Time Domain MLE of the Parameters of FBM Traffic

A. Vidács[†1] and J.T. Virtamo[‡]

[†]High Speed Networks Laboratories
Dept. of Telecommunications and Telematics
Technical University of Budapest
Pázmány Péter sétány 1/D, H-1117 Budapest, Hungary
Tel: +361 463 3422, Fax: +361 463 3107
E-mail: vidacs@ttt-atm.ttt.bme.hu

[‡]Laboratory of Telecommunications Technology
Helsinki University of Technology
P.O.B. 3000, FIN-02015 HUT, Finland
Tel: +358 9 451 2461, Fax: +358 9 451 2474
E-mail: jorma.virtamo@hut.fi

May 20, 1999

[1]while working at the Lab. of Telecommunications Technology, Helsinki University of Technology.

Abstract

This document is the collection of mathematical expressions and formulas related to the time domain maximum likelihood estimation (MLE) of the parameters of fractional Brownian traffic (fBt). No detailed descriptions and explanations for the formulas or proofs for theorems are given, only brief notes and the structure of the document link them together. This compact structure is believed to serve the purpose of this paper, namely, to give a useful handbook of formulas when one is working on this topic thoroughly.

For a more detailed presentation of the main ideas and results of this work the reader is referred to [20]. An extensive list of references is provided to give an overview of the related literature.

Tiivistelmä

Tämä raportti sisältää kokoelman matemaattisia kaavoja, joita tarvitaan sovellettaessa suurimman uskottavuuden menetelmää (maximum likelihood estimation, MLE) fraktionaaliseen Brownin liikkeeseen (fractional Brownian traffic, fBt) perustuvan liikennemallin parametrien estimointiin. Kaavoja ei tässä raportissa johdeta tai perustella yksityiskohtaisesti, eikä teoreemoille anneta todistuksia; ainoastaan lyhyet selitykset ja dokumentin rakenne sitovat ne toisiinsa. Tällaisen tiiviin esitystavan uskomme parhaiten palvelevan raportin tarkoitusta toimia hyödyllisenä käsikirjana niille, jotka työskentelevät tämän aiheen parissa.

Yksityiskohtaisempi kuvaus tämän työn taustalla olevista pääideoista ja saavutetuista tuloksista löytyy viitteestä [20]. Raporttiin liittyy myös laajahko luettelo aiheeseen liittyvästä muusta kirjallisuudesta.

Contents

1 Definitions, properties and characterization

1.1 Fractional Brownian motion

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 [16]:

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.

$Z(t)$ is a self-similar process whose scaling behaviour is defined by the Hurst-parameter H as follows

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

One can deduce

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

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 [16].

1.2 Fractional Brownian traffic

Fractional Brownian motion is a popular model for long-range dependent traffic. I. Norros [16] 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 $Z(t)$.

The process $X(t)$ has the following properties:

$$\mathrm{E}[X(t)] = mt; \tag{4}$$

$$\mathrm{Var}[X(t)] = at^{2H}; \tag{5}$$

$$\mathrm{Cov}[X(t_1), X(t_2)] = \frac{a}{2}\left\{t_1^{2H} + t_2^{2H} - |t_2 - t_1|^{2H}\right\}. \tag{6}$$

It follows from Eq.(3) and the above properties of $Z(t)$ that $(X(t+1) - X(t))$, the number of arrivals in the unit interval $[t, t+1)$, has the following mean and variance:

$$\mathrm{E}[X(t+1) - X(t)] = m, \tag{7}$$

$$\mathrm{Var}[X(t+1) - X(t)] = a. \tag{8}$$

Thus, the scale factor a gives the variance of arrivals over *one unit of the chosen time-scale.*

2 Exact Gaussian Maximum Likelihood Estimation

We use the notation of Beran [2]. Assume the traffic has been observed at n time instances forming the vector $\mathbf{t} = (t_1, \ldots, t_n)^{\mathrm{t}}$. And let $\mathbf{X} = (X(t_1), \ldots, X(t_n))^{\mathrm{t}}$ be the vector of observed traffic values at these instances. Since $X(t)$ is Gaussian, the joint distribution function of $\mathbf{X}$ is equal to

$$h(\mathbf{x}) = (2\pi)^{-\frac{n}{2}} |\Gamma|^{-\frac{1}{2}} e^{-\frac{1}{2}(\mathbf{x}-\mathbf{m})^{\mathrm{t}}\mathbf{\Gamma}^{-1}(\mathbf{x}-\mathbf{m})}, \tag{9}$$

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

$$\Gamma = \Big[\mathrm{Cov}\left[X(t_i), X(t_j)\right]\Big]_{i,j=1,\ldots,n}. \tag{10}$$

2.1 MLE $(\hat{m}|\, a, H)$

The log-likelihood function is given by

$$\log h(\mathbf{X}; m) = -\frac{n}{2}\log 2\pi - \frac{1}{2}\log|\Gamma| - \frac{1}{2}(\mathbf{X} - m\mathbf{t})^{\mathrm{t}}\,\Gamma^{-1}(\mathbf{X} - m\mathbf{t}). \tag{11}$$

The MLE of m is obtained by maximizing $\log h(\mathbf{X}; m)$ with respect to m. Or, equivalently, by minimizing the function

$$L(\mathbf{X}; m) = (\mathbf{X} - m\mathbf{t})^{\mathrm{t}}\,\Gamma^{-1}(\mathbf{X} - m\mathbf{t}). \tag{12}$$

This minimization problem can be reformulated in terms of the first partial derivatives. The MLE $\hat{m}$ is the solution of

$$\begin{aligned} L'(\mathbf{X}; \hat{m}) &= \frac{\partial}{\partial m} L(\mathbf{X}; \hat{m}) \\ &= -2\hat{m}\mathbf{t}^{\mathrm{t}}\,\Gamma^{-1}\mathbf{t} - 2\mathbf{t}^{\mathrm{t}}\,\Gamma^{-1}\mathbf{X} = 0 \end{aligned} \tag{13}$$

and we get

$$\hat{m} = \hat{m}(H) = \frac{\mathbf{t}^{\mathrm{t}}\,\Gamma^{-1}\mathbf{X}}{\mathbf{t}^{\mathrm{t}}\,\Gamma^{-1}\mathbf{t}}. \tag{14}$$

2.2 MLE $(\hat{a}|\, m, H)$

$\Gamma = \Gamma(a)$ is a simple linear function of a:

$$\Gamma = \Gamma(a) = a\,\Gamma_H \tag{15}$$

where

$$\Gamma_H = \mathrm{E}\left[\mathbf{Z}\mathbf{Z}^{\mathrm{t}}\right] = \Big[\mathrm{Cov}\left[Z(t_i), Z(t_j)\right]\Big]_{i,j=1,\ldots,n}. \tag{16}$$

The log-likelihood function is given by

$$\log h(\mathbf{X}; a) = -\frac{n}{2}\log 2\pi - \frac{1}{2}\log a^n |\Gamma_H| - \frac{1}{2a}(\mathbf{X} - \mathbf{m})^{\mathrm{t}}\,\Gamma_H^{-1}(\mathbf{X} - \mathbf{m}). \tag{17}$$

The MLE of a is obtained by minimizing the a-dependent part of the log-likelihood function multiplied by -2, and that is

$$L(\mathbf{X}; a) = n\log a + \frac{1}{a}(\mathbf{X} - \mathbf{m})^{\mathrm{t}}\,\Gamma_H^{-1}(\mathbf{X} - \mathbf{m}). \tag{18}$$

The MLE $\hat{a}$ is the solution of

$$\begin{aligned} L'(\mathbf{X};\hat{a}) &= \frac{\partial}{\partial a} L(\mathbf{X};\hat{a}) \\ &= \frac{n}{\hat{a}} - \frac{1}{\hat{a}^2}(\mathbf{X}-\mathbf{m})^{\mathrm{t}}\,\Gamma_H^{-1}\,(\mathbf{X}-\mathbf{m}) = 0 \end{aligned} \tag{19}$$

and from this we get

$$\hat{a} = \hat{a}(m,H) = \frac{1}{n}(\mathbf{X}-\mathbf{m})^{\mathrm{t}}\,\Gamma_H^{-1}\,(\mathbf{X}-\mathbf{m}). \tag{20}$$

2.3 MLE $(\hat{m}, \hat{a}|\, H)$

If we don't know the mean input rate m in advance, in Eq.(20) $\mathbf{m}$ should be replaced by $\hat{m}\mathbf{t}$. Using Eq.(14) and Eq.(20) we get

$$\begin{aligned} \hat{a}(H) &= \frac{1}{n}(\mathbf{X}-\hat{m}\mathbf{t})^{\mathrm{t}}\,\Gamma_H^{-1}\,(\mathbf{X}-\hat{m}\mathbf{t}) \\ &= \frac{1}{n}\frac{(\mathbf{X}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{X})(\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{t}) - (\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{X})^2}{\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{t}}. \end{aligned} \tag{21}$$

2.4 MLE $(\hat{m}, \hat{a}, \hat{H})$

Finally, we are left with the maximization of the log-likelihood function

$$\log h(\mathbf{X};H) = -\frac{n}{2}\log 2\pi - \frac{1}{2}\log|\hat{a}(H)\,\Gamma_H| - \frac{1}{2\hat{a}(H)}(\mathbf{X}-\hat{m}(H)\mathbf{t})^{\mathrm{t}}\,\Gamma_H^{-1}\,(\mathbf{X}-\hat{m}(H)\mathbf{t}) \tag{22}$$

or equivalently, with the minimization of

$$\begin{aligned} L(\mathbf{X};H) &= \log|\hat{a}(H)\,\Gamma_H| + \frac{1}{\hat{a}(H)}(\mathbf{X}-\hat{m}(H)\mathbf{t})^{\mathrm{t}}\,\Gamma_H^{-1}\,(\mathbf{X}-\hat{m}(H)\mathbf{t}) \\ &= \log \hat{a}^n(H)\,|\Gamma_H| + \frac{1}{\hat{a}(H)}\frac{(\mathbf{X}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{X})(\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{t}) - (\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{X})^2}{\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{t}} \\ &= n\log\frac{(\mathbf{X}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{X})(\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{t}) - (\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{X})^2}{\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{t}} - n\log n + \log|\Gamma_H| + n \to \min, \end{aligned} \tag{23}$$

i.e., essentially we have to minimize

$$\tilde{L}(\mathbf{X};H) = |\Gamma_H|^{1/n}\frac{(\mathbf{X}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{X})(\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{t}) - (\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{X})^2}{\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{t}}. \tag{24}$$

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 MLE estimate; the corresponding MLE estimates for m and a are $\hat{m} = m(\hat{H})$ and $\hat{a} = a(\hat{H})$.

