BUDAPEST UNIVERSITY OF TECHNOLOGY AND ECONOMICS
DEPT. OF TELECOMMUNICATIONS AND TELEMATICS

FRACTAL CHARACTERIZATION OF NETWORK TRAFFIC: FROM PARAMETER ESTIMATION TO APPLICATION

Attila Vidács

Ph.D Dissertation

Supervised by

Dr. Sándor Molnár
High Speed Networks Laboratory
Dept. of Telecommunications and Telematics
Budapest University of Technology and Economics

Budapest, Hungary
2000

Attila Vidács
High Speed Networks Laboratory
Dept. of Telecommunications and Telematics
Budapest University of Technology and Economics[1]

[1]The reviews and the minutes of the Ph.D. Defense are available from the Dean's Office.

BUDAPESTI MŰSZAKI ÉS GAZDASÁGTUDOMÁNYI EGYETEM
TÁVKÖZLÉSI ÉS TELEMATIKAI TANSZÉK

HÁLÓZATI FORGALOM FRAKTÁLIS JELLEMZÉSE: A PARAMETÉRBECSLÉSTŐL AZ ALKALMAZÁSIG

Vidács Attila

Ph.D Dissertation

Tudományos vezető

Dr. Molnár Sándor
Nagysebességű Hálózatok Laboratóriuma
Távközlési és Telematikai Tanszék
Budapesti Műszaki és Gazdaságtudományi Egyetem

Budapest
2000

To my family

Table of Contents

List of Tables

List of Figures

Abstract

The real promise of the fractal characterization framework is that it suggests a very simple way to handle the complex structure of network traffic. This work does not aim at arguing in favor of or against the fractal-type traffic modeling. Rather, it tries to answer some practical questions. The objective is threefold. Firstly, an efficient parameter estimation method of a traffic model based on fractional Brownian motion is proposed. Secondly, the performance implications of long-range dependence in networking is investigated. Finally, a real-time queue length monitoring method is proposed to approximately estimate the effective bandwidth of the traffic.

Kivonat

A fraktális forgalomleírás nagy ígérete, hogy képes lehet a hálózati forgalom összetett struktúráját egyszerű formulákkal megragadni. A jelen dolgozatnak nem célja, hogy a fraktális forgalom-modellezés mellett vagy éppen ellene érveljen, inkább néhány gyakorlati kérdésre próbál meg választ adni. A cél hármas. Elsőként egy a fraktális Brown-mozgáson alapuló forgalommodell paraméterbecslésére ad hatékony módszert. Másodszor a hosszúidejű összefüggőség hálózatok teljesítményére gyakorolt hatását vizsgálja meg. Végül a forgalom effektív sávszélesség becslésére ad egy valósidejű sorhossz-monitorozáson alapuló eljárást.

Acknowledgements

A number of people have directly or indirectly contributed to this thesis. To all these people I express my deepest gratitude.

The supervisor of this Ph.D. Dissertation has been Dr. Sándor Molnár. First of all I wish to thank him for his encouraging support and guidance.

My research work started five years ago at the High Speed Networks Laboratory at the Dept. of Telecommunications and Telematics, Technical University of Budapest. Special thanks to all members and Ph.D. students of the department, and in particular, to the members of HSN Lab. I also would like to express my appreciation to the head of our laboratory, Dr. Tamás Henk.

Nine months of research work was done at the Laboratory of Telecommunications Technology, Helsinki University of Technology. I wish to thank Prof. Jorma Virtamo for this opportunity and for his continuous guidance during my stay in Finland. His door was always open to me.

Another fruitful half year was spent at the Multimedia Networks Laboratories, Nippon Telegraph and Telephone Corp., Tokyo, Japan. I would like to acknowledge Dr. Hiroshi Saito for his supervision during this time.

My colleagues and friends, István Maricza and Szabolcs Malomsoky deserve special thanks for all the help they have provided during the years.

I am very thankful to my family. Throughout the years of my study, they provided me a stable background and there is no doubt that their inspiration and love made it possible for me to carry out my research.

Last but not least, special thanks are due to all those people who helped me and inspired me during this work either in profession or in private life.

Budapest, Attila Vidács
December 1, 2000

Introduction

Understanding the nature of traffic in high speed networks is essential for engineering, operation, and performance evaluation. Traffic characteristics in high speed packet based networks differ substantially from those in telephone networks. The complexity inherent in this new type of traffic arises from the wide range of applications and services provided. Therefore, the problem of obtaining an accurate and tractable characterization of packet traffic is of particular importance.

Statistical analysis of a large number of traffic traces taken from a variety of networking environments revealed that the traffic variations are dominant over a wide range of time scales [3, 19, 24, 31, 44]. These variations can be parsimoniously described using the concepts of long-range dependence (LRD) and self-similarity [23, 42, 55]. The real promise of the self-similar traffic modeling approach is that it suggests a very simple way to handle the extremely complex structure of network traffic.

However, this work does not aim at arguing in favor of or against the self-similar concept in traffic modeling. Rather, it tries to answer some practical questions such as how to parameterize a self-similar traffic model, or what are the performance implications of long-range dependence, and how to apply the results of self-similar traffic modeling in networking.

Once the traffic is decided to be modeled as a self-similar process, one of the simplest and most studied models for aggregated data traffic is the fractional Brownian motion (fBm) model, which is the only self-similar Gaussian process [42]. In its basic form the model has only three parameters, the mean rate m, the variance parameter a and the Hurst parameter H describing the scaling behavior of the traffic. However, the estimation of even a small number of parameters poses a problem for long-range dependent traffic. The problem arises, for example, in the estimation of the Hurst parameter H. As H describes the scaling behavior of the traffic variability, a large number of sample points may be required 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. Therefore to develop an algorithm which is able to do the parameter estimation of the model efficiently is of great importance.

When we deal with fractal-like traffic within the networks, first we want to detect long-range dependence and any presence of self-similar features. Once identified, these properties should be somehow quantified. The estimation and interpretation of the Hurst parameter H as a descriptor of the degree of self-similarity (or long-range dependence) is far from being a trivial task. Some pitfalls must be avoided at any cost, since blind application of different test methods may lead to useless (or even misleading!) results. Once H is captured and

under control, we want to look for various networking mechanisms to intervene and influence the fractal behavior (e.g., to get rid of it if it is harmful for certain reasons.) Unfortunately, the long-range dependence property is much more robust than one would think, therefore removing it is rather difficult. Finally, if—despite our efforts—we have to live together with long-range dependence within the network, we would want to know exactly its impact on key quality of service parameters such as cell loss ratio, for example.

In asynchronous transfer mode (ATM) networks, cell loss ratio (CLR), cell delay and cell delay variation (CDV) are considered to be the major quality of service (QoS) factors. Cell loss and cell delay mainly occur in the output buffers of the network nodes. Traditional queueing analyses based on parametric models have the drawback that since the traffic pattern of ATM streams may be quite complex, the appropriate statistical model seems to be difficult to identify (if possible at all). Moreover, the presence of a scaling property in the traffic has a strong impact on queueing behavior [22, 26]. In order to avoid this difficulty traffic control methods based on real time measurements have been proposed recently [20, 49, 54]. If the actual cell loss performance of an ATM output buffer could be determined in real time, the rate of the server (that is, the VP bandwidth) could be adjusted such that the cell loss would be smaller than a pre-determined threshold. The effective bandwidth approach is one possibility to deal with this problem. A simplified effective bandwidth formula with the CLR objective and some easy-to-measure quantities as parameters could be a useful tool in network management.

Outline of the dissertation

The objective of the dissertation is threefold.

- In the first part the goal was to develop an efficient parameter estimation method of a traffic model based on a Gaussian self-similar process called fractional Brownian motion.

At first, Chapter 1 briefly summarizes the mathematical background of the concepts of self-similarity and long-range dependence. The fractional Brownian motion process is also introduced.

Chapter 2 contains the proposed maximum likelihood estimation (MLE) method to estimate the parameters of the fractional Brownian traffic model. After the introduction of the model in the first section, Section 2.2 gives the detailed description of the exact MLE method for general sampling scheme. Next, results for the ordinary linear sampling are given, and in Section 2.4 a geometrical sampling scheme is investigated and different approximate MLE methods are proposed for this case. In Section 2.5 computer simulation results are given to validate the proposed approximations. The approximate MLE with geometrical sampling is also compared to the linear sampling case as well as to the wavelet-based Abry-Veitch parameter estimation method. Finally, Section 2.6 concludes the chapter.

- In the second part the performance implications of long-range dependence in networking was investigated.

Chapter 3 deals with this topic. The results are based on experimental studies using real data. The first section describes the measured data sets and the measuring configuration, as well as the Hurst-parameter estimation techniques used. Section 3.2 examines some robustness issues of H estimation (i.e., dependence on estimation technique, time scale, nonstationary level shifts). Section 3.3 investigates the effects of network mechanisms such as shaping, policing and multiplexing on the H estimates. Section 3.5 reveals the impacts of scaling behavior of input traffic on cell loss in queueing. A new set of ATM measurements was used for this purpose and are described in this section.

- In the third part the aim was to estimate the effective bandwidth of the traffic using real time traffic measurements. An algorithm to estimate the CLR in real time based on buffer measurements was also proposed, which works for both short-range and long-range dependent input traffic.

The effective bandwidth approach is widely used in network dimensioning and operation. In Section 4.1 a very simple (and thus tractable!) approximate effective bandwidth formula is presented. To evaluate this formula based on on-line measurements, a three-point buffer monitoring method is proposed in Section 4.2 to estimate the CLR. The validation of the proposed method is also given using simulations. Section 4.3 improves the previously mentioned effective bandwidth formula by incorporating the self-similarity parameter H into it.

Chapter 5 summarizes the main results of the dissertation.

Appendix A and B give some technical derivations for the results of the MLE parameter estimation method in Chapter 2. Appendix C presents an approximate method to calculate the effective bandwidth function for Chapter 4.

Chapter 1

Self-Similarity and Long-Range Dependence

The following brief summary is mainly based on [48], [4] and [31].

Self-similar processes[1] are invariant in distribution under scaling of time and space. The scaling coefficient or index of self-similarity is a non-negative number denoted H. They are important in probability because of their connection to limit theorems and they are of great interest in modeling [48]. Lamperti [30] showed that self-similarity arises in a natural way from limit theorems for sums of random variables. The theorem (see [30] and also in [4]) essentially says that whenever a process is the limit of normalized partial sums of random variables, it is necessarily self-similar. Thus the role of self-similar processes among stochastic processes is analogous to the central role of stable distributions among probability distributions [4].

A non-degenerate self-similar process cannot be stationary, but can have stationary increments (H-sssi). The increments of H-sssi processes can display long-range dependence, i.e., the covariances decay very slowly to zero, like a power function.

The existence of moments of H-sssi processes limits the possible values of H. For example, for finite variance processes H is between zero and one. For a given $H \in (0, 1)$, there is a single Gaussian H-sssi process, namely fractional Brownian motion (fBm). The increments of fBm (also called fractional Gaussian noise) exhibit long-range dependence, when $0.5 < H < 1$.

