameter estimation for $1/f$ processes using scale stationary models," in *Proc., IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP-96)*, Atlanta, Georgia, May 1996, vol. 5, pp. 2841–2844.

[11] B. Yazıcı and R.L. Kashyap, "A class of second-order stationary self-similar processes for $1/f$ phenomena," *IEEE Trans. on Signal Processing*, vol. 45, no. 2, pp. 396–410, Feb. 1997.

[12] G. Samorodnitsky and M.S. Taqqu, *Stable Non-Gaussian Processes: Stochastic Models with Infinite Variance*, Chapman & Hall, New York, London, 1994.

[13] P. Rózsa, *Linear Algebra and its Applications*, University Press, Budapest, 3rd edition, 1991, (in Hungarian).

[14] I. Daubechies, *Ten Lectures on Wavelets*, SIAM, Philadelphia (PA), 1992.

[15] L. Delbeke and W. Van Assche, "A wavelet based estimator for the parameter of self-similarity of fractional Brownian motion," in *Proc., 3rd International Conference on Approximation and Optimization in the Caribbean*, Puebla, Mexico, Oct. 1995, *Applied and Computational Harmonic Analysis*, in press, 1997.

[16] P. Abry and D. Veitch, "Wavelet analysis of long range dependent traffic," *IEEE Trans. Inform. Theory*, vol. 44, no. 1, pp. 2–15, Jan. 1998.

[17] I. Norros, P. Mannersalo and J.L. Wang, "Simulation of Fractional Brownian Motion with Conditionalized Random Midpoint Displacement" *Adv. Performance Anal.*, vol. 2, pp. 77–101, 1999.
http://www.vtt.fi/tte/tte23/cost257/