Alternatively, $\hat{H}$ can be calculated as the solution of

$$L'(\mathbf{X},H) = \frac{\partial}{\partial H}\Big(\log|\Gamma_H|\Big) + n\frac{\partial}{\partial H}\log\left[(\mathbf{X}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{X}) - \frac{(\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{X})^2}{\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{t}}\right] = 0. \tag{25}$$

Using the notations $\mathbf{u} = (\mathbf{t}^{\mathrm{t}}\Gamma_H^{-1}\mathbf{t})\mathbf{X}$ and $\mathbf{v} = (\mathbf{t}^{\mathrm{t}}\Gamma_H^{-1}\mathbf{X})\mathbf{t}$, after some calculations we have

$$L'(\mathbf{X}, H) = \frac{\partial}{\partial H}\Big(\log|\Gamma_H|\Big) + n\frac{(\mathbf{u}+\mathbf{v})^{\mathrm{t}}\left(\frac{\partial}{\partial H}\Gamma_H^{-1}\right)(\mathbf{u}+\mathbf{v})}{(\mathbf{u}+\mathbf{v})^{\mathrm{t}}\Gamma_H^{-1}(\mathbf{u}-\mathbf{v})} = 0. \tag{26}$$

If we use the relationships

$$\frac{\partial}{\partial\theta}\mathbf{A}^{-1} = -\mathbf{A}^{-1}\left(\frac{\partial}{\partial\theta}\mathbf{A}\right)\mathbf{A}^{-1}, \tag{27}$$

$$\frac{\partial}{\partial\theta}\log|\mathbf{A}| = \mathrm{Tr}\left(\mathbf{A}^{-1}\frac{\partial}{\partial\theta}\mathbf{A}\right) \tag{28}$$

valid for any matrix $\mathbf{A}$ depending on a parameter θ [8] we have (with $\Gamma'_H = \frac{\partial}{\partial H}\Gamma_H$)

$$L'(\mathbf{X}, H) = \mathrm{Tr}(\Gamma_H^{-1}\Gamma'_H) + n\frac{(\mathbf{v}+\mathbf{u})^{\mathrm{t}}\left(\Gamma_H^{-1}\Gamma'_H\Gamma_H^{-1}\right)(\mathbf{v}+\mathbf{u})}{(\mathbf{v}+\mathbf{u})^{\mathrm{t}}\Gamma_H^{-1}(\mathbf{v}-\mathbf{u})} = 0. \tag{29}$$

Note, that solving Eq.(29) we do not need to calculate the determinant of Γ_H.

2.5 Some remarks on the estimates

Let us first consider the estimate Eq.(14) for the mean $m = m(\hat{H})$. Since the expectation of $\mathbf{X}$ is $m\mathbf{t}$ we have

$$\mathrm{E}[\hat{m}] = \frac{\mathbf{t}^{\mathrm{t}}\hat{\Gamma}_H^{-1}\mathrm{E}[\mathbf{X}]}{\mathbf{t}^{\mathrm{t}}\hat{\Gamma}_H^{-1}\mathbf{t}} = m, \tag{30}$$

where $\hat{\Gamma}_H^{-1} \stackrel{def}{=} \Gamma_H^{-1}(\hat{H})$, i.e. the estimate is unbiased (irrespective of whether our estimate for H is correct or not).

The variance of $\hat{m}$ can also be calculated. For the time being we assume that H is known exactly, $\hat{H} = H$. (It will be interesting and important to consider also the case where $\hat{H}$ itself is a random variable!) We have

$$\begin{aligned}
\mathrm{Var}[\hat{m}] &= \mathrm{E}\left[(\hat{m} - \mathrm{E}[\hat{m}])^2\right] \\
&= a \cdot \mathrm{E}\left[\left(\frac{\mathbf{t}^{\mathrm{t}}\Gamma_H^{-1}\mathbf{Z}}{\mathbf{t}^{\mathrm{t}}\Gamma_H^{-1}\mathbf{t}}\right)^2\right] \\
&= a \cdot \frac{\mathrm{E}\left[(\mathbf{t}^{\mathrm{t}}\hat{\Gamma}_H^{-1}\mathbf{Z})(\mathbf{Z}^{\mathrm{t}}\hat{\Gamma}_H^{-1}\mathbf{t})\right]}{(\mathbf{t}^{\mathrm{t}}\hat{\Gamma}_H^{-1}\mathbf{t})^2} \\
&= a \cdot \frac{\mathbf{t}^{\mathrm{t}}\Gamma_H^{-1}\mathrm{E}[\mathbf{Z}\mathbf{Z}^{\mathrm{t}}]\Gamma_H^{-1}\mathbf{t}}{(\mathbf{t}^{\mathrm{t}}\Gamma_H^{-1}\mathbf{t})^2} \\
&= a \cdot \frac{\mathbf{t}^{\mathrm{t}}\Gamma_H^{-1}\mathbf{t}}{(\mathbf{t}^{\mathrm{t}}\Gamma_H^{-1}\mathbf{t})^2} \\
&= \frac{a}{\mathbf{t}^{\mathrm{t}}\Gamma_H^{-1}\mathbf{t}}.
\end{aligned} \tag{31}$$

where we have made use of the definition $\mathrm{E}[\mathbf{Z}\mathbf{Z}^t] = \Gamma_H$. The variance of our estimator is smaller than the estimator based on the total sample mean, by the factor in the denominator.

Next consider the estimator Eq.(21) for a, $\hat{a} = a(H)$. Again, for the time being we assume that H is known exactly and calculate the expectation of $\hat{a}$,

$$
\begin{aligned}
n\mathrm{E}[\hat{a}] &= \mathrm{E}\left[\mathbf{X}^t\Gamma_H^{-1}\mathbf{X}\right] - \frac{\mathrm{E}\left[(\mathbf{t}^t\Gamma_H^{-1}\mathbf{X})^2\right]}{\mathbf{t}^t\Gamma_H^{-1}\mathbf{t}} \\
&= \mathrm{E}\left[(m\mathbf{t}+\sqrt{a}\mathbf{Z})^t\Gamma_H^{-1}(m\mathbf{t}+\sqrt{a}\mathbf{Z})\right] - \frac{(\Gamma_H^{-1}\mathbf{t})^t\,\mathrm{E}[\mathbf{X}\mathbf{X}^t]\,(\Gamma_H^{-1}\mathbf{t})}{\mathbf{t}^t\Gamma_H^{-1}\mathbf{t}} \\
&= m^2(\mathbf{t}^t\Gamma_H^{-1}\mathbf{t}) + a\mathrm{E}\left[\mathbf{Z}^t\Gamma_H^{-1}\mathbf{Z}\right] - \frac{(\Gamma_H^{-1}\mathbf{t})^t(\mathrm{Var}[\mathbf{X}\mathbf{X}^t] + \mathrm{E}[\mathbf{X}]\,\mathrm{E}[\mathbf{X}^t])(\Gamma_H^{-1}\mathbf{t})}{\mathbf{t}^t\Gamma_H^{-1}\mathbf{t}} \\
&= m^2(\mathbf{t}^t\Gamma_H^{-1}\mathbf{t}) + a\mathrm{E}\left[\mathbf{Z}^t\Gamma_H^{-1}\mathbf{Z}\right] - \frac{\mathbf{t}^t\Gamma_H^{-1}(a\Gamma_H + m^2\mathbf{t}\mathbf{t}^t)\Gamma_H^{-1}\mathbf{t}}{\mathbf{t}^t\Gamma_H^{-1}\mathbf{t}} \\
&= (n-1)a,
\end{aligned}
\tag{32}
$$

where we have used $\mathrm{E}\left[\mathbf{Z}^t\Gamma_H^{-1}\mathbf{Z}\right] = \mathrm{E}[\mathbf{N}^t\mathbf{N}] = n$ since $\mathbf{Z} \sim \Gamma_H^{1/2}\mathbf{N}$ where $\mathbf{N}$ is a vector of independent standard Gaussian variables. Thus the variance estimator has the "normal" $(n-1)/n$ bias. The "naive" estimator [18] had a much larger bias because the correlations between the samples had not been "decorrelated".

The next step is the calculation of the variance of $\hat{a}$:

$$
\begin{aligned}
\mathrm{Var}[\hat{a}] &= \mathrm{E}\left[\hat{a}^2\right] - \mathrm{E}[\hat{a}]^2 \\
&= \mathrm{E}\left[\hat{a}^2\right] - \frac{(n-1)^2}{n^2}a^2.
\end{aligned}
\tag{33}
$$

To calculate $\mathrm{E}[\hat{a}^2]$ first we rewrite Eq.(21) as follows:

$$
\begin{aligned}
\hat{a} &= \frac{1}{n(\mathbf{t}^t\Gamma_H^{-1}\mathbf{t})}\left[(\mathbf{X}^t\Gamma_H^{-1}\mathbf{X})(\mathbf{t}^t\Gamma_H^{-1}\mathbf{t}) - (\mathbf{t}^t\Gamma_H^{-1}\mathbf{X})^2\right] \\
&= \frac{a}{n(\mathbf{t}^t\Gamma_H^{-1}\mathbf{t})}\underbrace{\left[(\mathbf{Z}^t\Gamma_H^{-1}\mathbf{Z})(\mathbf{t}^t\Gamma_H^{-1}\mathbf{t}) - (\mathbf{t}^t\Gamma_H^{-1}\mathbf{Z})^2\right]}_{A(\mathbf{Z})}
\end{aligned}
\tag{34}
$$

and now we have

$$
\mathrm{E}\left[\hat{a}^2\right] = \frac{a^2}{n^2(\mathbf{t}^t\Gamma_H^{-1}\mathbf{t})^2}\cdot\mathrm{E}\left[A(\mathbf{Z})^2\right].
\tag{35}
$$

To calculate the expectation $\mathrm{E}[A(\mathbf{Z})^2]$ we use the following equation:

$$
\begin{aligned}
\mathrm{E}\left[A(\mathbf{Z})^2\right] &= \mathrm{E}\left[A\left(\frac{\partial}{\partial\mathbf{s}}\right)^2 e^{\mathbf{s}^t\mathbf{Z}}\right]_{\mathbf{s}=0} \\
&= \left[A\left(\frac{\partial}{\partial\mathbf{s}}\right)^2 \mathrm{E}\left[e^{\mathbf{s}^t\mathbf{Z}}\right]\right]_{\mathbf{s}=0}
\end{aligned}
\tag{36}
$$

$$= \left[A \left(\frac{\partial}{\partial \mathbf{s}} \right)^2 M(\mathbf{s}) \right]_{\mathbf{s}=0}$$

with

$$M(\mathbf{s}) = \mathrm{E}\left[e^{\mathbf{s}^{\mathrm{t}}\mathbf{Z}} \right] = e^{\frac{1}{2}\mathbf{s}^{\mathrm{t}}\mathbf{\Gamma}_H \mathbf{s}}. \tag{37}$$

To proceed further, it is useful to derive the following expressions (with $\nabla_{\mathbf{s}}^{\mathrm{t}} = \partial/\partial\mathbf{s}$ and using $\nabla_{\mathbf{s}} M(\mathbf{s}) = (\mathbf{s}^{\mathrm{t}}\,\Gamma_H) M(\mathbf{s})$):

$$\begin{aligned} (\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\nabla_{\mathbf{s}}) M(\mathbf{s}) &= (\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\Gamma_H\,\mathbf{s}) M(\mathbf{s}) \\ &= (\mathbf{t}^{\mathrm{t}}\,\mathbf{s}) M(\mathbf{s}); \end{aligned} \tag{38}$$

$$(\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\nabla_{\mathbf{s}})^2 M(\mathbf{s}) = \left[(\mathbf{t}^{\mathrm{t}}\,\mathbf{s})^2 + (\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{t}) \right] M(\mathbf{s}); \tag{39}$$

$$\begin{aligned} (\nabla_{\mathbf{s}}^{\mathrm{t}}\,\Gamma_H^{-1}\nabla_{\mathbf{s}}) M(\mathbf{s}) &= (\nabla_{\mathbf{s}}^{\mathrm{t}}\,\Gamma_H^{-1}\,\Gamma_H\,\mathbf{s}) M(\mathbf{s}) \\ &= (\nabla_{\mathbf{s}}^{\mathrm{t}}\,\mathbf{s}) M(\mathbf{s}) \\ &= \left[n + (\mathbf{s}^{\mathrm{t}}\,\Gamma_H\,\mathbf{s}) \right] M(\mathbf{s}). \end{aligned} \tag{40}$$