1.1 Self-similarity

Definition 1.1.1. *[48] The real-valued process $\{X(t), t \in \mathbb{R}\}$ is self-similar with index $H > 0$ (H-ss) if for all $a > 0$, the finite-dimensional distributions of $\{X(at), t \in \mathbb{R}\}$ are*

[1] Note, that the term self-similar is also used in the context of the scaling of non-random objects, such as fractals [33].

identical to the finite-dimensional distributions of $\{a^H X(t), t \in \mathbb{R}\}$*; i.e., if for any* $d \geq 1$, $t_1, t_2, \ldots, t_d \in \mathbb{R}$ *and any* $a > 0$,

$$(X(at_1), X(at_2), \ldots, X(at_d)) \stackrel{d}{=} (a^H X(t_1), a^H X(t_2), \ldots, a^H X(t_d)). \tag{1.1}$$

Note: Eq.(1.1) will be expressed succinctly as follows[2]:

$$\{X(at), t \in \mathbb{R}\} \stackrel{d}{=} \{a^H X(t), t \in \mathbb{R}\}. \tag{1.2}$$

Eq.(1.1) states that a change of the time scale is equivalent to a change in the state space scale. Thus, typical sample paths of a self-similar process look qualitatively the same, irrespective of the time-scale from which we look at them. (It does not mean that the same picture repeats itself exactly, it is rather the general impression that remains the same due to the same statistical characteristics [4].)

A non-degenerate H-ss process cannot be stationary. However, there is an important correspondence between self-similar and stationary processes.

Theorem 1.1.2. *[48] If* $\{X(t), 0 < t < \infty\}$ *is* H*-ss, then*

$$Y(t) = e^{-tH} X(e^t), \quad -\infty < t < \infty, \tag{1.3}$$

is stationary. Conversely, if $\{Y(t), -\infty < t < \infty\}$ *is stationary, then*

$$X(t) = t^H Y(\ln t), \quad 0 < t < \infty, \tag{1.4}$$

is H*-ss.*

Theorem 1.1.2 shows that there are many different self-similar processes. From the applications point of view, those that have stationary increments are of great interest because they give rise to stationary sequences with remarkable features.

Definition 1.1.3. *[48] A real-valued process* $\{X(t), t \in \mathbb{R}\}$ *has stationary increments if*

$$\{X(t+h) - X(h), t \in \mathbb{R}\} \stackrel{d}{=} \{X(t) - X(0), t \in \mathbb{R}\}, \quad \text{for all } h \in \mathbb{R}. \tag{1.5}$$

Definition 1.1.4. *[48] The process* $\{X(t), t \in \mathbb{R}\}$ *is called* H*-sssi if it is self-similar with index* H *and has stationary increments.*

[2]The shorthand notation $X(at) \sim a^H X(t)$ will also be used sometimes.

The existence of moments limits the possible values of H.

Lemma 1.1.5. *[48] Suppose that $\{X(t), t \in \mathbb{R}\}$ is a (non-degenerate) H-sssi finite variance process. Then*

$$0 < H \leq 1, \tag{1.6}$$

$$X(0) = 0 \text{ almost surely}, \tag{1.7}$$

and

$$\mathrm{Cov}\left[X(t_1), X(t_2)\right] = \frac{1}{2}\left\{|t_1|^{2H} + |t_2|^{2H} - |t_1 - t_2|^{2H}\right\} \mathrm{Var}\left[X(1)\right]. \tag{1.8}$$

Moreover, in the case $0 < H < 1$ from Eq.(1.2) we get

$$\mathrm{E}\left[X(t)\right] = 0. \tag{1.9}$$

In the following, let $\{X(t), t \in \mathbb{R}\}$ be a non-degenerate H-sssi finite variance process with $0 < H < 1$. The increment sequence of $\{X(t), t \in \mathbb{R}\}$ in discrete time can be defined as

$$Y_k = X(k+1) - X(k), \quad k \in \mathbb{Z}. \tag{1.10}$$

The autocovariances of the increments are of the form

$$\mathrm{Cov}\left[Y_i, Y_{i+k}\right] = \frac{1}{2}\left\{|k+1|^{2H} - 2|k|^{2H} + |k-1|^{2H}\right\} \mathrm{Var}\left[Y_1\right]. \tag{1.11}$$

1.2 Long-range dependence

Let $\gamma(\cdot)$ denote the autocovariance function of $\{Y_i, i \in \mathbb{Z}\}$ defined by

$$\gamma(k) = \mathrm{Cov}\left[Y_i, Y_{i+k}\right], \quad k \in \mathbb{Z}, \tag{1.12}$$

and $\rho(\cdot)$ denote the autocorrelation function of the process given by $\rho(k) = \gamma(k)/\gamma(0)$. The spectral density $f(\cdot)$ of $\{Y_i, i \in \mathbb{Z}\}$ can be defined as

$$f(\lambda) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma(k) e^{\mathbf{i}k\lambda}, \quad \lambda \in [-\pi, \pi] \tag{1.13}$$

where $\mathbf{i} = \sqrt{-1}$.

Definition 1.2.1. *[4]* $\{Y_i, i \in \mathbb{Z}\}$ *is called a stationary process with short-range dependence (or short memory) if there exists a constant* $0 < c_0 < 1$ *such that*

$$\lim_{k\to\infty} \frac{\rho(k)}{c_0^k} = 1. \tag{1.14}$$

Definition 1.2.2. *[4]* $\{Y_i, i \in \mathbb{Z}\}$ *is called a stationary process with long-range dependence (or long memory) if there exists a real number* $\alpha \in (0,1)$ *and a constant* $c_\rho > 0$ *such that*

$$\lim_{k\to\infty} \frac{\rho(k)}{c_\rho k^{-\alpha}} = 1, \tag{1.15}$$

or equivalently (by virtue of the Tauberian theorem [4]), if there exists a real number $\beta \in (0,1)$ *and a constant* $c_f > 0$ *such that*

$$\lim_{\lambda\to 0} \frac{f(\lambda)}{c_f |\lambda|^{-\beta}} = 1. \tag{1.16}$$

It is important to note that the definition of long-range dependence is an asymptotic definition. It only tells us something about the ultimate behavior of the correlations as the lag tends to infinity. In this generality, it does not specify the correlations for any fixed finite lag. Moreover, it determines only the rate of convergence, not the absolute size which is given by the constants c_ρ and c_f. Each individual correlation can be arbitrarily small, only the decay of the correlations is slow. This makes the detection of long-range dependence difficult [4].

The asymptotic behavior of $\rho(\cdot)$ (see Eq.(1.11)) follows by Taylor expansion:

$$\rho(k) = H(2H-1)k^{2H-2} + o(k^{2H-2}). \tag{1.17}$$

According to Definition 1.2.2 the process $\{Y_i, i \in \mathbb{Z}\}$ with $0.5 < H < 1$ is long-range dependent with parameter $\alpha = 2 - 2H$ in Eq.(1.15) or with $\beta = 2H - 1$ in Eq.(1.16). This also means that the correlations are nonsummable:

$$\sum_{k=-\infty}^{\infty} \rho(k) = \infty. \tag{1.18}$$

(For $H = 0.5$ the increments are uncorrelated, and for $0 < H < 0.5$ the correlations sum up to zero [4].)

The spectral density of $\{Y_i, i \in \mathbb{Z}\}$ is given by [4]

$$f(\lambda) = 2c_f(1-\cos\lambda) \sum_{j=-\infty}^{\infty} |2\pi j + \lambda|^{-2H-1}, \quad \lambda \in [-\pi, \pi] \tag{1.19}$$

with

$$c_f = \frac{1}{2\pi} \sin(\pi H)\, \Gamma(2H+1) \mathrm{Var}\left[Y_1\right]. \tag{1.20}$$

The behavior of $f(\cdot)$ near the origin follows by Taylor expansion at zero [4]:

$$f(\lambda) = c_f |\lambda|^{1-2H} + O\left(|\lambda|^{\min(3-2H,2)}\right). \tag{1.21}$$

The approximation of $f(\lambda)$ by $c_f|\lambda|^{1-2H}$ is in fact very good even for relatively large frequencies.

Eq.(1.21) is used when estimating H in the frequency domain (see periodogram plot method in Section 3.1.2).

1.3 Aggregated process

Let $\{Y_k, k \in \mathbb{Z}\}$ be a stationary stochastic process in discrete time with finite variance. For each $m = 1, 2, \ldots$, let $\{Y_k^{(m)}, k \in \mathbb{Z}\}$ denote a new time series obtained by averaging (and rescaling) the original series Y_k over non-overlapping blocks of size m. That is,

$$Y_k^{(m)} = \frac{1}{m^H} \sum_{i=(k-1)m+1}^{km} Y_i, \quad k \in \mathbb{Z} \tag{1.22}$$

with some $0 < H < 1$. Note that for each m, the aggregated process is also stationary and has finite variance. Let $\rho^{(m)}(\cdot)$ denote its autocorrelation function.

Definition 1.3.1. *The process $\{Y_k, k \in \mathbb{Z}\}$ is called second-order self-similar with parameter H, if*

$$\rho^{(m)}(k) = \rho(k), \quad m = 1, 2, \ldots . \tag{1.23}$$

In other words, the process is second-order self-similar if the aggregated processes $\{Y_k^{(m)}, k \in \mathbb{Z}\}$ are indistinguishable from $\{Y_k, k \in \mathbb{Z}\}$ as far as their second order statistical properties are concerned [31].

Note that in the above definition the stationary process $\{Y_k, k \in \mathbb{Z}\}$ is called self-similar. This is seemingly in contrast to the definitions in the previous sections, where nonstationary processes in continuous time fulfilled the requirements of self-similarity. The link between the two can be explained as follows. Assume now that the process $\{Y_k, k \in \mathbb{Z}\}$ is the increment process of some process $\{X(t), t \in \mathbb{R}\}$. If $\{X(t), t \in \mathbb{R}\}$ is a H-sssi process then

its increment process fulfills the requirements of Definition 1.3.1. Moreover, in case of H-sssi processes a more stringent statement is valid: their aggregated increment processes satisfy

$$\{Y_k, k \in \mathbb{Z}\} \stackrel{d}{=} \{Y_k^{(m)}, k \in \mathbb{Z}\}. \tag{1.24}$$

It follows from Eq.(1.23) that the autocorrelation function of a second-order self-similar process is of the form

$$\rho(k) = r(k) \stackrel{\text{def}}{=} \frac{1}{2}\left\{|k+1|^{2H} - 2|k|^{2H} + |k-1|^{2H}\right\}. \tag{1.25}$$

Note, that $r(k)$ is the autocorrelation function of the increment process of a H-sssi process (see Eq.(1.11)). Hence a second-order self-similar process with $0.5 < H < 1$ is long-range dependent.

Eq.(1.23) is a rather strict constraint to $\rho^{(m)}(\cdot)$ and, as a result, $\rho(k)$ is determined for all the values of k. If we only require that $\rho^{(m)}(\cdot)$ approaches $r(\cdot)$ as the aggregation level increases we get the following definition:

Definition 1.3.2. *The process $\{Y_k, k \in \mathbb{Z}\}$ is called asymptotically second-order self-similar with parameter H if*

$$\lim_{m\to\infty} \rho^{(m)}(k) = r(k). \tag{1.26}$$

The striking feature is that the aggregated processes possess a nondegenerate correlation structure as m tends to infinity. As a result, the processes whose autocorrelation function satisfies Eq.(1.26) will satisfy Eq.(1.15) in Definition 1.2.2. In other words, asymptotical second-order self-similarity with $0.5 < H < 1$ and long-range dependence are equivalent concepts.

As a summary, Figure 1.1 shows the connections between the self-similarity related definitions. (The label 'H-fix' denotes the set of processes which satisfy Eq.(1.24), called as fixed-point self-similar processes.)

1.4 Cumulative arrival process

This section provides the basics of how self-similar processes can be used in traffic modeling.