Next, we derive

$$\begin{aligned} A(\nabla_{\mathbf{s}}) M(\mathbf{s}) &= \left[(\nabla_{\mathbf{s}}^{\mathrm{t}}\,\Gamma_H^{-1}\nabla_{\mathbf{s}})(\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{t}) - (\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\nabla_{\mathbf{s}})^2 \right] M(\mathbf{s}) \\ &= \left\{ (\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{t}) \left[n + (\mathbf{s}^{\mathrm{t}}\,\Gamma_H\,\mathbf{s}) \right] - \left[(\mathbf{t}^{\mathrm{t}}\,\mathbf{s})^2 + (\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{t}) \right] \right\} M(\mathbf{s}) \\ &= \left\{ (\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{t}) \left[(n-1) + (\mathbf{s}^{\mathrm{t}}\,\Gamma_H\,\mathbf{s}) \right] - (\mathbf{t}^{\mathrm{t}}\,\mathbf{s})^2 \right\} M(\mathbf{s}), \end{aligned} \tag{41}$$

and now we are ready to calculate

$$\begin{aligned} A(\nabla_{\mathbf{s}})^2 M(\mathbf{s}) &= \underbrace{(n-1)(\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{t}) A(\nabla_{\mathbf{s}}) M(\mathbf{s})}_{T_1(\mathbf{s})} \\ &+ \underbrace{A(\nabla_{\mathbf{s}}) \left[(\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{t})(\mathbf{s}^{\mathrm{t}}\,\Gamma_H\,\mathbf{s}) - (\mathbf{t}^{\mathrm{t}}\,\mathbf{s})^2 \right] M(\mathbf{s})}_{T_2(\mathbf{s})} \end{aligned} \tag{42}$$

According to Eq.(36) we need to get $T_1(0)$ and $T_2(0)$, As for the first term we have

$$T_1(0) = (n-1)^2 (\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{t})^2. \tag{43}$$

However, for $T_2(0)$ we have

$$\begin{aligned} T_2(\mathbf{s}) &= A(\nabla_{\mathbf{s}}) \underbrace{\left[(\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{t})(\mathbf{s}^{\mathrm{t}}\,\Gamma_H\,\mathbf{s}) - (\mathbf{t}^{\mathrm{t}}\,\mathbf{s})^2 \right]}_{T_3(\mathbf{s})} M(\mathbf{s}) \\ &= \left[(\nabla_{\mathbf{s}}^{\mathrm{t}}\,\Gamma_H^{-1}\nabla_{\mathbf{s}})(\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{t}) - (\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\nabla_{\mathbf{s}})^2 \right] T_3(\mathbf{s}) M(\mathbf{s}). \end{aligned} \tag{44}$$

To solve Eq.(44) it is useful to calculate the following terms:

$$
\begin{aligned}
(\mathbf{t}^{\mathrm{t}} ,_H^{-1} \nabla_{\mathbf{s}}) T_3(\mathbf{s}) M(\mathbf{s}) &= (\mathbf{t}^{\mathrm{t}}\, \mathbf{s}) T_3(\mathbf{s}) M(\mathbf{s}) + \underbrace{\left[2(\mathbf{t}^{\mathrm{t}} ,_H^{-1} \mathbf{t})(\mathbf{t}^{\mathrm{t}} ,_H^{-1} ,_H \mathbf{s}) - 2(\mathbf{t}^{\mathrm{t}} ,_H^{-1} \mathbf{t})(\mathbf{t}^{\mathrm{t}}\, \mathbf{s})\right]}_{0} M(\mathbf{s}) \\
&= \underbrace{\left[(\mathbf{t}^{\mathrm{t}}\, \mathbf{s})(\mathbf{t}^{\mathrm{t}} ,_H^{-1} \mathbf{t})(\mathbf{s}^{\mathrm{t}} ,_H \mathbf{s}) - (\mathbf{t}^{\mathrm{t}}\, \mathbf{s})^3\right]}_{T_4(\mathbf{s})} M(\mathbf{s}); \qquad (45)
\end{aligned}
$$

$$
\begin{aligned}
(\mathbf{t}^{\mathrm{t}} ,_H^{-1} \nabla_{\mathbf{s}})^2 T_3(\mathbf{s}) M(\mathbf{s}) &= (\mathbf{t}^{\mathrm{t}} ,_H^{-1} \nabla_{\mathbf{s}}) T_4(\mathbf{s}) M(\mathbf{s}) \qquad (46) \\
&= (\mathbf{t}^{\mathrm{t}}\, \mathbf{s}) T_4(\mathbf{s}) M(\mathbf{s}) + 3(\mathbf{t}^{\mathrm{t}} ,_H^{-1} \mathbf{t}) \left[(\mathbf{t}^{\mathrm{t}} ,_H^{-1} \mathbf{t})(\mathbf{s}^{\mathrm{t}} ,_H \mathbf{s}) - (\mathbf{t}^{\mathrm{t}}\, \mathbf{s})^2\right] M(\mathbf{s});
\end{aligned}
$$

$$
\begin{aligned}
(\nabla_{\mathbf{s}} ,_H^{-1} \nabla_{\mathbf{s}}) T_3(\mathbf{s}) M(\mathbf{s}) &= T_3(\mathbf{s})(\nabla_{\mathbf{s}}^{\mathrm{t}} \mathbf{s}) M(\mathbf{s}) + 2(\mathbf{t}^{\mathrm{t}} ,_H^{-1} \mathbf{t})(\nabla_{\mathbf{s}}^{\mathrm{t}} \mathbf{s}) M(\mathbf{s}) - 2(\mathbf{t}^{\mathrm{t}} ,_H^{-1} \nabla_{\mathbf{s}})(\mathbf{t}^{\mathrm{t}}\, \mathbf{s}) M(\mathbf{s}) \\
&= T_3(\mathbf{s}) \left[n + (\mathbf{s}^{\mathrm{t}} ,_H \mathbf{s})\right] M(\mathbf{s}) + 2(\mathbf{t}^{\mathrm{t}} ,_H^{-1} \mathbf{t}) \left[n + (\mathbf{s}^{\mathrm{t}} ,_H \mathbf{s})\right] M(\mathbf{s}) \qquad (47) \\
&- 2\left[(\mathbf{t}^{\mathrm{t}} ,_H^{-1} \mathbf{t}) + (\mathbf{t}^{\mathrm{t}}\, \mathbf{s})^2\right] M(\mathbf{s}).
\end{aligned}
$$

Note, that in Eq.(47) we used the results from Eq.(40). Since all terms in Eq.(46) are zero when $\mathbf{s} = 0$, from Eq.(44) using Eq.(47) and the fact that $T_3(0) = 0$ we get

$$T_2(0) = 2(n-1)(\mathbf{t}^{\mathrm{t}} ,_H^{-1} \mathbf{t})^2. \qquad (48)$$

Substituting this result and Eq.(43) into Eq.(36) using Eq.(42) we have

$$\mathrm{E}\left[A(\mathbf{Z})\right] = T_1(0) + T_2(0) = (n^2-1)(\mathbf{t}^{\mathrm{t}} ,_H^{-1} \mathbf{t})^2, \qquad (49)$$

and from Eq.(35) we get

$$\mathrm{E}\left[\hat{a}^2\right] = \frac{n^2-1}{n^2} a^2. \qquad (50)$$

Finally, from Eq.(33) we get the variance of the estimate $\hat{a}$ as

$$\mathrm{Var}\left[\hat{a}\right] = \frac{2(n-1)}{n^2} a^2. \qquad (51)$$

If we use the bias-corrected estimator $(n/(n-1))a$, the variance is

$$\mathrm{Var}\left[\frac{n}{n-1}\,\hat{a}\right] = \frac{2a^2}{n-1}. \qquad (52)$$

3 Linear sampling

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

$$t_i = \frac{i}{n}, \quad i = 1, 2, \ldots, n. \qquad (53)$$

3.1 The increment process

Define the increment process $Y = (Y_1, Y_2, \ldots, Y_n)$ as

$$
\begin{aligned}
Y_i &= X(t_i) - X(t_{i-1}) && (54)\\
&= X\left(\frac{i}{n}\right) - X\left(\frac{i-1}{n}\right) && (55)\\
&= \frac{m}{n} + \sqrt{a}\left[Z\left(\frac{i}{n}\right) - Z\left(\frac{i-1}{n}\right)\right], \quad i = 1, 2, \ldots, n. && (56)
\end{aligned}
$$

The process Y is a strongly correlated stationary sequence with

$$
\begin{aligned}
\mathrm{E}[Y] &= \frac{m}{n}, && (57)\\
\mathrm{Var}[Y] &= an^{-2H}, && (58)\\
\mathrm{Cov}[Y_i, Y_j] &= \frac{1}{2}an^{-2H}\left(|i-j+1|^{2H} + |i-j-1|^{2H} - 2|i-j|^{2H}\right), \quad i, j = 1, 2, \ldots, n && (59)
\end{aligned}
$$

Define the aggregated process $Y^{(\mu)}$ as

$$
Y_i^{(\mu)} = \sum_{j=(i-1)\mu+1}^{i\mu} Y_j, \quad i = 1, 2, \ldots, \left\lfloor \frac{n}{\mu} \right\rfloor. \tag{60}
$$

The expectation and variance of the aggregated process is given by

$$
\begin{aligned}
\mathrm{E}\left[Y^{(\mu)}\right] &= \mu\,\mathrm{E}[Y], && (61)\\
\mathrm{Var}\left[Y^{(\mu)}\right] &= \mu^{2H}\mathrm{Var}[Y], && (62)\\
\mathrm{Cov}\left[Y_i^{(\mu)}, Y_j^{(\mu)}\right] &= \mu^{2H}\mathrm{Cov}[Y_i, Y_j], \quad i, j = 1, 2, \ldots, n. && (63)
\end{aligned}
$$

3.2 The increments of the increment process

Define the process $B = (B_1, B_2, \ldots, B_{n-1})$ as

$$
\begin{aligned}
B_i &= Y_{i+1} - Y_i && (64)\\
&= X(t_{i+1}) - 2X(t_i) + X(t_{i-1})\\
&= a\left[Z\left(\frac{i+1}{n}\right) - 2Z\left(\frac{i}{n}\right) + Z\left(\frac{i-1}{n}\right)\right].
\end{aligned}
$$

The process B is stationary and has the following properties

$$
\begin{aligned}
\mathrm{E}[B] &= 0, && (65)\\
\mathrm{Var}[B] &= an^{-2H}\left(4 - 2^{2H}\right),\\
\mathrm{Cov}[B_i, B_j] &= \frac{1}{2}an^{-2H}\Big(-|i-j-2|^{2H} + 4|i-j-1|^{2H} - 6|i-j|^{2H}\\
&+ 4|i-j+1|^{2H} - |i-j+2|^{2H}\Big).
\end{aligned}
$$

The variance of B can be estimated by

$$S_B^2 = \frac{1}{n-1}\sum_{i=1}^{n-1} B_i. \tag{66}$$

The expectation $\mathrm{E}[S_B^2]$ can be calculated as

$$\mathrm{E}\left[S_B^2\right] = \frac{1}{n-1}\sum_{i=1}^{n-1}\mathrm{E}[B_i] = \mathrm{Var}[B]. \tag{67}$$

3.3 Sample variance

$$S_{(1)}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \frac{1}{n}\sum_{j=1}^{n} Y_j\right)^2. \tag{68}$$

With straightforward calculation, we have [18]

$$\begin{aligned}
\mathrm{E}\left[S_{(1)}^2\right] &= \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\left[\left(Y_i - \frac{1}{n}X(1)\right)^2\right] \qquad (69)\\
&= \frac{1}{n}\sum_{i=1}^{n}\left(\mathrm{E}\left[Y_i^2\right] - \frac{2}{n}\mathrm{E}[Y_iX(1)] + \frac{1}{n^2}\mathrm{E}\left[X(1)^2\right]\right)\\
&= \frac{1}{n}\sum_{i=1}^{n}\left(a\left(\frac{1}{n}\right)^{2H} - \frac{2a}{n}\mathrm{Cov}\left[Z\left(\frac{i}{n}\right) - Z\left(\frac{i-1}{n}\right), Z(1)\right] + \frac{a}{n^2}\right)\\
&= \frac{a}{n}\sum_{i=1}^{n}\left(n^{-2H} - c(n,i)n^{-2}\right)
\end{aligned}$$

with

$$c(n,i) = 2n\,\mathrm{Cov}\left[Z\left(\frac{i}{n}\right) - Z\left(\frac{i-1}{n}\right), Z(1)\right] - 1, \quad i = 1,2,\ldots,n. \tag{70}$$

Using the fact

$$\begin{aligned}
\sum_{i=1}^{n} c(n,i) &= 2n\sum_{i=1}^{n}\left(\mathrm{Cov}\left[Z\left(\frac{i}{n}\right), Z(1)\right] - \mathrm{Cov}\left[Z\left(\frac{i-1}{n}\right), Z(1)\right]\right) - n \qquad (71)\\
&= n^{1-2H}\sum_{i=1}^{n}\left(i^{2H} - |n-i|^{2H} - (i-1)^{2H} + |n-i+1|^{2H}\right) - n\\
&= n^{1-2H}\left(2n^{2H}\right) - n = n,
\end{aligned}$$

we get

$$\begin{aligned}
\mathrm{E}\left[S_{(1)}^2\right] &= \mathrm{Var}[Y] - \frac{a}{n^2} \qquad (72)\\
&= a\left(n^{-2H} - n^{-2}\right).
\end{aligned}$$

Similarly, for the aggregated process $Y^{(\mu)}$ we have (assuming n/μ is integer)

$$\begin{aligned} S^2_{(\mu)} &= \frac{\mu}{n}\sum_{i=1}^{n/\mu}\left(Y_i^{(\mu)} - \frac{\mu}{n}\sum_{j=1}^{n/\mu} Y_j^{(\mu)}\right)^2 && (73) \\ &= \frac{\mu}{n}\sum_{i=1}^{n/\mu}\left[X\left(\frac{i\mu}{n}\right) - X\left(\frac{(i-1)\mu}{n}\right) - \frac{\mu}{n}X(1)\right]^2 && (74) \end{aligned}$$

and

$$\begin{aligned} \mathrm{E}\left[S^2_{(\mu)}\right] &= \mathrm{Var}\left[Y^{(\mu)}\right] - a\left(\frac{\mu}{n}\right)^2 && (75) \\ &= a\left[\left(\frac{\mu}{n}\right)^{2H} - \left(\frac{\mu}{n}\right)^2\right]. \end{aligned}$$

3.4 Variance based estimation of parameters H and a

Assume that the mean rate is estimated by $\hat{m} = X(1)$. From Eq.(73) with $\mu = n/2^k$ the sums of squares [18]

$$S_k^2 = \frac{1}{2^k}\sum_{i=1}^{2^k}\left[X\left(\frac{i}{2^k}\right) - X\left(\frac{i-1}{2^k}\right) - \frac{\hat{m}}{2^k}\right]^2, \quad k = 1, 2, \ldots, \lfloor \log_2 n \rfloor \tag{76}$$

are calculated. Then the parameters H and a can be consistently estimated by, e.g.,

$$\hat{H}_k = \frac{1}{2}\log_2 \frac{2S_{k-1}^2}{S_k^2} \tag{77}$$

and

$$\hat{a}_k = S_k^2 \frac{(2^k)^{2\hat{H}_k}}{1 - (2^k)^{2\hat{H}_k - 2}}. \tag{78}$$

3.5 Exact MLE

$$h(\mathbf{y}) = (2\pi)^{-\frac{n}{2}} |\mathbf{\Sigma}|^{-\frac{1}{2}} e^{-\frac{1}{2}(\mathbf{y}-\mathbf{m})^{\mathrm{t}}\mathbf{\Sigma}^{-1}(\mathbf{y}-\mathbf{m})}, \tag{79}$$

where $\mathbf{y} = (y_1, y_2, \ldots, y_n)^{\mathrm{t}} \in \mathbb{R}^n$, $\mathbf{m} = (m/n)\mathbf{1}$, and $|\mathbf{\Sigma}|$ is the determinant of the covariance matrix

$$\mathbf{\Sigma} = \left[\mathrm{Cov}\left[Y_i, Y_j\right]\right]_{i,j=1,2,\ldots,n}. \tag{80}$$

3.5.1 MLE($\hat{m}|a, H$)

$$\hat{m} = \hat{m}(H) = \frac{\mathbf{1}^{\mathrm{t}}\,\mathbf{\Sigma}^{-1}\,\mathbf{Y}}{\mathbf{1}^{\mathrm{t}}\,\mathbf{\Sigma}^{-1}\,\mathbf{1}} \cdot n. \tag{81}$$

$$\begin{aligned} \mathrm{E}\left[\hat{m}\right] &= m, && (82) \\ \mathrm{Var}\left[\hat{m}\right] &= \frac{a}{\mathbf{1}\,\mathbf{\Sigma}_H^{-1}\,\mathbf{1}} \cdot n^2. && (83) \end{aligned}$$

3.5.2 MLE($\hat{m}, \hat{a}|H$)

$$\mathbf{\Sigma} = \mathbf{\Sigma}(a) = a\mathbf{\Sigma}_H. \tag{84}$$

$$\mathbf{\Sigma}_H = \Big[\mathrm{Cov}\left[Z(t_i) - Z(t_{i-1}), Z(t_j) - Z(t_{j-1})\right]\Big]_{i,j=1,2,\ldots,n}. \tag{85}$$

$$\hat{a}(H) = \frac{1}{n}\left(\mathbf{Y}^{\mathrm{t}}\,\mathbf{\Sigma}_H^{-1}\,\mathbf{Y} - \frac{\left(\mathbf{1}^{\mathrm{t}}\,\mathbf{\Sigma}^{-1}\,\mathbf{Y}\right)^2}{\mathbf{1}^{\mathrm{t}}\,\mathbf{\Sigma}^{-1}\,\mathbf{1}}\right). \tag{86}$$

$$\mathrm{E}\left[\hat{a}\right] = \frac{n-1}{n}a, \tag{87}$$

$$\mathrm{Var}\left[\frac{n}{n-1}\hat{a}\right] = \frac{2a^2}{n-1}. \tag{88}$$

3.5.3 MLE($\hat{m}, \hat{a}, \hat{H}$)

We have to minimize

$$\tilde{L}(\mathbf{Y}; H) = |\mathbf{\Sigma}_H|^{1/n}\left(\mathbf{Y}^{\mathrm{t}}\,\mathbf{\Sigma}_H^{-1}\,\mathbf{Y} - \frac{\left(\mathbf{1}^{\mathrm{t}}\,\mathbf{\Sigma}^{-1}\,\mathbf{Y}\right)^2}{\mathbf{1}^{\mathrm{t}}\,\mathbf{\Sigma}^{-1}\,\mathbf{1}}\right). \tag{89}$$

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

4 Descaled process

$Z(t)$ has the self-similar property $Z(\alpha t) \sim \alpha^H Z(t)$. Now consider the 'descaled' process

$$\tilde{Y}(t) \stackrel{d}{=} t^{-H} Z(t) \tag{90}$$

which has the scaling property

$$\tilde{Y}(\alpha t) \sim (\alpha t)^{-H} Z(\alpha t) = t^{-H} Z(t) = \tilde{Y}(t). \tag{91}$$

Further let us take a new time variable $u = -\log t$ and denote

$$Y(u) \stackrel{d}{=} \tilde{Y}(e^{-u}) = \tilde{Y}(t). \tag{92}$$

Now we have

$$Y(u - \log\alpha) = \tilde{Y}(e^{-u+\log\alpha}) = \tilde{Y}(\alpha e^{-u}) = \tilde{Y}(\alpha t) \sim \tilde{Y}(t) = Y(u). \tag{93}$$

Thus the process $Y(u)$ is stationary and has the following properties:

$$\mathrm{E}\left[Y(u)\right] = 0; \tag{94}$$

$$\mathrm{Var}\left[Y(u)\right] = 1; \tag{95}$$

$$\begin{aligned}\mathrm{Cov}\left[Y(u_1), Y(u_2)\right] &= \frac{1}{2}e^{H(u_2-u_1)}\left\{1 + e^{-2H(u_2-u_1)} - \left(1 - e^{-(u_2-u_1)}\right)^{2H}\right\} \\ &= g(\tau), \quad \text{where } \tau = u_2 - u_1, \quad 0 < u_1 < u_2.\end{aligned} \tag{96}$$

4.1 Descaled fractional Brownian traffic

If we 'descale' the process $X(t)$ we get

$$\tilde{W}(t) \stackrel{d}{=} t^{-H} X(t) = m t^{1-H} + \sqrt{a}\tilde{Y}(t), \tag{97}$$

and using $u = -\log t$ we finally have

$$W(u) \stackrel{d}{=} \tilde{W}(e^{-u}) = m e^{(H-1)u} + \sqrt{a} Y(u). \tag{98}$$

Thus the process $W(u)$ has the following properties:

$$\mathrm{E}\left[W(u)\right] = m e^{(H-1)u}; \tag{99}$$

$$\mathrm{Var}\left[W(u)\right] = a; \tag{100}$$

$$\mathrm{Cov}\left[W(u_1), W(u_2)\right] = \frac{a}{2} e^{H(u_2-u_1)} \left\{ 1 + e^{-2H(u_2-u_1)} - \left(1 - e^{-(u_2-u_1)}\right)^{2H} \right\}. \tag{101}$$

4.2 Geometrical sampling

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

$$t_i = \alpha^{i-1}, \quad i = 1, 2, \ldots, n, \quad 0 < \alpha < 1. \tag{102}$$