Let $A(t)$ represent the amount of traffic arrived within the interval $(0, t]$. The cumulative arrival process $\{A(t), t > 0\}$ can be modeled by

$$A(t) = mt + X(t), \quad t > 0 \tag{1.27}$$

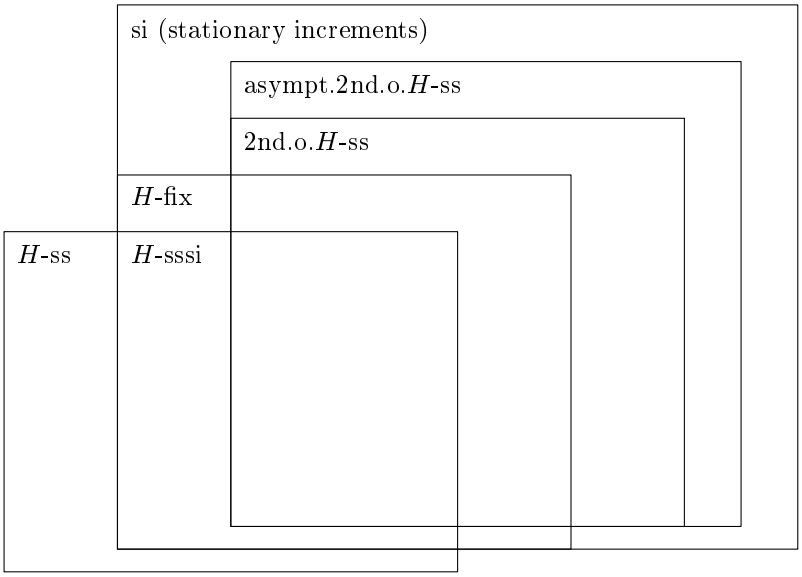

Figure 1.1: Connections between self-similarity related definitions.

where $m > 0$ is the mean arrival rate, and $\{X(t), t \in \mathbb{R}\}$ is a H-sssi process with variance $\mathrm{Var}[X(1)] = \sigma^2$. The process $\{A(t), t > 0\}$ has the following mean and variance:

$$\mathrm{E}[A(t)] = mt, \tag{1.28}$$
$$\mathrm{Var}[A(t)] = t^{2H}\sigma^2. \tag{1.29}$$

The following properties of $A(t)$ are useful when one tries to estimate the parameter H (see later in Section 3.1.2).

The variance of $A(t)/t$ can be written as

$$\mathrm{Var}[A(t)/t] = t^{2H-2}\sigma^2. \tag{1.30}$$

The property in Eq.(1.30) forms the basis of the variance-time method to estimate H (see Section 3.1.2).

The index of dispersion for counts (IDC) defined by

$$IDC(t) = \frac{\mathrm{Var}[A(t)]}{\mathrm{E}[A(t)]} \tag{1.31}$$

can be calculated as

$$IDC(t) = t^{2H-1}m^{-1}\sigma^2. \tag{1.32}$$

This property forms the basis of the IDC plot method (see Section 3.1.2).

1.5 Fractional Brownian motion

Suppose that $\{Z(t), t \in \mathbb{R}\}$ is a H-sssi process with $0 < H < 1$. The expected value of the increment process $Y_i = Z(i+1) - Z(i)$ $(i = 0, 1, 2 \dots)$ is zero (see Eq.(1.9)), and the covariances of Y_i are given by Eq.(1.11). Furthermore, suppose that the increment process is Gaussian. Then the distribution of the process is fully specified by the mean and covariances. Therefore for each value of $H \in (0, 1)$ there is exactly one Gaussian process $\{Y_k, k \in \mathbb{Z}\}$ that is the stationary increment of a self-similar process $Z(t)$. The increment process is called fractional Gaussian noise (fGn) and the corresponding self-similar process $\{Z(t), t \in \mathbb{R}\}$ is called fractional Brownian motion. By definition [48]:

Definition 1.5.1. *[48] A Gaussian H-sssi process, $0 < H \leq 1$, is called fractional Brownian motion (fBm) and is denoted $\{Z(t), t \in \mathbb{R}\}$. It is called standard fractional Brownian motion if* $\mathrm{Var}\,[Z(1)] = 1$.

The fBm process is widely applied in modeling, particularly in the context of long-range dependence, when $0.5 < H < 1$. A standard fBm with Hurst-parameter $H \in [0.5, 1)$ is characterized by the following properties [41]:

1. $Z(t)$ has stationary increments;
2. $Z(0) = 0$, and $\mathrm{E}\,[Z(t)] = 0$ for all t;
3. $\mathrm{Var}\,[Z(t)] = \mathrm{E}\left[Z(t)^2\right] = |t|^{2H}$ for all t;
4. $Z(t)$ has continuous sample 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.

The fBm process will play a central role in this dissertation.

Chapter 2

Parameter Estimation of Fractional Brownian Traffic

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 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 Abry-Veitch estimator show that the geometrical sampling improves indeed the accuracy of the estimate $\hat{H}$ with a given number of samples.

2.1 Fractional Brownian traffic

One of the simplest and most studied models for aggregated data traffic is the fractional Brownian motion (fBm) model [45], which is a model for 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 [42]. In its basic form the model contains only three parameters, and a small number of traffic parameters is a very desirable feature from the applicability point of view for traffic engineering purposes.

Norros [41] has suggested the following model

$$X(t) = mt + \sqrt{a}Z(t), \tag{2.1}$$

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 the standard fractional Brownian motion $Z(t)$.

Since $Z(t)$ is a H-sssi process, its scaling behavior is defined by the Hurst parameter H as follows

$$Z(\alpha t) \sim \alpha^H Z(t). \tag{2.2}$$

The covariance structure of the process is given by

$$\mathrm{Cov}\left[Z(t_1), Z(t_2)\right] = \frac{1}{2}\left\{t_1^{2H} + t_2^{2H} - |t_2 - t_1|^{2H}\right\}. \tag{2.3}$$

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

$$\begin{aligned} \mathrm{E}\left[X(t)\right] &= mt; & (2.4)\\ \mathrm{Var}\left[X(t)\right] &= at^{2H}; & (2.5)\\ \mathrm{Cov}\left[X(t_1), X(t_2)\right] &= \frac{a}{2}\left\{t_1^{2H} + t_2^{2H} - |t_2 - t_1|^{2H}\right\}. & (2.6) \end{aligned}$$

It follows from Eq.(2.1) and the above properties of $Z(t)$ that $(X(t+1) - X(t))$, the amount of traffic arrived in the unit interval $[t, t+1)$, has the following mean and variance:

$$\begin{aligned} \mathrm{E}\left[X(t+1) - X(t)\right] &= m, & (2.7)\\ \mathrm{Var}\left[X(t+1) - X(t)\right] &= a. & (2.8) \end{aligned}$$

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

2.2 Exact Gaussian MLE for general sampling scheme

Throughout this work the maximum likelihood estimation (MLE) method [4] is applied. (The MLE method was previously applied to this problem by Deriche and Tewfik [15] and Ninness [40] 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.

Assume that 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 probability density function of $\mathbf{X}$ is equal to

$$h(\mathbf{x}) = (2\pi)^{-\frac{n}{2}} |\Gamma|^{-\frac{1}{2}} e^{-\frac{1}{2}(\mathbf{x}-\mathbf{m})^{\mathrm{t}}\Gamma^{-1}(\mathbf{x}-\mathbf{m})}, \tag{2.9}$$

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

$$\Gamma = \Big[\mathrm{Cov}\left[X(t_i), X(t_j)\right]\Big]_{i,j=1,\ldots,n}. \tag{2.10}$$

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

The ML estimate $\hat{m}$ can be derived as follows. The log-likelihood function is given by

$$\log h(\mathbf{X}; m) = -\frac{n}{2}\log 2\pi - \frac{1}{2}\log|\Gamma| - \frac{1}{2}(\mathbf{X} - m\mathbf{t})^{\mathrm{t}}\,\Gamma^{-1}\,(\mathbf{X} - m\mathbf{t}). \tag{2.11}$$

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

$$L(\mathbf{X}; m) = (\mathbf{X} - m\mathbf{t})^{\mathrm{t}}\,\Gamma^{-1}\,(\mathbf{X} - m\mathbf{t}). \tag{2.12}$$

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

$$\begin{aligned} L'(\mathbf{X}; \hat{m}) &= \frac{\partial}{\partial m} L(\mathbf{X}; \hat{m}) \\ &= 2\hat{m}\mathbf{t}^{\mathrm{t}}\,\Gamma^{-1}\,\mathbf{t} - 2\mathbf{t}^{\mathrm{t}}\,\Gamma^{-1}\,\mathbf{X} = 0 \end{aligned} \tag{2.13}$$

and using the relationships $(\mathbf{AB})^{\mathrm{t}} = \mathbf{B}^{\mathrm{t}}\mathbf{A}^{\mathrm{t}}$ and $(\mathbf{A}^{\mathrm{t}})^{-1} = (\mathbf{A}^{-1})^{\mathrm{t}}$ valid for any matrices $\mathbf{A}$ and $\mathbf{B}$, and the fact that Γ is symmetric, we get

$$\hat{m} = \hat{m}(H) = \frac{\mathbf{t}^{\mathrm{t}}\,\Gamma^{-1}\mathbf{X}}{\mathbf{t}^{\mathrm{t}}\,\Gamma^{-1}\,\mathbf{t}}. \tag{2.14}$$

The expectation and variance are important properties of an estimator. Since the expectation of $\mathbf{X}$ is $m\mathbf{t}$ we have

$$\mathrm{E}\,[\hat{m}] = \frac{\mathbf{t}^{\mathrm{t}}\,\hat{\Gamma}_H^{-1}\,\mathrm{E}\,[\mathbf{X}]}{\mathbf{t}^{\mathrm{t}}\,\hat{\Gamma}_H^{-1}\,\mathbf{t}} = m, \tag{2.15}$$

where $\hat{\Gamma}_H^{-1} \stackrel{\mathrm{def}}{=} \Gamma_H^{-1}(\hat{H})$, 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!) Finally we get (see Appendix A):

$$\mathrm{Var}\,[\hat{m}] = \frac{a}{\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{t}}. \tag{2.16}$$

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

Next, consider the estimator for a. Note, that $\Gamma = \Gamma(a)$ is a simple linear function of a:

$$\Gamma = \Gamma(a) = a\,\Gamma_H \tag{2.17}$$

where, using the notation $\mathbf{Z} = (Z(t_1), \ldots, Z(t_n))^{\mathrm{t}}$, Γ_H is given as

$$\Gamma_H = \mathrm{E}\left[\mathbf{Z}\mathbf{Z}^{\mathrm{t}}\right] = \Big[\mathrm{Cov}\left[Z(t_i), Z(t_j)\right]\Big]_{i,j=1,\ldots,n}. \tag{2.18}$$

Hence the log-likelihood function is given by

$$\log h(\mathbf{X}; a) = -\frac{n}{2}\log 2\pi - \frac{1}{2}\log a^n\,|\Gamma_H| - \frac{1}{2a}(\mathbf{X}-\mathbf{m})^{\mathrm{t}}\,\Gamma_H^{-1}\,(\mathbf{X}-\mathbf{m}). \tag{2.19}$$

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

$$L(\mathbf{X}; a) = n\log a + \frac{1}{a}(\mathbf{X}-\mathbf{m})^{\mathrm{t}}\,\Gamma_H^{-1}\,(\mathbf{X}-\mathbf{m}). \tag{2.20}$$