The auto-covariance matrix $\tilde{\Gamma}$ of the descaled samples $\mathbf{W} = (W(u_1), W(u_2), \ldots, W(u_n))^{\mathrm{t}}$ with $u_i = -\log t_i = (1-i)\log\alpha$ can be written as

$$\tilde{\Gamma} = \mathrm{E}\left[\mathbf{W}\mathbf{W}^{\mathrm{t}}\right] = a \cdot \mathrm{E}\left[\mathbf{Y}\mathbf{Y}^{\mathrm{t}}\right]. \tag{103}$$

Note, that our geometrical grid is now equally spaced with regards to u. Thus, if we use the notation $Y_i = Y(u_i)$ the process $Y = (Y_1, Y_2, \ldots, Y_n)$ is a stationary process in discrete time with zero mean and unit variance and its auto-correlation function $\rho(k)$ can be defined as

$$\rho(i-j) = \mathrm{Cov}\left[Y_i, Y_j\right], \quad i, j = 1, 2, \ldots \tag{104}$$

and thus

$$\tilde{\Gamma}_{ij} = a\rho(i-j), \quad i, j = 1, 2, \ldots, n. \tag{105}$$

From Eq.(94) and Eq.(102) we get

$$\begin{aligned} \rho(i-j) &= \frac{1}{2} e^{-H|i-j|\log\alpha} \left\{ 1 + e^{2H|i-j|\log\alpha} - \left(1 - e^{|i-j|\log\alpha}\right)^{2H} \right\} \\ &= \frac{1}{2} \alpha^{-H|i-j|} \left\{ 1 + \alpha^{2H|i-j|} - \left(1 - \alpha^{|i-j|}\right)^{2H} \right\}. \end{aligned} \tag{106}$$

With the notations $g(x) = (1 + x^{2H} - (1-x)^{2H})/2$ and $\beta = \alpha^{-H}$ the matrix $\tilde{\Gamma}$ has the following structure:

$$\tilde{\Gamma} = a \cdot \begin{pmatrix} 1 & \beta g(\alpha) & \beta^2 g(\alpha^2) & \cdots & \beta^{n-1} g(\alpha^{n-1}) \\ \beta g(\alpha) & 1 & \beta g(\alpha) & \ddots & \vdots \\ \beta^2 g(\alpha^2) & \beta g(\alpha) & 1 & \ddots & \beta^2 g(\alpha^2) \\ \vdots & \ddots & \ddots & \ddots & \beta g(\alpha) \\ \beta^{n-1} g(\alpha^{n-1}) & \cdots & \beta^2 g(\alpha^2) & \beta g(\alpha) & 1 \end{pmatrix}. \tag{107}$$

4.3 Descaled MLE

When doing the maximum likelihood estimation of the model parameters m, a and H, the calculation of the inverse and determinant of the covariance matrix Γ can be problematic. To ease the calculations, 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 proceed as

$$\begin{aligned} \tilde{\Gamma} &= \mathrm{E}\left[\mathbf{W}\mathbf{W}^{\mathrm{t}}\right] \qquad (108) \\ &= \mathrm{E}\left[(\mathbf{D}\mathbf{X})(\mathbf{D}\mathbf{X})^{\mathrm{t}}\right] \\ &= \mathbf{D}\,\mathrm{E}\left[\mathbf{X}\mathbf{X}^{\mathrm{t}}\right]\mathbf{D} \\ &= \mathbf{D}\Gamma\mathbf{D}. \end{aligned}$$

From this one can easily deduce that $\Gamma^{-1} = \mathbf{D}\tilde{\Gamma}^{-1}\mathbf{D}$.

4.3.1 MLE($\hat{m}|a, H$)

The MLE $\hat{m}$ is the solution of

$$\hat{m}(H) = \frac{\mathbf{t}^{\mathrm{t}}\,\mathbf{D}\tilde{\Gamma}^{-1}\mathbf{D}\,\mathbf{X}}{\mathbf{t}^{\mathrm{t}}\,\mathbf{D}\tilde{\Gamma}^{-1}\mathbf{D}\,\mathbf{t}}. \tag{109}$$

4.3.2 MLE($\hat{m}|\hat{a}, H$)

Similarly, the same result holds when estimating a as in Eq.(21) with $\Gamma_H^{-1} = \mathbf{D}\tilde{\Gamma}_H^{-1}\mathbf{D}$, namely,

$$\begin{aligned} \hat{a}(H) &= \frac{1}{n}(\mathbf{X} - \hat{m}\mathbf{t})^{\mathrm{t}}\,\mathbf{D}\tilde{\Gamma}_H^{-1}\mathbf{D}\,(\mathbf{X} - \hat{m}\mathbf{t}) \qquad (110) \\ &= \frac{1}{n}\frac{(\mathbf{X}^{\mathrm{t}}\,\mathbf{D}\tilde{\Gamma}_H^{-1}\mathbf{D}\,\mathbf{X})(\mathbf{t}^{\mathrm{t}}\,\mathbf{D}\tilde{\Gamma}_H^{-1}\mathbf{D}\,\mathbf{t}) - (\mathbf{t}^{\mathrm{t}}\,\mathbf{D}\tilde{\Gamma}_H^{-1}\mathbf{D}\,\mathbf{X})^2}{\mathbf{t}^{\mathrm{t}}\,\mathbf{D}\tilde{\Gamma}_H^{-1}\mathbf{D}\,\mathbf{t}}. \end{aligned}$$

4.3.3 MLE($\hat{m}, \hat{a}, \hat{H}$)

Finally, we have to minimize

$$\tilde{L}(\mathbf{X}; H) = |\Gamma_H|^{1/n}\frac{(\mathbf{X}^{\mathrm{t}}\,\mathbf{D}\tilde{\Gamma}_H^{-1}\mathbf{D}\,\mathbf{X})(\mathbf{t}^{\mathrm{t}}\,\mathbf{D}\tilde{\Gamma}_H^{-1}\mathbf{D}\,\mathbf{t}) - (\mathbf{t}^{\mathrm{t}}\,\mathbf{D}\tilde{\Gamma}_H^{-1}\mathbf{D}\,\mathbf{X})^2}{\mathbf{t}^{\mathrm{t}}\,\mathbf{D}\tilde{\Gamma}_H^{-1}\mathbf{D}\,\mathbf{t}}. \tag{111}$$

The determinant $|\Gamma_H|$ can be also calculated as

$$\begin{aligned}
|\Gamma_H| &= |\mathbf{D}^{-1}\tilde{\Gamma}_H\mathbf{D}^{-1}| \\
&= \left(\prod_{i=1}^{n} t_i^H\right)|\tilde{\Gamma}_H\mathbf{D}^{-1}| \\
&= \left(\prod_{i=1}^{n} t_i^{2H}\right)|\tilde{\Gamma}_H| \\
&= \left(\prod_{i=1}^{n} \alpha^{2H(i-1)}\right)|\tilde{\Gamma}_H| \\
&= \alpha^{Hn(n-1)}|\tilde{\Gamma}_H|
\end{aligned} \tag{112}$$

The minimum of Eq.(111) is obtained for some value $\hat{H}$ which is the MLE estimate; the corresponding MLE estimates for m and a are $\hat{m} = m(\hat{H})$ and $\hat{a} = a(\hat{H})$.

5 Robustness

5.1 Additive Gaussian white noise model

$$X^*(t) = mt + \sqrt{a}Z(t) + N(t), \tag{113}$$

where $N(t)$ is a zero mean Gaussian white noise with variance σ_N^2, i.e.,

$$\begin{aligned}
\mathrm{E}\,[N(t)] &= 0; && (114) \\
\mathrm{Var}\,[N(t)] &= \sigma_N^2; && (115) \\
\mathrm{Cov}\,[N(t_1), N(t_2)] &= 0, \quad t_1 \neq t_2. && (116)
\end{aligned}$$

Let $\mathbf{X}^* = (X^*(t_1), \ldots, X^*(t_n))^{\mathrm{t}}$ be the observed noisy traffic values at n instances forming the vector $\mathbf{t} = (t_1, \ldots, t_n)^{\mathrm{t}}$. Since $X(t)$ and $N(t)$ are independent Gaussian processes (which means that their convolutions are also Gaussian), the joint distribution of $\mathbf{X}^*$ can be written as

$$h(\mathbf{x}) = (2\pi)^{-\frac{n}{2}}|\Gamma^*|^{-\frac{1}{2}}e^{-\frac{1}{2}(\mathbf{x}-m\mathbf{t})^{\mathrm{t}}\Gamma^{*-1}(\mathbf{x}-m\mathbf{t})}, \quad \mathbf{x} \in \mathbb{R}^n, \tag{117}$$

with

$$\Gamma^* = \Gamma + \sigma_N^2\mathbf{I}, \tag{118}$$

where $\mathbf{I}$ is the identity matrix.

Note, that this noise model is rather unrealistic at low time scales, and the MLE estimate of H is strongly biased.

5.2 Additive Gaussian noise

$$X^*(t) = mt + \sqrt{a}Z(t) + \sigma_N B(t), \tag{119}$$

where $B(t)$ is the normalized Brownian motion with zero mean and unit variance. Since $X(t)$ and $N(t)$ are independent Gaussian processes the joint distribution of $\mathbf{X}^*$ can be written as

$$h(\mathbf{x}) = (2\pi)^{-\frac{n}{2}} |,^*|^{-\frac{1}{2}} e^{-\frac{1}{2}(\mathbf{x}-m\mathbf{t})^t}, ^{*-1}(\mathbf{x} - m\mathbf{t}), \quad \mathbf{x} \in \mathbb{R}^n, \tag{120}$$

with

$$,^* = a,_H + \sigma_N^2,_{0.5}, \tag{121}$$

where

$$,_{0.5} = [\min(t_i, t_j)]_{i,j=1,\ldots,n}. \tag{122}$$

6 Numerical approximations

...to calculate $|\tilde{,}|$ and $\tilde{,}^{-1}$.

6.1 Linear approximation of $g(x)$

We use the approximation

$$g(x) = \frac{1}{2}(1 + x^{2H} - (1-x)^{2H}) \approx x, \tag{123}$$

6.1.1 Approximations for $\rho(k)$ and $f(\lambda)$

The autocorrelation function of Y can be approximated as

$$\begin{aligned} \rho(k) &= \frac{1}{2}\alpha^{-H|k|}\left\{1 + \alpha^{2H|k|} - (1-\alpha^{|k|})^{2H}\right\} \\ &\approx \alpha^{(1-H)|k|}. \end{aligned} \tag{124}$$

With the notation $\gamma = \alpha^{1-H}$ we get

$$\begin{aligned} f(\lambda; H) &= \frac{1}{2\pi}\sum_{k=-\infty}^{\infty} \gamma^{|k|} e^{ik\lambda} \\ &= \frac{1}{2\pi}\left[1 + \sum_{k=1}^{\infty}\left(\gamma e^{i\lambda}\right)^k + \sum_{k=1}^{\infty}\left(\gamma e^{-i\lambda}\right)^k\right] \\ &= \frac{1}{2\pi}\left[1 + \frac{\gamma e^{i\lambda}}{1-\gamma e^{i\lambda}} + \frac{\gamma e^{-i\lambda}}{1-\gamma e^{-i\lambda}}\right] \\ &= \frac{1}{2\pi}\left[1 + \frac{2\gamma\cos\lambda - 2\gamma^2}{1 - 2\gamma\cos\lambda + \gamma^2}\right] \\ &= \frac{1}{2\pi}\frac{1-\gamma^2}{1-2\gamma\cos\lambda+\gamma^2}. \end{aligned} \tag{125}$$

6.1.2 Approximations for $\tilde{,}^{-1}$ and $|\tilde{,}|$

Using the approximation Eq.(123), from Eq.(107) $\tilde{,}$ can be approximated as $\tilde{,} \approx a\mathbf{R}$, where $\mathbf{R}$ is a Toeplitz-type matrix of the form (with $\gamma = \alpha^{1-H}$)

$$\mathbf{R} = \begin{pmatrix} 1 & \gamma & \gamma^2 & \cdots & \gamma^{n-1} \\ \gamma & 1 & \gamma & \ddots & \vdots \\ \gamma^2 & \gamma & 1 & \ddots & \gamma^2 \\ \vdots & \ddots & \ddots & \ddots & \gamma \\ \gamma^{n-1} & \cdots & \gamma^2 & \gamma & 1 \end{pmatrix}. \tag{126}$$

The inverse of $\mathbf{R}$ can be easily calculated as [19]

$$\mathbf{R}^{-1} = \frac{1}{\frac{1}{\gamma} - \gamma} \begin{pmatrix} \frac{1}{\gamma} & -1 & 0 & \cdots & 0 \\ -1 & \gamma + \frac{1}{\gamma} & -1 & \ddots & \vdots \\ 0 & -1 & \gamma + \frac{1}{\gamma} & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & -1 \\ 0 & \cdots & 0 & -1 & \frac{1}{\gamma} \end{pmatrix}, \tag{127}$$

and the determinant of $\mathbf{R}$ is given by [19]

$$|\mathbf{R}| = (-1)^{n-1} \prod_{i=1}^{n-1} \begin{vmatrix} \gamma^{1-i} & \gamma^{i-1} \\ \gamma^{-i} & \gamma^{i} \end{vmatrix} \cdot \gamma^{n-1} = (1-\gamma^2)^{n-1}. \tag{128}$$

6.1.3 Approximate MLE

The approximate MLE $\hat{m}$ is the solution of

$$\hat{m}(H) \approx \frac{\mathbf{t}^{\mathrm{t}}\,\mathbf{D}\mathbf{R}^{-1}\mathbf{D}\,\mathbf{X}}{\mathbf{t}^{\mathrm{t}}\,\mathbf{D}\mathbf{R}^{-1}\mathbf{D}\,\mathbf{t}}. \tag{129}$$

Using the fact that $\mathbf{t}^{\mathrm{t}}\,\mathbf{D}\mathbf{R}^{-1}\mathbf{D}\,\mathbf{t} = 1$ and $\mathbf{t}^{\mathrm{t}}\,\mathbf{D}\mathbf{R}^{-1}\mathbf{D} = (1, 0, \ldots, 0)$, we get

$$\hat{m}(H) = X_1. \tag{130}$$

Similarly,

$$\hat{a}(H) = \frac{1}{n}\left(\mathbf{X}^{\mathrm{t}}\,\mathbf{D}\,\mathbf{R}^{-1}\mathbf{D}\,\mathbf{X} - X_1^2\right). \tag{131}$$

Finally, we have to minimize

$$L(\mathbf{X}; H) = \frac{n-1}{n}\log\left(\alpha^{nH}(1-\alpha^{2-2H})\right) + \log\left(\mathbf{X}^{\mathrm{t}}\,\mathbf{D}\,\mathbf{R}^{-1}\mathbf{D}\,\mathbf{X} - X_1^2\right). \tag{132}$$

Note, that the term $\mathbf{X}^{\mathrm{t}}\,\mathbf{D}\,\mathbf{R}^{-1}\mathbf{D}\,\mathbf{X}$ can be calculated as

$$\begin{aligned}
\mathbf{X}^{\mathrm{t}}\,\mathbf{D}\,\mathbf{R}^{-1}\mathbf{D}\,\mathbf{X} &= (\mathbf{D}\,\mathbf{X})^{\mathrm{t}}\,\mathbf{R}^{-1}(\mathbf{D}\,\mathbf{X}) \\
&= 2\sum_{i=1}^{n-1}\frac{\gamma}{\gamma^2-1}\left(t_i^{-H}X_i\right)\left(t_{i+1}^{-H}X_{i+1}\right)+\frac{X_1^2+\left(t_n^{-H}X_n\right)^2}{1-\gamma^2}+\sum_{i=2}^{n-1}\frac{1+\gamma^2}{1-\gamma^2}\left(t_i^{-H}X_i\right)^2 \\
&= 2\sum_{i=1}^{n-1}\frac{\gamma\alpha^{-H(2i-1)}}{\gamma^2-1}X_iX_{i+1}+\frac{X_1^2+\alpha^{-2H(n-1)}X_n^2}{1-\gamma^2}+\sum_{i=2}^{n-1}\frac{(1+\gamma^2)\alpha^{-2H(i-1)}}{1-\gamma^2}X_i^2
\end{aligned} \tag{133}$$

6.2 Parameter fitting—Approximation for $\tilde{\Gamma}^{-1}$

$\tilde{\Gamma}_H^{-1} \approx \mathbf{C}$ so that

$$\mathbf{C} = \begin{pmatrix}
\hat{\mathbf{c}}_{\mathbf{1}} & \hat{\mathbf{c}}_{\mathbf{2}} & \hat{\mathbf{c}}_{\mathbf{3}} & \cdots & \hat{\mathbf{c}}_{\mathbf{p-1}} & c_p & 0 & \cdots & 0 \\
\hat{\mathbf{c}}_{\mathbf{2}} & \hat{\mathbf{c}}_{\mathbf{1}} & c_2 & \cdots & c_{p-2} & c_{p-1} & c_p & \ddots & \vdots \\
\hat{\mathbf{c}}_{\mathbf{3}} & c_2 & c_1 & \cdots & c_{p-3} & c_{p-2} & c_{p-1} & \ddots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & c_p \\
\hat{\mathbf{c}}_{\mathbf{p-1}} & c_{p-2} & c_{p-3} & \cdots & c_1 & c_2 & c_3 & & \hat{\mathbf{c}}_{\mathbf{p-1}} \\
c_p & c_{p-1} & c_{p-2} & \cdots & c_2 & c_1 & c_2 & \ddots & \vdots \\
0 & c_p & c_{p-1} & \cdots & c_3 & c_2 & c_1 & \ddots & \hat{\mathbf{c}}_{\mathbf{3}} \\
\vdots & \ddots & \ddots & \ddots & & \ddots & \ddots & \ddots & \hat{\mathbf{c}}_{\mathbf{2}} \\
0 & \cdots & 0 & c_p & \hat{\mathbf{c}}_{\mathbf{p-1}} & \cdots & \hat{\mathbf{c}}_{\mathbf{3}} & \hat{\mathbf{c}}_{\mathbf{2}} & \hat{\mathbf{c}}_{\mathbf{1}}
\end{pmatrix}. \tag{134}$$

With

$$\tilde{\Gamma} = \left[\left(\tilde{\Gamma}_H\right)_{ij}\right]_{i,j=1,\ldots,2p-1} \tag{135}$$

and from $\mathbf{C}\tilde{\Gamma}_H \approx \mathbf{E}$ we get

$$(c_p,\ldots,c_2,c_1,c_2,\ldots,c_p)\cdot\tilde{\Gamma} = \mathbf{e_p}, \tag{136}$$

and from this we have

$$c_i = \tilde{\Gamma}^{-1}_{p(p+i-1)}, \quad i=1,2,\ldots,p. \tag{137}$$

6.3 Approximations from Whittle's approximate MLE

Suppose that $\mathbf{\Sigma}$ is the covariance matrix of a stationary Gaussian process. Let its spectral density $f(\lambda)$ be characterized by an unknown finite dimensional parameter vector

$$\boldsymbol{\theta}^o = (\sigma_o^2, H^o, \theta_3^o, \ldots, \theta_M^o). \tag{138}$$

Thus, we assume that the spectral density comes from a parametric family of densities $f(\lambda) = f(\lambda;\boldsymbol{\theta})$, where $\boldsymbol{\theta} \in \mathbb{R}^M$.

Approximation of $\log|\boldsymbol{\Sigma}(\boldsymbol{\theta}^o)|$**:** It was shown by Grenander and Szegő [10] that

$$\lim_{n\to\infty}\frac{1}{n}\log|\boldsymbol{\Sigma}(\boldsymbol{\theta})| = \frac{1}{2\pi}\int_{-\pi}^{\pi}\log f(\lambda;\boldsymbol{\theta})\,d\lambda. \tag{139}$$

Therefore, we replace $\log|\boldsymbol{\Sigma}(\boldsymbol{\theta})|$ by

$$\log|\boldsymbol{\Sigma}(\boldsymbol{\theta})| \approx \frac{n}{2\pi}\int_{-\pi}^{\pi}\log f(\lambda;\boldsymbol{\theta})\,d\lambda. \tag{140}$$

Approximation of $\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})$**:** The matrix $\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})$ can be replaced by a matrix whose elements are easier to calculate. Define

$$\mathbf{S}(\boldsymbol{\theta}) = [s(j-l)]_{j,l=1,\dots,n} \tag{141}$$

to be the $n\times n$-matrix with elements

$$s(j-l) = \frac{1}{(2\pi)^2}\int_{-\pi}^{\pi}\frac{1}{f(\lambda;\boldsymbol{\theta})}e^{i(j-l)\lambda}\,d\lambda. \tag{142}$$

The matrix $\mathbf{S}$ is asymptotically the inverse of $\boldsymbol{\Sigma}$.