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

$$\begin{aligned} L'(\mathbf{X}; \hat{a}) &= \frac{\partial}{\partial a} L(\mathbf{X}; \hat{a}) \\ &= \frac{n}{\hat{a}} - \frac{1}{\hat{a}^2}(\mathbf{X}-\mathbf{m})^{\mathrm{t}}\,\Gamma_H^{-1}\,(\mathbf{X}-\mathbf{m}) = 0 \end{aligned} \tag{2.21}$$

and from this we get

$$\hat{a} = \hat{a}(m, H) = \frac{1}{n}(\mathbf{X}-\mathbf{m})^{\mathrm{t}}\,\Gamma_H^{-1}\,(\mathbf{X}-\mathbf{m}). \tag{2.22}$$

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

If we do not know the mean input rate m in advance, in Eq.(2.22) $\mathbf{m}$ should be replaced by $\hat{m}\mathbf{t}$. Using Eq.(2.14) and Eq.(2.22) we get

$$\begin{aligned} \hat{a}(H) &= \frac{1}{n}(\mathbf{X}-\hat{m}\mathbf{t})^{\mathrm{t}}\,\Gamma_H^{-1}\,(\mathbf{X}-\hat{m}\mathbf{t}) \\ &= \frac{1}{n}\frac{(\mathbf{X}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{X})(\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{t}) - (\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{X})^2}{\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{t}}. \end{aligned} \tag{2.23}$$

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

$$n\mathrm{E}\left[\hat{a}\right] = (n-1)a. \tag{2.24}$$

Thus the variance estimator has the "usual" $(n-1)/n$ bias. The "naïve" estimator [45] 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 $\hat{a}$ (see Appendix A):

$$\mathrm{Var}\left[\hat{a}\right] = \frac{2(n-1)}{n^2}a^2. \tag{2.25}$$

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

$$\mathrm{Var}\left[\frac{n}{n-1}\,\hat{a}\right] = \frac{2a^2}{n-1}. \tag{2.26}$$

2.2.4 Likelihood function and its derivative for $\hat{H}$

Finally, we are left with the estimation of H. The task is to maximize the log-likelihood function

$$\log h(\mathbf{X};H) = -\frac{n}{2}\log 2\pi - \frac{1}{2}\log|\hat{a}(H)\,\Gamma_H)| - \frac{1}{2\hat{a}(H)}(\mathbf{X}-\hat{m}(H)\mathbf{t})^{\mathrm{t}}\,\Gamma_H^{-1}\,(\mathbf{X}-\hat{m}(H)\mathbf{t}) \tag{2.27}$$

or equivalently, the minimization of

$$\begin{aligned} L(\mathbf{X};H) &= \log|\hat{a}(H)\,\Gamma_H| + \frac{1}{\hat{a}(H)}(\mathbf{X}-\hat{m}(H)\mathbf{t})^{\mathrm{t}}\,\Gamma_H^{-1}\,(\mathbf{X}-\hat{m}(H)\mathbf{t}) \qquad (2.28)\\ &= \log\hat{a}^n(H)\,|\Gamma_H| + \frac{1}{\hat{a}(H)}\frac{(\mathbf{X}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{X})(\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{t}) - (\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{X})^2}{\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{t}}\\ &= n\log\frac{(\mathbf{X}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{X})(\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{t}) - (\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{X})^2}{\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{t}} - n\log n + \log|\Gamma_H| + n \to \min, \end{aligned}$$

i.e., essentially we have to minimize

$$\tilde{L}(\mathbf{X};H) = |\Gamma_H|^{1/n}\frac{(\mathbf{X}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{X})(\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{t}) - (\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{X})^2}{\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{t}}. \tag{2.29}$$

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 corresponding ML estimates for m and a are $\hat{m} = m(\hat{H})$ and $\hat{a} = a(\hat{H})$.

The major difficulty in this method is the calculation of the inverse and determinant of the covariance matrix appearing in the likelihood function.

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

$$L'(\mathbf{X},H) = \frac{\partial}{\partial H}\Big(\log|\Gamma_H|\Big) + n\frac{\partial}{\partial H}\log\left[(\mathbf{X}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{X}) - \frac{(\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{X})^2}{\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{t}}\right] = 0. \tag{2.30}$$

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

$$L'(\mathbf{X},H) = \frac{\partial}{\partial H}\Big(\log|\Gamma_H|\Big) + n\frac{(\mathbf{u}+\mathbf{v})^{\mathrm{t}}\,\left(\frac{\partial}{\partial H}\Gamma_H^{-1}\right)\,(\mathbf{u}+\mathbf{v})}{(\mathbf{u}+\mathbf{v})^{\mathrm{t}}\,\Gamma_H^{-1}\,(\mathbf{u}-\mathbf{v})} = 0. \tag{2.31}$$

If we use the relationships (where $\mathrm{Tr}(\cdot)$ denotes the trace of a matrix)

$$\frac{\partial}{\partial\theta}\mathbf{A}_\theta^{-1} = -\mathbf{A}_\theta^{-1}\left(\frac{\partial}{\partial\theta}\mathbf{A}_\theta\right)\mathbf{A}_\theta^{-1}, \tag{2.32}$$

$$\frac{\partial}{\partial\theta}\log|\mathbf{A}_\theta| = \mathrm{Tr}\left(\mathbf{A}_\theta^{-1}\frac{\partial}{\partial\theta}\mathbf{A}_\theta\right) \tag{2.33}$$

valid for any matrix $\mathbf{A}_\theta$ depending on a parameter θ [16] we have (with $\Gamma'_H = \frac{\partial}{\partial H}\Gamma_H$)

$$L'(\mathbf{X}, H) = \mathrm{Tr}(\Gamma_H^{-1}\,\Gamma'_H) + n\frac{(\mathbf{v}+\mathbf{u})^{\mathrm{t}}\left(\Gamma_H^{-1}\,\Gamma'_H\,\Gamma_H^{-1}\right)(\mathbf{v}+\mathbf{u})}{(\mathbf{v}+\mathbf{u})^{\mathrm{t}}\,\Gamma_H^{-1}\,(\mathbf{v}-\mathbf{u})} = 0. \tag{2.34}$$

It is important that in order to solve Eq.(2.34) we do not need to calculate the determinant of Γ_H.

2.3 Exact Gaussian MLE using linear sampling

A special (and most common) case of the general sampling scheme is when the samples are taken uniformly in time. Let $\mathbf{X} = (X(t_1), X(t_2), \dots, X(t_n))^{\mathrm{t}}$ be the vector of observed traffic values at instances

$$t_i = \frac{i}{n}, \quad i = 1, 2, \dots, n. \tag{2.35}$$

The increment sequence $(Y_1, Y_2, \dots)$ 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

$$\mathrm{Cov}\,[Y_i, Y_j] = \frac{1}{2}an^{-2H}\left(|i-j+1|^{2H} + |i-j-1|^{2H} - 2|i-j|^{2H}\right) \tag{2.36}$$

for $i, j = 1, 2, \dots, n$. The formulas for the exact Gaussian MLE for this increment process are nearly the same as in Section 2.2, we only need to replace the covariance matrix Γ with $\Sigma = [\mathrm{Cov}\,[Y_i, Y_j]]_{i,j=1,2,\dots,n}$, and the vector $\mathbf{t}$ with the vector $(1/n, 1/n, \dots, 1/n)^{\mathrm{t}}$. After some minor simplifications we get an estimate for m

$$\hat{m} = \hat{m}(H) = \frac{\mathbf{1}^{\mathrm{t}}\,\Sigma^{-1}\,\mathbf{Y}}{\mathbf{1}^{\mathrm{t}}\,\Sigma^{-1}\,\mathbf{1}} \cdot n \tag{2.37}$$

where $\mathbf{1}$ is a vector of ones, and $\Sigma = a\Sigma_H$. For a we have the estimator

$$\hat{a}(H) = \frac{1}{n}\left(\mathbf{Y}^{\mathrm{t}}\,\Sigma_H^{-1}\,\mathbf{Y} - \frac{\left(\mathbf{1}^{\mathrm{t}}\,\Sigma^{-1}\,\mathbf{Y}\right)^2}{\mathbf{1}^{\mathrm{t}}\,\Sigma^{-1}\,\mathbf{1}}\right). \tag{2.38}$$

Again, finally we have to minimize

$$\tilde{L}(\mathbf{Y}; H) = \log|\Sigma_H| + n\log\left(\mathbf{Y}^{\mathrm{t}}\,\Sigma_H^{-1}\,\mathbf{Y} - \frac{\left(\mathbf{1}^{\mathrm{t}}\,\Sigma^{-1}\,\mathbf{Y}\right)^2}{\mathbf{1}^{\mathrm{t}}\,\Sigma^{-1}\,\mathbf{1}}\right). \tag{2.39}$$

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

2.4 Exact Gaussian MLE using geometrical sampling

The Hurst parameter H describes the scaling behavior of the traffic. Therefore, in order to determine its value from measured data, the sample points must cover several time scales, i.e., the total time of the measurements must 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 a very large number of samples. 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, a geometric sequence of sampling points can be introduced: $t_i = \alpha^i$, $i = 1, \ldots, n$, 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 with the self-similar behavior of the fBm traffic. By a simple transformation we can obtain from the fBm process another process which is a stationary process on logarithmic time scale. 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.)

2.4.1 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 (see Theorem 1.1.2 in Section 1.1). This transformation can be achieved by first *descaling* the process and then distorting the time axis exponentially. The descaled process can be obtained from the original process by multiplying it with the term t^{-H}, and the time distortion can be achieved by changing the process' index to logarithmic time.

$Z(t)$ has the self-similar property $Z(\alpha t) \sim \alpha^H Z(t)$. Now consider the descaled process

$$\tilde{Y}(t) \stackrel{\text{def}}{=} t^{-H} Z(t) \tag{2.40}$$

which has the scaling property

$$\tilde{Y}(\alpha t) \sim (\alpha t)^{-H} Z(\alpha t) = t^{-H} Z(t) = \tilde{Y}(t). \tag{2.41}$$

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

$$Y(u) \stackrel{\text{def}}{=} \tilde{Y}(e^{-u}) = \tilde{Y}(t). \tag{2.42}$$

Now we have

$$Y(u - \log \alpha) = \tilde{Y}(e^{-u+\log \alpha}) = \tilde{Y}(\alpha e^{-u}) = \tilde{Y}(\alpha t) \sim \tilde{Y}(t) = Y(u). \tag{2.43}$$

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

$$\begin{aligned} \mathrm{E}\left[Y(u)\right] &= 0; && (2.44)\\ \mathrm{Var}\left[Y(u)\right] &= 1; && (2.45)\\ \mathrm{Cov}\left[Y(u_1), Y(u_2)\right] &= \frac{1}{2} e^{H(u_2-u_1)} \left\{ 1 + e^{-2H(u_2-u_1)} - \left(1 - e^{-(u_2-u_1)}\right)^{2H} \right\} && (2.46)\\ &= g(\tau), \quad \text{where } \tau = u_2 - u_1, \quad 0 < u_1 < u_2. \end{aligned}$$

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

$$\tilde{W}(t) \stackrel{\text{def}}{=} t^{-H} X(t) = m t^{1-H} + \sqrt{a}\tilde{Y}(t), \tag{2.47}$$

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

$$W(u) \stackrel{\text{def}}{=} \tilde{W}(e^{-u}) = m e^{(H-1)u} + \sqrt{a} Y(u). \tag{2.48}$$

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

$$\begin{aligned} \mathrm{E}\left[W(u)\right] &= m e^{(H-1)u}; && (2.49)\\ \mathrm{Var}\left[W(u)\right] &= a; && (2.50)\\ \mathrm{Cov}\left[W(u_1), W(u_2)\right] &= \frac{a}{2} e^{H(u_2-u_1)} \left\{ 1 + e^{-2H(u_2-u_1)} - \left(1 - e^{-(u_2-u_1)}\right)^{2H} \right\}. && (2.51) \end{aligned}$$

2.4.2 Geometrical sampling

The idea here is that a geometrical 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.