The process Y is a stationary Gaussian process with zero mean, unit variance and covariance matrix $\tilde{\Gamma}_H$. The spectral density of the process can be written as

$$\begin{aligned} f(\lambda;H) &= \frac{1}{2\pi}\sum_{k=-\infty}^{\infty}\frac{1}{2}\alpha^{-H|k|}\left\{1+\alpha^{2H|k|}-(1-\alpha^{|k|})^{2H}\right\}e^{ik\lambda} \\ &= \frac{1}{2\pi}\left[1+\sum_{k=1}^{\infty}\frac{1}{2}\alpha^{-Hk}\left\{1+\alpha^{2Hk}-(1-\alpha^{k})^{2H}\right\}\left(e^{ik\lambda}+e^{-ik\lambda}\right)\right] \\ &\approx \frac{1}{2\pi}\left[1+\sum_{k=1}^{M}\alpha^{-Hk}\left\{1+\alpha^{2Hk}-(1-\alpha^{k})^{2H}\right\}\cos(k\lambda)\right] \end{aligned} \tag{143}$$

with some M large enough. If we approximate the integrals in Eq.(140) and Eq.(142) with a simple Riemann sum with

$$\lambda_{j,m} = \frac{2\pi j}{m}, \quad j = 1,2,\dots,\left\lfloor\frac{m-1}{2}\right\rfloor, \tag{144}$$

we get the following approximations:

$$\begin{aligned} \log|\tilde{\Gamma}_H| &\approx \frac{n}{2\pi}\int_{-\pi}^{\pi}\log f(\lambda)\,d\lambda \\ &\approx \frac{n}{2\pi}\left[\log f(0)\frac{2\pi}{m}+2\sum_{j=1}^{\lfloor\frac{m-1}{2}\rfloor}\log f(\lambda_{j,m})\frac{2\pi}{m}\right] \\ &= \frac{n}{m}\log f(0)+\frac{2n}{m}\sum_{j=1}^{\lfloor\frac{m-1}{2}\rfloor}\log f\left(\frac{2\pi j}{m}\right) \end{aligned} \tag{145}$$

and

$$\tilde{\Gamma}_H^{-1} \approx \mathbf{S}(H) = [s(j-l)]_{j,l=1,\ldots,n}, \tag{146}$$

where

$$\begin{aligned} s(k) &= \frac{1}{(2\pi)^2} \int_{-\pi}^{\pi} \frac{1}{f(\lambda)} e^{ik\lambda}\, d\lambda \qquad (147) \\ &\approx \frac{1}{(2\pi)^2} \left[\frac{1}{f(0)} \frac{2\pi}{m} + \sum_{j=1}^{\lfloor \frac{m-1}{2} \rfloor} \frac{1}{f(\lambda_{j,m})} \left(e^{ik\lambda_{j,m}} + e^{-ik\lambda_{j,m}} \right) \frac{2\pi}{m} \right] \\ &= \frac{1}{2\pi m f(0)} + \frac{1}{\pi m} \sum_{j=1}^{\lfloor \frac{m-1}{2} \rfloor} \frac{1}{f\left(\frac{2\pi j}{m}\right)} \cos \frac{2\pi j k}{m}. \end{aligned}$$

6.4 Linear approximation with Whittle

$$\begin{aligned} \log |\tilde{\Gamma}_H| &\approx \frac{n}{2\pi} \int_{-\pi}^{\pi} \log f(\lambda) d\lambda \qquad (148) \\ &\approx \frac{n}{2\pi} \int_{-\pi}^{\pi} \frac{1}{2\pi} \frac{1-\gamma^2}{1-2\gamma\cos\lambda+\gamma^2} d\lambda \\ &= \frac{n}{2\pi} \left[2\pi \log \frac{1-\gamma}{2\pi(1+\gamma)} + 2\gamma \int_{-\pi}^{\pi} \frac{\lambda \sin\lambda}{1-2\gamma\cos\lambda+\gamma^2} d\lambda \right] \\ &= n \log \frac{1-\gamma}{2\pi(1+\gamma)} + \frac{n\gamma}{\pi} \int_{-\pi}^{\pi} \frac{\lambda\sin\lambda}{1-2\gamma\cos\lambda+\gamma^2} d\lambda. \end{aligned}$$

If n is large enough, $\gamma \approx 1$.

$$\begin{aligned} \log |\tilde{\Gamma}_H| &\approx n \left[\log(1-\gamma) - \log(2\pi(1+\gamma)) + \frac{\gamma}{\pi} \int_{-\pi}^{\pi} \frac{\lambda\sin\lambda}{1-2\gamma\cos\lambda+\gamma^2} d\lambda \right] \qquad (149) \\ &\approx n \left[\log(1-\gamma) - \log(2\pi \cdot 2) + \frac{1}{\pi} \int_{-\pi}^{\pi} \frac{\lambda\sin\lambda}{2-2\cos\lambda} d\lambda \right] \\ &\approx n\left[\log(1-\gamma) - 1.145\right] \end{aligned}$$

From Eq.(128) we get

$$\begin{aligned} \log |\tilde{\Gamma}_H| &= \log(1-\gamma^2)^{n-1} \qquad (150) \\ &= (n-1)\left[\log(1-\gamma) + \log(1+\gamma)\right] \\ &\approx n\left[\log(1-\gamma) + \log 2\right] \\ &\approx n\left[\log(1-\gamma) + 0.693\right] \end{aligned}$$

$$s(k) = \frac{1}{(2\pi)^2} \int_{-\pi}^{\pi} \frac{1}{f(\lambda)} e^{ik\lambda} d\lambda \tag{151}$$

$$
\begin{aligned}
&\approx \frac{1}{(2\pi)^2}\int_{-\pi}^{\pi} 2\pi \frac{1-2\gamma\cos\lambda+\gamma^2}{1-\gamma^2} e^{ik\lambda} d\lambda \\
&= \begin{cases} \dfrac{1+\gamma^2}{1-\gamma^2}, & k=0 \\ \dfrac{\gamma}{\gamma^2-1}, & k=1 \\ 0, & k\geq 2 \end{cases}
\end{aligned}
$$

$$
\mathbf{S} = \frac{1}{\frac{1}{\gamma}-\gamma}\begin{pmatrix} \gamma+\frac{1}{\gamma} & -1 & 0 & \cdots & 0 \\ -1 & \gamma+\frac{1}{\gamma} & -1 & \ddots & \vdots \\ 0 & -1 & \gamma+\frac{1}{\gamma} & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & -1 \\ 0 & \cdots & 0 & -1 & \gamma+\frac{1}{\gamma} \end{pmatrix}, \tag{152}
$$

A Related math for inversion of band matrices

A.1 General case

$$
\begin{matrix} r) \\ q) \end{matrix}\left[\begin{array}{c|c} \mathbf{A} & \mathbf{B} \\ \hline \mathbf{C} & \mathbf{D} \end{array}\right]^{-1} = \left[\begin{array}{c|c} \mathbf{U} & \mathbf{L} \\ \hline \mathbf{S} & \mathbf{V} \end{array}\right]\begin{matrix} (q \\ (r \end{matrix} \tag{153}
$$

$$
|\mathbf{C}| = (-1)^{rq}\begin{vmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{vmatrix} \cdot |\mathbf{S}| \tag{154}
$$

$$
\mathbf{C} = \begin{pmatrix} c_1^{(1)} & c_1^{(2)} & \cdots & c_1^{(p)} & 0 & \cdots & 0 \\ c_1^{(2)} & c_2^{(1)} & \cdots & c_2^{(p-1)} & c_2^{(p)} & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & 0 \\ c_1^{(p)} & c_2^{(p-1)} & \cdots & c_p^{(1)} & c_p^{(2)} & \ddots & c_{n-p+1}^{(p)} \\ 0 & c_2^{(p)} & \cdots & c_p^{(2)} & c_{p+1}^{(1)} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & c_{n-1}^{(2)} \\ 0 & \cdots & 0 & c_{n-p+1}^{(p)} & \cdots & c_{n-1}^{(2)} & c_n^{(1)} \end{pmatrix}. \tag{155}
$$

$$\mathbf{A}_{(p-1)\times(n+p-1)} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots & & \vdots \\ 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \end{pmatrix}, \quad \mathbf{D}_{(n+p-1)\times(p-1)} = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix}. \tag{156}$$

$$\begin{vmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{vmatrix} = \prod_{i=1}^{n-p+1} c_i^{(p)}. \tag{157}$$

$$|\mathbf{C}| = (-1)^{n(p-1)} \cdot |\mathbf{S}| \cdot \prod_{i=1}^{n-p+1} c_i^{(p)}. \tag{158}$$

A.2 $p = 2$ case

Let

$$\mathbf{C} = \begin{pmatrix} a_1 & b_1 & 0 & \cdots & 0 \\ b_1 & a_2 & b_2 & \ddots & \vdots \\ 0 & b_2 & a_3 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & b_{n-1} \\ 0 & \cdots & 0 & b_{n-1} & a_n \end{pmatrix} \tag{159}$$

with $b_i \neq 0$.

$$\left(\begin{array}{cccccc|c} 1 & 0 & 0 & \cdots & 0 & & 0 \\ \hline a_1 & b_1 & 0 & \cdots & 0 & & 0 \\ b_1 & a_2 & b_2 & \ddots & \vdots & & \vdots \\ 0 & b_2 & a_3 & \ddots & 0 & & 0 \\ \vdots & \ddots & \ddots & \ddots & b_{n-1} & & 0 \\ 0 & \cdots & 0 & b_{n-1} & a_n & & 1 \end{array}\right)^{-1} = \left(\begin{array}{c|cccc} u_1 & 0 & \cdots & 0 & 0 \\ u_2 & l_{21} & \ddots & \vdots & \vdots \\ \vdots & \vdots & \ddots & 0 & 0 \\ u_n & l_{n1} & \cdots & l_{n(n-1)} & 0 \\ \hline s & v_1 & \cdots & v_{n-1} & v_n \end{array}\right) \tag{160}$$

$$\begin{pmatrix} \mathbf{e}_1^{\mathrm{t}} & 0 \\ \mathbf{C} & \mathbf{e}_n \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{u} & \mathbf{L} \\ s & \mathbf{v}^{\mathrm{t}} \end{pmatrix}. \tag{161}$$

s can be calculated using the following recursive equations:

$$\begin{aligned} u_1 &= 1, \\ u_2 &= -\frac{1}{b_1} a_1, \end{aligned} \tag{162}$$

$$
\begin{aligned}
u_{i+1} &= -\frac{1}{b_i}(a_i u_i + b_{i-1} u_{i-1}), \quad i = 2, 3, \ldots, n-1, \\
s &= a_n u_n + b_{n-1} u_{n-1}.
\end{aligned}
$$

The determinant of $\mathbf{C}$ can be calculated as

$$
|\mathbf{C}| = (-1)^n \cdot s \cdot \prod_i^{n-1} b_i \approx \prod_i^{n-1} |b_i|. \tag{163}
$$

A.3 $p = 3$ case

$$
\left(\begin{array}{cccccc|cc}
1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\
\hline
a_1 & b_1 & c_1 & 0 & \cdots & 0 & 0 & 0 \\
b_1 & a_2 & b_2 & c_2 & \ddots & \vdots & \vdots & \vdots \\
c_1 & b_2 & a_3 & b_3 & \ddots & 0 & 0 & 0 \\
0 & c_2 & b_3 & a_4 & \ddots & c_{n-2} & 0 & 0 \\
\vdots & \ddots & \ddots & \ddots & \ddots & b_{n-1} & 1 & 0 \\
0 & \cdots & 0 & c_{n-2} & b_{n-1} & a_n & 0 & 1
\end{array}\right)^{-1}
=
\left(\begin{array}{cc|ccccc}
u_{11} & 0 & 0 & \cdots & 0 & 0 & 0 \\
u_{21} & u_{22} & 0 & \cdots & 0 & 0 & 0 \\
u_{31} & u_{32} & l_{31} & \ddots & \vdots & \vdots & \vdots \\
\vdots & \vdots & \vdots & \ddots & 0 & 0 & 0 \\
u_{n1} & u_{n2} & l_{n1} & \cdots & l_{n(n-2)} & 0 & 0 \\
\hline
s_{11} & s_{12} & v_{11} & \cdots & v_{1(n-2)} & v_{1(n-1)} & 0 \\
s_{21} & s_{22} & v_{21} & \cdots & v_{2(n-2)} & v_{2(n-1)} & v_{2n}
\end{array}\right) \tag{164}
$$

$$
\begin{aligned}
u_{11} &= u_{22} = 1 \qquad (165) \\
u_{12} &= u_{21} = 0 \\
u_{3j} &= -\frac{1}{c_1}(a_1 u_{1j} + b_1 u_{2j}), \quad j = 1, 2 \\
u_{4j} &= -\frac{1}{c_2}(b_1 u_{1j} + a_2 u_{2j} + b_2 u_{3j}), \quad j = 1, 2 \\
u_{(i+1)j} &= -\frac{1}{c_{i-1}}(c_{i-3} u_{(i-3)j} + b_{i-2} u_{(i-2)j} + a_{i-1} u_{(i-1)j} + b_{i-1} u_{ij}), \quad i = 4, 5, \ldots, n-1, \quad j = 1, 2 \\
s_{1j} &= -(b_{n-1} u_{nj} + a_{n-1} u_{(n-1)j} + b_{n-2} u_{(n-2)j} + c_{n-3} u_{(n-3)j}), \quad j = 1, 2 \\
s_{2j} &= -(a_n u_{nj} + b_{n-1} u_{(n-1)j} + c_{n-2} u_{(n-2)j}), \quad j = 1, 2
\end{aligned}
$$

$$
|\mathbf{C}| = c_1 c_2 \cdots c_{n-2} |\mathbf{S}|. \tag{166}
$$

B Toeplitz type matrices

Consider the matrix , of the form (thus , is Toeplitz)

$$
, = \begin{pmatrix}
\gamma_0 & \gamma_1 & \gamma_2 & \cdots & \gamma_{n-1} \\
\gamma_1 & \gamma_0 & \gamma_1 & \ddots & \vdots \\
\gamma_2 & \gamma_1 & \gamma_0 & \ddots & \gamma_2 \\
\vdots & \ddots & \ddots & \ddots & \gamma_1 \\
\gamma_{n-1} & \cdots & \gamma_2 & \gamma_1 & \gamma_0
\end{pmatrix}. \tag{167}
$$