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

$$t_i = \alpha^{i-1}, \quad i = 1, 2, \ldots, n, \quad 0 < \alpha < 1. \tag{2.52}$$

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

$$\tilde{\Gamma} = \mathrm{E}\left[\mathbf{W}\mathbf{W}^{\mathrm{t}}\right] = a \cdot \mathrm{E}\left[\mathbf{Y}\mathbf{Y}^{\mathrm{t}}\right]. \tag{2.53}$$

Note that our geometrical grid is now equally spaced with regard 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

$$\rho(i-j) = \mathrm{Cov}\left[Y_i, Y_j\right], \quad i, j = 1, 2, \ldots \tag{2.54}$$

and thus

$$\tilde{\Gamma}_{ij} = a\rho(i-j), \quad i,j = 1,2,\ldots,n. \tag{2.55}$$

From Eq.(2.44) and Eq.(2.52) we get

$$\begin{aligned} \rho(i-j) &= \frac{1}{2}e^{-H|i-j|\log\alpha}\left\{1 + e^{2H|i-j|\log\alpha} - \left(1 - e^{|i-j|\log\alpha}\right)^{2H}\right\} \\ &= \frac{1}{2}\alpha^{-H|i-j|}\left\{1 + \alpha^{2H|i-j|} - \left(1 - \alpha^{|i-j|}\right)^{2H}\right\}. \end{aligned} \tag{2.56}$$

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

$$\tilde{\Gamma} = a \cdot \begin{pmatrix} 1 & \beta g(\alpha) & \beta^2 g(\alpha^2) & \cdots & \beta^{n-1} g(\alpha^{n-1}) \\ \beta g(\alpha) & 1 & \beta g(\alpha) & \ddots & \vdots \\ \beta^2 g(\alpha^2) & \beta g(\alpha) & 1 & \ddots & \beta^2 g(\alpha^2) \\ \vdots & \ddots & \ddots & \ddots & \beta g(\alpha) \\ \beta^{n-1} g(\alpha^{n-1}) & \cdots & \beta^2 g(\alpha^2) & \beta g(\alpha) & 1 \end{pmatrix}. \tag{2.57}$$

2.4.3 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 follows:

$$\begin{aligned} \tilde{\Gamma} &= \mathrm{E}\left[\mathbf{W}\mathbf{W}^{\mathrm{t}}\right] \\ &= \mathrm{E}\left[(\mathbf{D}\mathbf{X})(\mathbf{D}\mathbf{X})^{\mathrm{t}}\right] \\ &= \mathbf{D}\,\mathrm{E}\left[\mathbf{X}\mathbf{X}^{\mathrm{t}}\right]\mathbf{D} \\ &= \mathbf{D}\Gamma\mathbf{D}. \end{aligned} \tag{2.58}$$

From this one can easily deduce that

$$\Gamma^{-1} = \mathbf{D}\,\tilde{\Gamma}^{-1}\mathbf{D}. \tag{2.59}$$

Substituting Eq.(2.59) into Eq.(2.14) the MLE $\hat{m}$ is

$$\hat{m}(H) = \frac{\mathbf{t}^{\mathrm{t}}\,\mathbf{D}\tilde{\Gamma}^{-1}\mathbf{D}\,\mathbf{X}}{\mathbf{t}^{\mathrm{t}}\,\mathbf{D}\tilde{\Gamma}^{-1}\mathbf{D}\,\mathbf{t}}. \tag{2.60}$$

Similarly, the same result holds when estimating a as in Eq.(2.23) with $\Gamma_H^{-1} = \mathbf{D}\tilde{\Gamma}_H^{-1}\mathbf{D}$, namely,

$$\begin{aligned} \hat{a}(H) &= \frac{1}{n}(\mathbf{X} - \hat{m}\mathbf{t})^{\mathrm{t}}\,\mathbf{D}\tilde{\Gamma}_H^{-1}\mathbf{D}\,(\mathbf{X} - \hat{m}\mathbf{t}) \qquad (2.61)\\ &= \frac{1}{n}\frac{(\mathbf{X}^{\mathrm{t}}\,\mathbf{D}\tilde{\Gamma}_H^{-1}\mathbf{D}\,\mathbf{X})(\mathbf{t}^{\mathrm{t}}\,\mathbf{D}\tilde{\Gamma}_H^{-1}\mathbf{D}\,\mathbf{t}) - (\mathbf{t}^{\mathrm{t}}\,\mathbf{D}\tilde{\Gamma}_H^{-1}\mathbf{D}\,\mathbf{X})^2}{\mathbf{t}^{\mathrm{t}}\,\mathbf{D}\tilde{\Gamma}_H^{-1}\mathbf{D}\,\mathbf{t}}. \end{aligned}$$

Finally, we have to minimize

$$\tilde{L}(\mathbf{X}; H) = |\Gamma_H|^{1/n}\frac{(\mathbf{X}^{\mathrm{t}}\,\mathbf{D}\tilde{\Gamma}_H^{-1}\mathbf{D}\,\mathbf{X})(\mathbf{t}^{\mathrm{t}}\,\mathbf{D}\tilde{\Gamma}_H^{-1}\mathbf{D}\,\mathbf{t}) - (\mathbf{t}^{\mathrm{t}}\,\mathbf{D}\tilde{\Gamma}_H^{-1}\mathbf{D}\,\mathbf{X})^2}{\mathbf{t}^{\mathrm{t}}\,\mathbf{D}\tilde{\Gamma}_H^{-1}\mathbf{D}\,\mathbf{t}}. \tag{2.62}$$

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

$$\begin{aligned} |\Gamma_H| &= |\mathbf{D}^{-1}\,\tilde{\Gamma}_H\,\mathbf{D}^{-1}| \qquad (2.63)\\ &= \left(\prod_{i=1}^{n} t_i^H\right)|\tilde{\Gamma}_H\,\mathbf{D}^{-1}|\\ &= \left(\prod_{i=1}^{n} t_i^{2H}\right)|\tilde{\Gamma}_H|\\ &= \left(\prod_{i=1}^{n} \alpha^{2H(i-1)}\right)|\tilde{\Gamma}_H|\\ &= \alpha^{Hn(n-1)}|\tilde{\Gamma}_H| \end{aligned}$$

The minimum of Eq.(2.62) 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 $\hat{a} = a(\hat{H})$.

2.4.4 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 [4], several possible approaches to approximating the Gaussian likelihood function are discussed, among them the well known Whittle's approximate MLE.

In our case the focus is on the properties of the covariance matrix Γ_H, and we try to take advantage of its special structure and to find efficient approximations for its inverse and determinant.

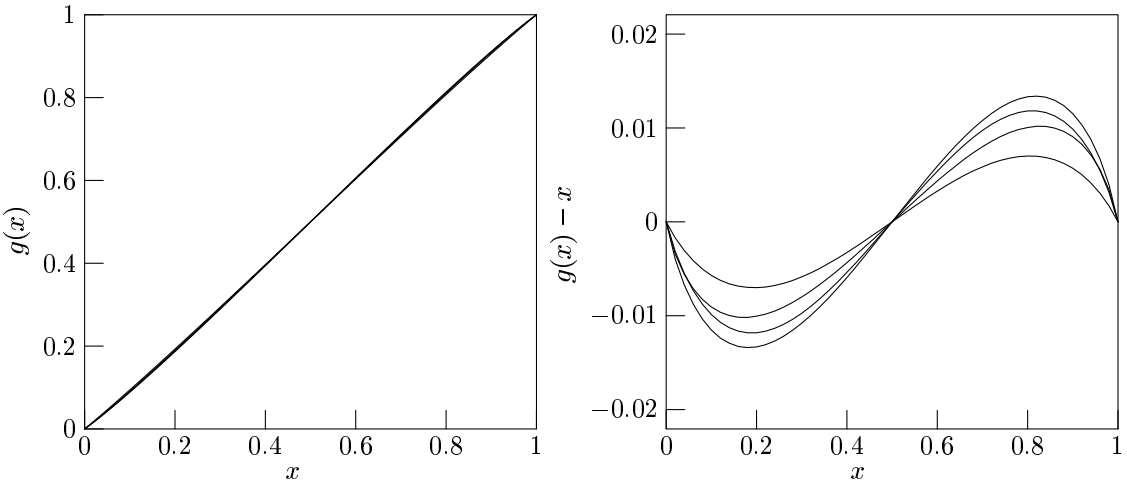

Figure 2.1: (a) The graphs of $g(x)$ for $H = 0.6$, 0.7, 0.8 and 0.9.; (b) Error of approximation $g(x) \approx x$ for $H = 0.6$, 0.7, 0.8 and 0.9.

Linear approximation of $g(x)$

Using the notation

$$g(x) = \frac{1}{2}\left(1 + x^{2H} - (1-x)^{2H}\right), \tag{2.64}$$

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

$$[\tilde{\Gamma}_H]_{i,j} = \alpha^{-H|i-j|} g(\alpha^{|i-j|}), \quad i,j = 1,2,\ldots,n. \tag{2.65}$$

It is interesting to note, that $g(x)$ is nearly completely linear for $x \in (0,1)$. Figure 2.1 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 $\tilde{\Gamma}_H$ can be approximated as $\tilde{\Gamma}_H \approx \mathbf{R}$, where $\mathbf{R}$ is a Toeplitz-type matrix of the form

$$[\mathbf{R}]_{ij} = \gamma^{|i-j|}, \quad i,j = 1,2,\ldots n, \tag{2.66}$$

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

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

Using the approximation $\tilde{\Gamma} \approx a\mathbf{R}$, the inverse of $\mathbf{R}$ can be easily calculated as [46]

$$\mathbf{R}^{-1} = \frac{1}{\frac{1}{\gamma} - \gamma} \begin{pmatrix} \frac{1}{\gamma} & -1 & 0 & \cdots & 0 \\ -1 & \gamma + \frac{1}{\gamma} & -1 & \ddots & \vdots \\ 0 & -1 & \gamma + \frac{1}{\gamma} & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & -1 \\ 0 & \cdots & 0 & -1 & \frac{1}{\gamma} \end{pmatrix}, \tag{2.67}$$

and the determinant of $\mathbf{R}$ is given by [46]

$$|\mathbf{R}| = (-1)^{n-1} \prod_{i=1}^{n-1} \begin{vmatrix} \gamma^{1-i} & \gamma^{i-1} \\ \gamma^{-i} & \gamma^{i} \end{vmatrix} \cdot \gamma^{n-1} = (1-\gamma^2)^{n-1}. \tag{2.68}$$

Approximate MLE

From Eq.(2.60) with $\tilde{\Gamma} \approx a\mathbf{R}$ the approximate MLE $\hat{m}$ is

$$\hat{m}(H) \;=\; \frac{\mathbf{t}^{\mathrm{t}}\,\mathbf{D}\mathbf{R}^{-1}\mathbf{D}\,\mathbf{X}}{\mathbf{t}^{\mathrm{t}}\,\mathbf{D}\mathbf{R}^{-1}\mathbf{D}\,\mathbf{t}}. \tag{2.69}$$

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

$$\hat{m}(H) = X(1). \tag{2.70}$$

Similarly, from Eq.(2.61) we get

$$\hat{a}(H) \;=\; \frac{1}{n}\left(\mathbf{X}^{\mathrm{t}}\,\mathbf{D}\,\mathbf{R}^{-1}\mathbf{D}\,\mathbf{X} - X_1^2\right). \tag{2.71}$$

Finally, we have to minimize (see Eq.(2.62)

$$L(\mathbf{X}; H) = \frac{n-1}{n}\log\left(\alpha^{nH}(1-\alpha^{2-2H})\right) + \log\left(\mathbf{X}^{\mathrm{t}}\,\mathbf{D}\,\mathbf{R}^{-1}\mathbf{D}\,\mathbf{X} - X_1^2\right) \tag{2.72}$$

to get the estimate $\hat{H}$.