B.1 Inversion

For any vector $\mathbf{a} \in \mathbb{R}^n$ let $\mathbf{L}(\mathbf{a})$ denote the $n \times n$ lower triangular Toeplitz matrix with first column equal to $\mathbf{a}$.

Lemma B.1 [8] Let $\mathbf{r} = (\gamma_0, \gamma_1, \ldots, \gamma_{n-1})^{\mathrm{t}}$ be the first column of , and $\mathbf{s} = (0, \gamma_1, \ldots, \gamma_{n-1})^{\mathrm{t}}$. Then,

$$, = \frac{1}{\gamma_0}\left(\mathbf{L}(\mathbf{r})^{\mathrm{t}}\mathbf{L}(\mathbf{r}) - \mathbf{L}(\mathbf{s})^{\mathrm{t}}\mathbf{L}(\mathbf{s})\right). \tag{168}$$

Lemma B.2 [8] Let $\mathbf{u} = (u_1, \ldots, u_n)^{\mathrm{t}}$ be the first column of $,^{-1}$ and $\mathbf{v} = (0, u_n, u_{n-1}, \ldots, u_2)^{\mathrm{t}}$. Then $,^{-1}$ has the representation

$$,^{-1} = \frac{1}{u_1}\left(\mathbf{L}(\mathbf{u})^{\mathrm{t}}\mathbf{L}(\mathbf{u}) - \mathbf{L}(\mathbf{v})^{\mathrm{t}}\mathbf{L}(\mathbf{v})\right). \tag{169}$$

The usefulness of Eq.(169) comes from the fact that $,^{-1}$ is completely determined by its first column $\mathbf{u}$. To get $\mathbf{u}$ we have to solve the following equation

$$, \cdot \mathbf{u} = (1, 0, \ldots, 0)^{\mathrm{t}} \tag{170}$$

Using the Levinson algorithm [21] we first solve

$$, \cdot \mathbf{w} = (\alpha_n, 0, \ldots, 0)^{\mathrm{t}} \tag{171}$$

and from that we have $\mathbf{u} = \mathbf{w}/\alpha_n$. With $\mathbf{w}^{(i)} = (w_1, \ldots, w_i)^{\mathrm{t}}$ the vector $\mathbf{w}$ can be determined iteratively as $\mathbf{w} \equiv \mathbf{w}^{(n)}$ in $\mathcal{O}(n^2)$ flops[1]. The algorithm is summarized below.

$\mathbf{w}^{(1)} \equiv w_1 = 1, \quad \alpha_1 = \gamma_0$

for $i = 1, \ldots, n-1$

$$\begin{aligned} \beta_i &= (\gamma_{i+1}, \ldots, \gamma_2) \cdot \mathbf{w}^{(i)} \\ K_i &= -\beta_i/\alpha_i \\ \mathbf{w}^{(i+1)} &= (w_1, \ldots, w_i, 0)^{\mathrm{t}} + K_i \cdot (0, w_i, \ldots, w_1)^{\mathrm{t}} \\ \alpha_{i+1} &= \alpha_i + K_i\beta_i \end{aligned}$$

B.2 Determinant

Let $,_k$ denote the upper-left $k \times k$ block of the matrix , for $k = 1, \ldots, n-1$, and $\boldsymbol{\gamma}_k$ denote the vector $(\gamma_1, \ldots, \gamma_k)^{\mathrm{t}}$. Then [8]

$$,_{k+1} = \begin{bmatrix} \gamma_0 & \boldsymbol{\gamma}_k^{\mathrm{t}} \\ \boldsymbol{\gamma}_k & ,_k \end{bmatrix}, \tag{172}$$

so that using the formula for the determinant of a block matrix ([19], p 289)

$$\begin{vmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{vmatrix} = |\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}| \cdot |\mathbf{D}|, \tag{173}$$

[1] $\mathbf{u}$ could be computed only in $\mathcal{O}(n \log^2 n)$ flops [8].

when $|\mathbf{D}| \neq 0$, we have

$$|,_{k+1}| = |,_k|(\gamma_0 - \boldsymbol{\gamma}_k^{\mathrm{t}} ,_k^{-1} \boldsymbol{\gamma}_k), \quad k = 1, \ldots, n-1. \tag{174}$$

($|,|$ can be obtained in $\mathcal{O}(n^2)$ flops [8].)
The algorithm:

$$\mathbf{w}^{(1)} \equiv w_1 = 1$$

$$\alpha_1 = \gamma_0$$

for $i = 1, \ldots, n-1$

$$\begin{aligned}
\beta_i &= (\gamma_{i+1}, \ldots, \gamma_2) \cdot \mathbf{w}^{(i)} \\
K_i &= -\beta_i / \alpha_i \\
\mathbf{u}^{(i)} &= \alpha_i^{-1} \cdot \mathbf{w}^{(i)} \\
\mathbf{v}^{(i)} &= (0, u_i, \ldots, u_2)^{\mathrm{t}} \\
,_i^{-1} &= u_1^{-1} \left(\mathbf{L}(\mathbf{u}^{(i)})^{\mathrm{t}} \mathbf{L}(\mathbf{u}^{(i)}) - \mathbf{L}(\mathbf{v}^{(i)})^{\mathrm{t}} \mathbf{L}(\mathbf{v}^{(i)}) \right) \\
|,_{i+1}| &= |,_i|(\gamma_0 - \boldsymbol{\gamma}_i^{\mathrm{t}} ,_i^{-1} \boldsymbol{\gamma}_i) \\
\mathbf{w}^{(i+1)} &= (w_1, \ldots, w_i, 0)^{\mathrm{t}} + K_i \cdot (0, w_i, \ldots, w_1)^{\mathrm{t}} \\
\alpha_{i+1} &= \alpha_i + K_i \beta_i
\end{aligned}$$

$$\begin{aligned}
\mathbf{u} &= \alpha_n^{-1} \cdot \mathbf{w}^{(n)} \\
\mathbf{v} &= (0, u_n, \ldots, u_2)^{\mathrm{t}} \\
,^{-1} &= u_1^{-1} (\mathbf{L}(\mathbf{u})^{\mathrm{t}} \mathbf{L}(\mathbf{u}) - \mathbf{L}(\mathbf{v})^{\mathrm{t}} \mathbf{L}(\mathbf{v})) \\
|,| &= |,_n|
\end{aligned}$$

References

[1] A.S. Balghonaim and J.M. Keller. Maximum likelihood estimator for fractal dimension of fractal signal plus noise. In *Proc., IEEE Canadian Conference on Electrical and Computer Engineering, Engineering Innovation: Voyage of Discovery*, volume 1, pages 223–228, St. John's, NF Canada, May 1997.

[2] Jan Beran. *Statistics for Long-Memory Processes.* Chapman & Hall, One Penn Plaza, New York, NY 10119, 1995.

[3] A. Le Breton. Filtering and parameter estimation in a simple linear system driven by a fractional Brownian motion. *Statistics and Probability Letters*, 38:263–274, 1998.

[4] T.F. Chan and P.C. Hansen. A look-ahead levinson algorithm for general toeplitz systems. *IEEE Trans. on Signal Processing*, 40(5):1079–1090, May 1992.

[5] 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)*, volume 5, pages 2841–2844, Atlanta, Georgia, May 1996.

[6] B. Yazıcı and R.L. Kashyap. A class of second-order stationary self-similar processes for $1/f$ phenomena. *IEEE Trans. on Signal Processing*, 45(2):396–410, February 1997.

[7] M. Deriche and A.A. Tewfik. Maximum likelihood estimation of the parameters of discrete fractionally differenced gaussian noise process. *IEEE Trans. Signal Proc.*, 41(10):2977–2989, October 1993.

[8] 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)*, pages 184–192. World Scientific, July 1993.

[9] P. Flandrin. On the spectrum of fractional Brownian motions. *IEEE Trans. on Information Theory*, 35(1):197–199, January 1989.

[10] U. Grenander and G. Szegő. *Toeplitz Forms and Their Application.* Univ. of California Press, Berkeley, 1958.

[11] W. Henkel. An extended berlekamp-massey algorithm for the inversion of toeplitz matrices. *IEEE Trans. on Communications, Transactions Letters*, 40(10):1557–1561, October 1992.

[12] W-C. Lau, A. Erramilli, J.L. Wang, and W. Willinger. Self-similar traffic generation: The random midpoint displacement algorithm and its properties. In *Proc., IEEE Int. Communications Conference (ICC'95)*, pages 466–472, Seattle, WA., June 1995.

[13] J-S. Leu and A. Papamarcou. On estimating the spectral exponent of fractional brownian motion. *IEEE Trans. on Information Theory*, 41(1):233–244, January 1995.

[14] P. Mannersalo and I. Norros. RMD-mn simulator: A fast and simple way to simulate gaussian processes. Technical Document COST257TD(97)31, COST257 Project, MC meeting, Espoo, Finland, May 1997.

[15] B. Ninness. Maximum likelihood estimation of the parameters of fractional brownian motions. In *Proc., 34th Conference on Decision & Control*, pages 4018–4023, New Orleans, LA, December 1995.

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

[17] I. Norros and J.T. Virtamo. Handbook of fbm formulae. COST257TD(96), September 1998.

[18] J. Roberts, U. Mocci, and J. Virtamo, editors. *Broadband Network Teletraffic—Performance Evaluation and Design of Broadband Multiservice Networks*, pages 354–369. Springer, 1996. Final Report of COST 242.

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

[20] A. Vidács and J.T. Virtamo. ML estimation of the parameters of FBM traffic with geometrical sampling. submitted for publication, March 1999.

[21] H-M. Zhang and P. Duhamel. On the methods for solving Yule-Walker equation. *IEEE Trans. on Signal Processing*, 40(12):2987–3000, December 1992.