Note that due to the relatively simple structure of $\mathbf{R}^{-1}$ the matrix product term $\mathbf{X}^{\mathrm{t}}\,\mathbf{D}\,\mathbf{R}^{-1}\mathbf{D}\,\mathbf{X}$ can be calculated as

$$\begin{aligned}\mathbf{X}^{\mathrm{t}}\,\mathbf{D}\,\mathbf{R}^{-1}\mathbf{D}\,\mathbf{X} &= (\mathbf{D}\,\mathbf{X})^{\mathrm{t}}\,\mathbf{R}^{-1}(\mathbf{D}\,\mathbf{X}) \qquad (2.73)\\ &= 2\sum_{i=1}^{n-1}\frac{\gamma}{\gamma^2-1}\left(t_i^{-H}X_i\right)\left(t_{i+1}^{-H}X_{i+1}\right)+\frac{X_1^2+\left(t_n^{-H}X_n\right)^2}{1-\gamma^2}+\sum_{i=2}^{n-1}\frac{1+\gamma^2}{1-\gamma^2}\left(t_i^{-H}X_i\right)^2\\ &= 2\sum_{i=1}^{n-1}\frac{\gamma\alpha^{-H(2i-1)}}{\gamma^2-1}X_iX_{i+1}+\frac{X_1^2+\alpha^{-2H(n-1)}X_n^2}{1-\gamma^2}+\sum_{i=2}^{n-1}\frac{(1+\gamma^2)\alpha^{-2H(i-1)}}{1-\gamma^2}X_i^2\end{aligned}$$

It should be noted that though the linear approximation $g(x)\approx x$ is rather accurate, the resulting inverse matrix $\mathbf{R}^{-1}$ of Eq.(2.67) is rather poor an approximation to $\tilde{\Gamma}^{-1}$ for large n. Nevertheless, the use of $\mathbf{R}^{-1}$ in the log-likelihood function Eq.(2.72), as we will see, yields a good estimate for H, while the estimate of a suffers more.

Improved approximation for $\tilde{\Gamma}^{-1}$

Since the matrix $\tilde{\Gamma}$ is a Toeplitz-type matrix with decreasing elements as we go farther from the diagonal, we expect that its inverse can be well approximated with a band matrix of the form:

$$\mathbf{C}=\begin{pmatrix}
\hat{\mathbf{c}}_{\mathbf{1}} & \hat{\mathbf{c}}_{\mathbf{2}} & \hat{\mathbf{c}}_{\mathbf{3}} & \cdots & \hat{\mathbf{c}}_{\mathbf{p-1}} & c_p & 0 & \cdots & 0\\
\hat{\mathbf{c}}_{\mathbf{2}} & \hat{\mathbf{c}}'_{\mathbf{1}} & c_2 & \cdots & c_{p-2} & c_{p-1} & c_p & \ddots & \vdots\\
\hat{\mathbf{c}}_{\mathbf{3}} & c_2 & c_1 & \cdots & c_{p-3} & c_{p-2} & c_{p-1} & \ddots & 0\\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & c_p\\
\hat{\mathbf{c}}_{\mathbf{p-1}} & c_{p-2} & c_{p-3} & \cdots & c_1 & c_2 & c_3 & & \hat{\mathbf{c}}_{\mathbf{p-1}}\\
c_p & c_{p-1} & c_{p-2} & \cdots & c_2 & c_1 & c_2 & \ddots & \vdots\\
0 & c_p & c_{p-1} & \cdots & c_3 & c_2 & c_1 & \ddots & \hat{\mathbf{c}}_{\mathbf{3}}\\
\vdots & \ddots & \ddots & \ddots & & \ddots & \ddots & \ddots & \hat{\mathbf{c}}_{\mathbf{2}}\\
0 & \cdots & 0 & c_p & \hat{\mathbf{c}}_{\mathbf{p-1}} & \cdots & \hat{\mathbf{c}}_{\mathbf{3}} & \hat{\mathbf{c}}_{\mathbf{2}} & \hat{\mathbf{c}}_{\mathbf{1}}
\end{pmatrix}. \qquad (2.74)$$

so that $\tilde{\Gamma}_H^{-1}\approx\mathbf{C}$. Our aim is to set the p parameters $c_1,\ldots,c_p$ to get $\mathbf{C}\tilde{\Gamma}_H\approx\mathbf{E}$. For example, this can be achieved by solving the equation

$$(c_p,\ldots,c_2,c_1,c_2,\ldots,c_p)\cdot\Gamma=(0,\ldots,0,1,0,\ldots,0), \qquad (2.75)$$

where $\mathbf{G} = (\tilde{\Gamma}_H)_{(2p-1)\times(2p-1)}$ and from this we have

$$c_i = \mathbf{G}^{-1}_{p(p+i-1)}, \quad i = 1, 2, \ldots, p. \tag{2.76}$$

With this approximation we only need to calculate the inverse of a $(2p-1) \times (2p-1)$ matrix[1]. To improve the approximate inverse, its elements in the upper-left and lower-right corners can be further corrected by solving the following set of equations:

$$\begin{aligned}
\mathbf{C}_{1,\cdot} \cdot (\tilde{\Gamma}_H)_{\cdot,1} &= (1, 0, \ldots, 0), \\
\mathbf{C}_{2,\cdot} \cdot (\tilde{\Gamma}_H)_{\cdot,2} &= (0, 1, 0, \ldots, 0), \\
&\vdots \\
\mathbf{C}_{p,\cdot} \cdot (\tilde{\Gamma}_H)_{\cdot,p} &= (\underbrace{0, \ldots, 0}_{p-1}, 1, 0, \ldots, 0),
\end{aligned} \tag{2.77}$$

where the notations $\mathbf{C}_{i,\cdot}$ and $(\tilde{\Gamma}_H)_{\cdot,j}$ denote the i-th row and j-th column of the matrices respectively.

Note, that to get $\mathbf{C}$ we need to calculate the inverse of a $(2p-1) \times (2p-1)$ matrix instead of an $n \times n$ matrix. As we increase the accuracy of the approximation (i.e., $p \to n$), we pay for this with increased calculation complexity. The corner-correction helps a lot when p is small.

Approximation for Σ_H^{-1} in case of linear sampling

However, to calculate the inverse and the determinant of Σ_H (see Section 2.3) the same problems arise as in the case of geometrical sampling with the covariant matrix $\tilde{\Gamma}_H$. Since Σ_H is also a Toeplitz type matrix, the same method as described in this section can be used to approximate Σ_H^{-1} with $\mathbf{C}$ of Eq.(2.74).

The only difference is that we need the determinant $|\Sigma_H|$, or at least an approximation of it, to calculate the estimate of H using Eq.(2.39). Fortunately, the determinant of $\mathbf{C}$ can be calculated (see Appendix B for the main idea and the solution for the $p = 2$ case), and we get $|\Sigma_H| \approx |\mathbf{C}|^{-1}$.

2.5 Validation with simulations

Computer simulations with Matlab were used to validate the proposed estimation methods and approximations, as well as to demonstrate the advantage of geometrical sampling over the traditional linear sampling scheme. In case of $\hat{H}$ the proposed estimation method was also compared to the wavelet-based method of Abry and Veitch [53] which is known to be a fast and efficient estimator.

This section describes the details and results of the simulations.

[1] More exactly—since the matrix is symmetric—, we only need to solve p equations.

Data sets

For small (up to 10^3) sample sizes the fBm samples were generated using the fact $\mathbf{Z} \sim \Gamma_H^{1/2}\mathbf{N}$ (or, correspondingly, $\mathbf{Z} \sim \Sigma_H^{1/2}\mathbf{N}$ 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 [35] was used. This method provides a fast and accurate approximation of fractional Brownian motion. (The RMD-mn simulator is available over the Internet [35].)

In the simulations 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.

The 95% confidence interval was obtained by repeating the simulations 100 times and calculating the sample variance of the estimates.

Parameter α

Using geometrical sampling the traffic is sampled at time instances

$$t_i = \alpha^{i-1}, \quad i = 1, 2, \ldots, n, \quad 0 < \alpha < 1. \tag{2.78}$$

Parameter α controls the number of samples and the total sampling time. (Note that Eq.(2.78) implies $t_1 = 1$ and $t_n = \alpha^{n-1} < 1$ resulting in decreasing t_i values. This reverse order of samples is used only to make the notations more simple. The total sampling time is also normalized to unity in this way.) 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 cannot 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^{-6} in all cases, except when the method was compared to the A-V estimator. The resolution in the latter case was 10^{-4}.

2.5.1 Geometrical vs. linear sampling

Estimates of H

Figure 2.2 shows the results of H estimates as a function of the number of sample points using both geometrical and linear sampling. In the geometrical case Eq.(2.72) was minimized while for the linear sampling we used the formula Eq.(2.39), where the inverse of Σ_H was approximated with the band matrix of type Eq.(2.74) with $p = 2$.

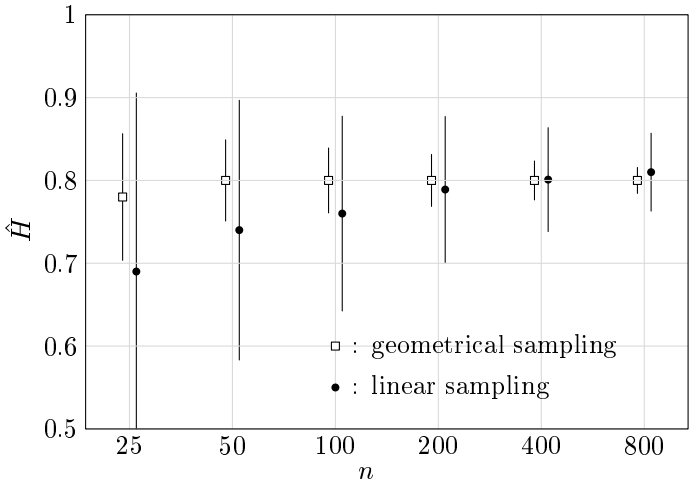

Figure 2.2: Estimates of H using geometrical and linear sampling.

The results show that the estimates using geometrical sampling have much smaller variance and are unbiased for sample sizes larger than 25. However, the variance of the estimates is always higher with linear sampling. For example, the variance for 800 samples using linear sampling is nearly the same as for only 50 geometrically sampled points.

Estimates of a

The next question was how the two different sampling methods affect the estimates for the variance parameter a. Figure 2.3 displays the results, assuming H is known. These simulations were useful to test whether our approximations in calculating the inverse and determinant of the covariance matrices are adequate or not. Figure 2.3 presents two different approximations for the geometrical sampling. First, the simple approximate inverse covariance matrix of Eq.(2.67) in Eq.(2.23) with Eq.(2.59) was used (denoted by light gray dots and labeled 'linear approximation' in the figure). As can be seen, the estimates of a are strongly biased and the bias is getting larger as the number of samples increases. So this estimate is clearly inadequate, therefore the approximation had to be refined. Next, we used the approximation of Eq.(2.74) for $\tilde{\Gamma}_H^{-1}$ with five parameters ($p = 5$). As we see from Figure 2.3, the strong bias from the $\hat{a}$ estimates disappeared and the variance of the estimates is only slightly higher than the theoretical value which can be calculated using Eq.(2.26). (Note, however, that the bias for sample sizes of 400 and 800 seems to be slightly increased.) Finally, the linear sampling method was used. Its estimates are asymptotically unbiased and have approximately the same variance as expected. The approximate inverse matrix used was as in Eq.(2.74) with only two parameters ($p = 2$).

Figure 2.4 shows the MLE $\hat{a}$ estimates without any *a priori* knowledge about the model

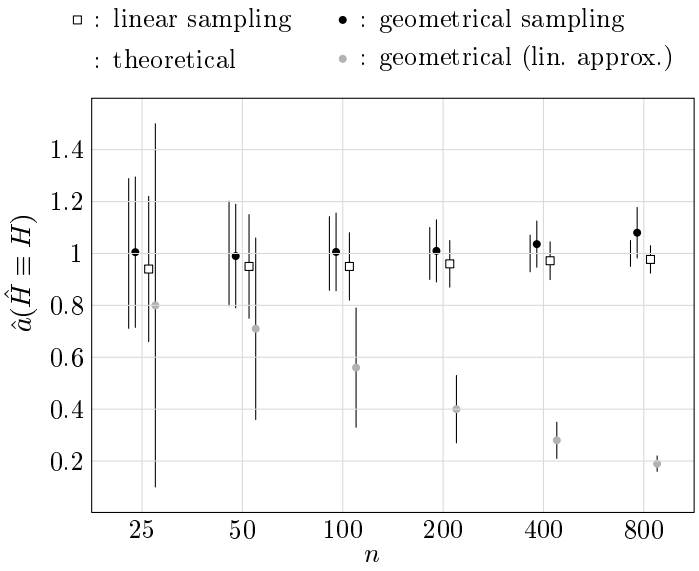

Figure 2.3: Estimates of a using geometrical and linear sampling and different approximations, assuming H is known.

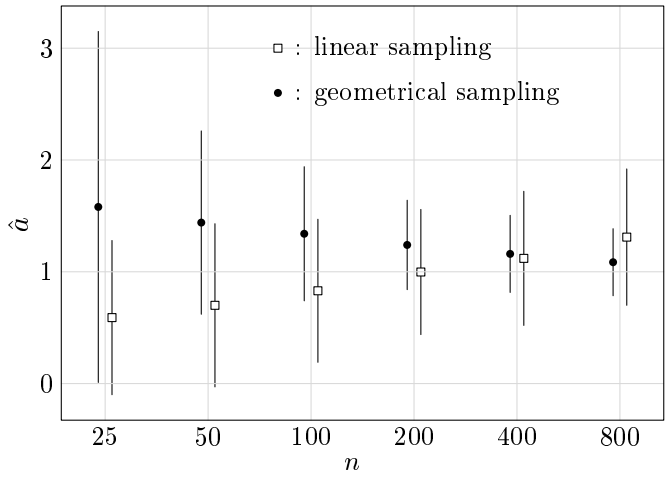

Figure 2.4: Estimates of a (when H is also estimated) using geometrical and linear sampling and different approximations.

parameters. All the approximations used here were the same as in the previous cases. Since H is not known and can only be estimated with a given variance, the estimates of a have larger variances than in the previous simulations. The question is how robust those estimates are when $\hat{H}$ can have a slight bias (see Figure 2.2). As for the geometrical sampling, the bias of $\hat{a}$ gets smaller and its variance is also decreasing rapidly as the sample size increases. On the other hand, for the linear sampling case the estimates seem to be biased for larger sample sizes and their variance does not seem to decrease. The reason for this behavior lies in the fact that the linear sampling for estimating H is less accurate than the geometrical sampling. The bias in $\hat{H}$ together with its higher variance is responsible for the bias and variance of $\hat{a}$, even if the linear sampling seems to be a better choice to estimate a than the geometrical one for known H (see Figure 2.3).

Estimates of m

As for the ML estimates for m the geometrical sampling does not give any extra advantage or disadvantage compared to the linear sampling. In fact, the approximate ML estimate gives back the sample mean as an estimate (see Eq.(2.70)). For example, using 400 samples the estimate $\hat{m}$ was 1.05 with variance 0.9.

2.5.2 Geometrical sampling vs. wavelet-based method

The focus here is on the H parameter estimation only.

Wavelet-based estimator

Here we use the notations of Veitch and Abry [53]. The continuous wavelet decomposition consists of the collection of coefficients

$$\{T_X(a,t) = \langle X, \psi_{a,t} \rangle,\ a \in \mathbb{R}^+, t \in \mathbb{R}\} \tag{2.79}$$

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

$$\left\{ \psi_{a,t}(u) \equiv \frac{1}{\sqrt{a}} \psi_0 \left(\frac{u-t}{a} \right),\ a \in \mathbb{R}^+, t \in \mathbb{R} \right\}. \tag{2.80}$$

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

$$X(t) \rightarrow \Big\{ \{a_X(J,k),\ k \in \mathbb{Z}\}, \{d_X(j,k), j = 1, \ldots, J,\ k \in \mathbb{Z}\} \Big\}. \tag{2.81}$$

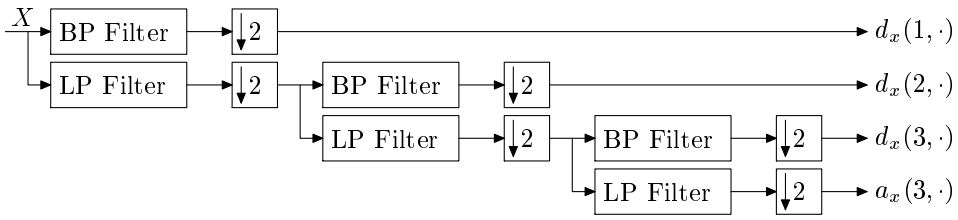

Figure 2.5: Filter bank

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 Figure 2.5) with low computational cost [13]. (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 [14], i.e.,

$$d_X(j,k) \stackrel{d}{=} 2^{j(H+0.5)} d_X(0,k) \tag{2.82}$$

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

$$\mathrm{E}\left[d_X(j,k)^2\right] = 2^{j(2H+1)}\, C(H,\psi_0), \tag{2.83}$$

with a constant $C(\cdot)$ dependent on H and the mother-wavelet, but 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 $\mathrm{E}\left[d_X(j,\cdot)^2\right]$

$$\mu_j = \frac{1}{n_j} \sum_{k=1}^{n_j} d_X(j,k)^2, \tag{2.84}$$

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 $\mathrm{E}\left[d_X(j,\cdot)^2\right]$ [1].

The power-law form in Eq.(2.83) 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 [53] is a weighted linear regression of $\log_2(\mu_j)$ on j.

Simulation results

The MLE method using geometrical sampling was compared to the wavelet-based Abry-Veitch estimator. Figure 2.6 shows similar results as previously, namely that the estimates

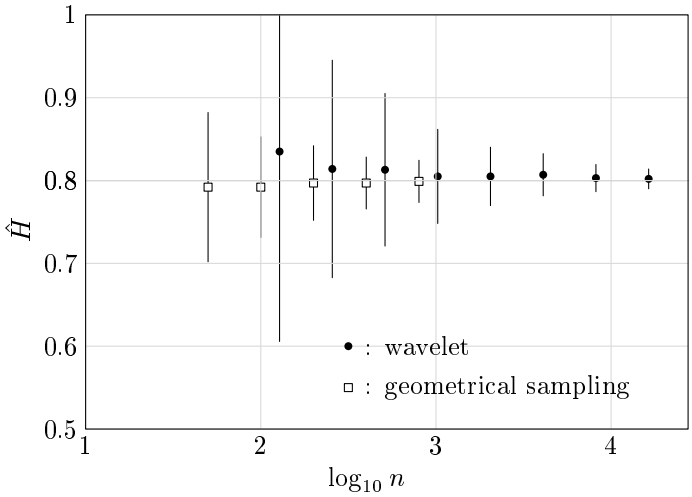

Figure 2.6: Estimates of H using MLE with geometrical sampling, and the wavelet-based Abry-Veitch estimator.

using geometrical sampling have much smaller variance and bias than the wavelet-based estimates. 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.

2.6 Conclusion

In this section the ML estimation of the parameters of fractional Brownian traffic was investigated. Closed form expressions for the estimators of m and a were derived as well as the log-likelihood function from which the estimator of H can be obtained. 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 idea of using geometrical sampling was also investigated. The intention with this sampling was to reduce the number of sampling points required for a given predefined confidence level. Intuitively, the geometrical sampling distributes the sampling points advantageously at different time scales, whereas linear sampling stresses the finest time scale.

The experiments with simulated traffic showed that the proposed approximate methods worked well, and 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.) For the estimation of a the geometrical sampling does not give any direct advantage, but as the estimator $\hat{a}$ actually depends on the estimator $\hat{H}$, the overall accuracy obtained is better. For the estimation of m, different sampling schemes give essentially the

same result, the estimate is basically the observed average rate.

It should, however, be noted that the experiments in this section were made only with simulated traffic with 'exact measurements'. If the measured values were 'noisy' in a sense that the traffic deviates from the fBm modeling assumption, then the estimates could be distorted as well. The wavelet-based Abry-Veitch estimator, for example, shows good performance in terms of robustness and computational cost. Similar investigations are necessary for the proposed method and are the topic of future study. As a first step, to get a glimpse how the proposed method works in practice, in Section 3.1.2 the approximate MLE with geometrical sampling is applied for estimating H of measured traffic.

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.

Chapter 3

Characterization and Control of Scaling Traffic

Practical questions and problems are investigated in this chapter, which are to be considered when using the self-similar modeling concept in case of real traffic.

Once self-similar processes are chosen to model the traffic, the most important question is the detection of self-similarity and the estimation of its parameter H. When performing the estimation, certain pitfalls must be avoided in any costs. In Sections 3.1 and 3.2 empirical studies are presented in this topic. In Section 3.3 the impacts of different network mechanisms (i.e., shaping, policing and multiplexing) on the H parameter estimates are investigated. Finally, in Section 3.5 queueing and shuffling analysis are performed in order to investigate the effects of scaling behaviour on cell loss in a simple queueing environment.

3.1 Hurst parameter estimation of real traffic

The estimation of the Hurst parameter is not easy in practice. The problem is that we always deal with finite data sets so it is not possible to check whether by definition a traffic trace is self-similar or not. Therefore it is only reasonable to speak about *self-similar behavior over a given time scale for a given data set.* I will refer to this phenomenon as "scaling property" in the following.

In practice, using measured data sets the estimated values of H obtained from different analysis methods are influenced by several factors (e.g., estimation technique, sample size, time scale, nonstationary traffic fluctuations). Four different H estimation method were investigated: the variance-time analysis, the closely related Index of Dispersion for Counts (IDC) analysis, the R/S analysis and the periodogram-based method.

3.1.1 ATM traffic measurements

Results reported here were obtained by statistical analysis of measured data. The measurements were made on the FUNET ATM WAN network.

The FUNET network

'FUNET' stands for 'Finnish University and Research Network', which provides primarily Internet services to its members based on TCP/IP-protocol. All these services are provided by CSC—Center for Scientific Computing which is a national service center that specializes in scientific computing and data communications providing modeling, computing and information services for universities, research institutes and industry. (The FUNET long-distance network was built on Telecom Finland's (now Sonera) ATM network at the time when the measurements were made.) All the Nordic national networks (FUNET, DENnet, ISnet, SUNET and UNINETT) are connected to the Nordic Backbone Network (NORDUnet) which has a connection point in Stockholm, Sweden. NORDUnet has connections to the US backbones, the European backbones and to networks in central and eastern Europe [32].

Measuring tool and configuration

The measurement was made at the CSC in Espoo, Otaniemi in 1996. This location is in the logical center of the whole network. All the international links start from here, including the main crosslink to Stockholm. Our measurement equipment was inserted between the network and the high-capacity ATM switch situated in Espoo (see Figure 3.1). From that point all the ATM traffic from the FUNET network transported through the switch and the traffic generated at the CSC and transmitted to the rest of the world could be monitored. The measurements were made by an HP Broadband Series Test System equipment. During our work only the cell capture capability of the measurement unit was used: 131,072 ATM cells mapped into a 155 Mbps SONET/SDH signal can be recorded into the 8MB of capture memory of this equipment. All cells are timestamped with the calendar time with resolution 0.1 μs.

The measured data

The aggregated traffic at the most heavily loaded point of the FUNET network was measured, including Internet traffic, data transfer and supercomputer usage. During the measurements, two types of data collections were made. In the first scenario the measured data was the time stamp of the arrival time instant for every single cell on the link. Because of the upper limit for the number of captured cells each measured data file contains 131,072 time stamps only, which corresponded to about 3–5 seconds according to the network load. For the long-term analysis longer measurement periods were needed, so in the

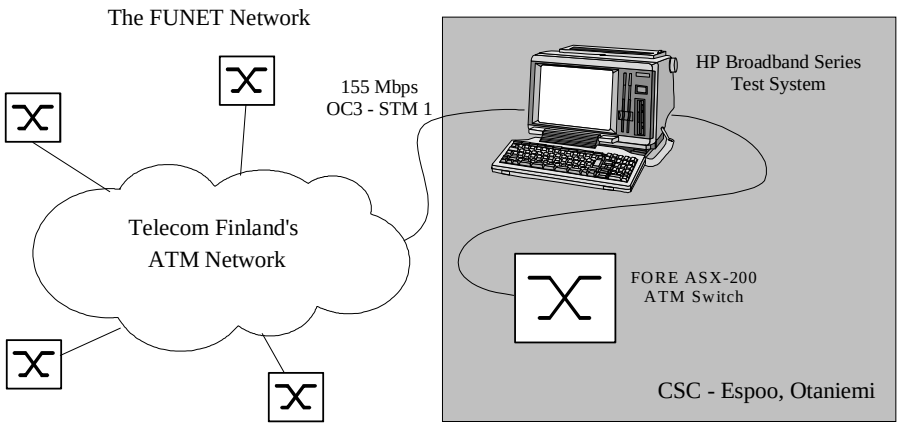

Figure 3.1: The FUNET measuring configuration.

Table 3.1: Quantitative description of the measured data sets.

Filename	#cells	Time (sec)
FUNET1	131,072	3.9
FUNET2	131,072	5.1
FUNET3	131,072	4.4
FUNET4	131,072	6.4
FUNETSTA.T3	14,807,546	425
FUNETSTA.T4	43,768,430	1964

second measurement scenario the recorded data was the number of cells received in a one second interval. In this case the time interval of the observation could take several minutes long. A summary of these data sets is given in Table 3.2. The files FUNET1, FUNET2 and FUNET3 contain traffic data captured from the incoming traffic from the whole country to the CSC, and the FUNET4 measurement was made on the outgoing link. In the case of the last two measurements in Table 3.1, the registered data was the number of cells received in every second on the incoming link. The average traffic load was about 14 Mbps for the first three measurements, and about 8 Mbps in the case of the FUNET4 data. In the following, we refer to the data above as the 'FUNET measurements'.

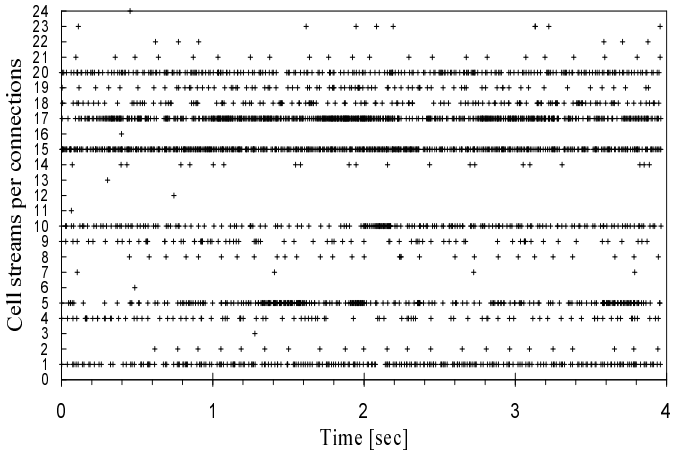

Figure 3.2: The structure of the FUNET1 traffic trace.

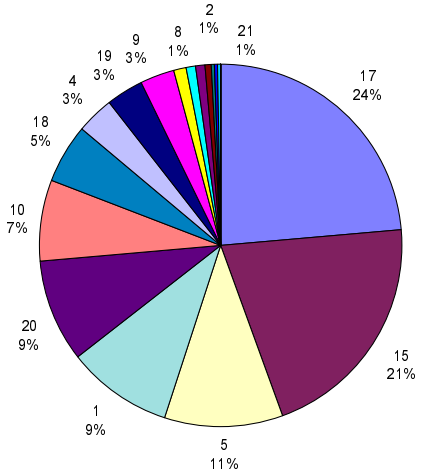

Figure 3.3: The bandwidths of connections in the FUNET1 data.

The traffic structure

As for the first four data sets, the measurement unit was able to register the VPI and VCI fields from the cell headers, too. Using this extra information we can reveal the structure of the aggregated traffic stream. Comparing the VPI/VCI fields the aggregated cell stream can be divided into independent connections. (Note that connection means a cell stream with common VPI and VCI fields in the headers. We do not have any information about the type of traffic carried by these cell streams.) The most important piece of information for us is the number of connections and their relative cell rate compared to each other. A detailed analysis was made for the FUNET1 data set. Figure 3.2 shows the separated cell streams schematically. (Because of the huge number of cells each hair-cross represents every 50th cell arrival in a connection.) As can be seen from the figure, during the 4 second measured time period 24 connections were in progress. The connection with highest rate contains about 30 000 cells which is about 24% of the whole aggregated traffic as well as the first dozen with highest intensity contain 99% of all the cells. Figure 3.3 shows the pie-chart of bandwidths of connections.

In our investigation the question of stationarity is fundamental. As far as it can be concluded from Figure 3.2 without a comprehensive stationarity analysis, the cell streams are homogeneous enough in time apart from the bursty nature of ATM traffic. There is no connection turned on or off in the middle of the measurement time and the rates apart from the burstiness are not changing considerably.

3.1.2 Hurst parameter estimation techniques

Our modeling assumption is that the samples are taken from the arrival process given by

$$A(t) = mt + X(t), \quad t > 0 \tag{3.1}$$

where $m > 0$ is the mean arrival rate, and $\{X(t), t \in \mathbb{R}\}$ is a H-sssi (or at least asymptotically self-similar) process with variance $\mathrm{Var}[X(1)] = \sigma^2$. The task is to estimate the H parameter of the self-similar component process. Next, five different statistical tests are presented.

Exact Gaussian MLE using geometrical sampling

The previously proposed exact Gaussian MLE using geometrical sampling (see Section 2.4) can be applied to estimate H only. Here we assume that $\{X(t), t \in \mathbb{R}\}$ in Eq.(3.1) is a fractional Brownian motion. Based on this assumption and a geometric sampling sequence, $\hat{H}$ is given as the minimizing argument of Eq.(2.62) on page 21.

The measured FUNET data was re-sampled geometrically, where 100 sample points were chosen with resolution 10^{-5}, i.e., the difference between the closest two measurement time instants was 10^{-5} while the total measurement time was normalized to unity. For each FUNET data set two such geometric sample series were generated. The time interval between the points is increasing exponentially in the first case, while in the second case the measuring time instants are getting denser towards the end of the measurement time. (I.e., the series of sampling instants are reversed in time.)

Table 3.2 shows the estimated $\hat{H}_{mle}$ values. The estimates consistently fall in the range of values typical for this kind of network traffic. Since the estimation method was developed and justified only for exact fractional Brownian traffic, no further conclusions can be drawn from the results obtained from real measurements. The estimated values can be validated by comparing them with results of other estimation methods. Note, however, that we used only 100 sample points for the estimation, which can be a great advantage of the method. The linear sampling scheme with 100 sample points was also examined as a comparison, but the method did not provide any results (i.e., the estimates of H were not even within the [0.5.1) interval.)

Index of dispersion for counts

This commonly used measure for capturing the variability of traffic over different time scales is provided by the *index of dispersion for counts* [12]. For a given time interval of length t, the index of dispersion for counts (IDC) is given by the variance of the number of arrivals $A(t)$ during the interval of length t divided by the expected value of the same quantity:

$$IDC(t) = \frac{\mathrm{Var}[A(t)]}{\mathrm{E}[A(t)]}. \tag{3.2}$$

Self-similar processes produce a monotonically increasing IDC (see Eq.(1.32) in Section 1.4) of the form

$$IDC(t) = m^{-1}t^{2H-1}\sigma^2. \tag{3.3}$$

Plotting $\log IDC(t)$ against $\log t$, this property results in an asymptotic straight line with slope $2H-1$ [31]. (This behaviour is in contrast to traditional processes where the IDC is bounded.)

For a finite data set, the variance of $A(t)$ can be calculated by dividing the whole series into nonoverlapping blocks of length t and treating them as different instances of $A(t)$.

Figure 3.4(a) depicts the IDC curve corresponding to the trace FUNET1. The sequence of cell counts in every 100μs interval was analyzed. The IDC curve for the FUNET1 file increases monotonically throughout a time span that covers 3–4 orders of magnitude and shows an asymptotic slope that is strictly different from the horizontal line and is estimated to be about 0.4, resulting in an estimate $\hat{H}$ of 0.7.

The same analysis was made for all the data sets. Table 3.2 shows the results: the values of the estimated Hurst-parameter $\hat{H}$. As can be seen from the table, the values of $\hat{H}$ are pretty much the same for all the data sets. It is remarkable that in the case of the last two data sets the analyzed process was the sequence of cell counts in each second instead of 100μs as in the case of the first four sets. In spite of the fact that the time scale was four orders of magnitude higher the Hurst-parameter remained the same.

Variance-time analysis

This method is based on the property that a self-similar process has slowly decaying variances (see Eq.(1.30) in Section 1.4). The so-called variance-time plot is obtained by plotting $\log \mathrm{Var}\,[A(t)/t]$ against $\log t$ and by fitting a simple least squares line through the resulting points in the plane, ignoring the small values of t. Values of the estimated asymptotic slope $\hat{\beta}$ between -1 and 0 suggest self-similarity, and the estimate for degree of self-similarity is given by $\hat{H} = 1 + \hat{\beta}/2$ [31].

The corresponding plot for the FUNET1 data set can be seen in Figure 3.4(b). The estimated values of $\hat{H}$ are listed in Table 3.2. Since the variance-time plots and the IDC diagrams are closely related statistical methods, the results obtained from this method are the same as in the previous subsection.

R/S analysis

This method tries to capture parameter H based on the rescaled adjusted range statistics. The R/S method is based on the followings.

Given an empirical time series $\{X_k : k = 1, \ldots, N\}$ of length N, the whole series is divided into K nonoverlapping blocks. Now, compute the so-called rescaled adjusted range $R(t_i, d)/S(t_i, d)$ for a number of values d, where $t_i = \lfloor N/K \rfloor (i-1) + 1$ are the starting

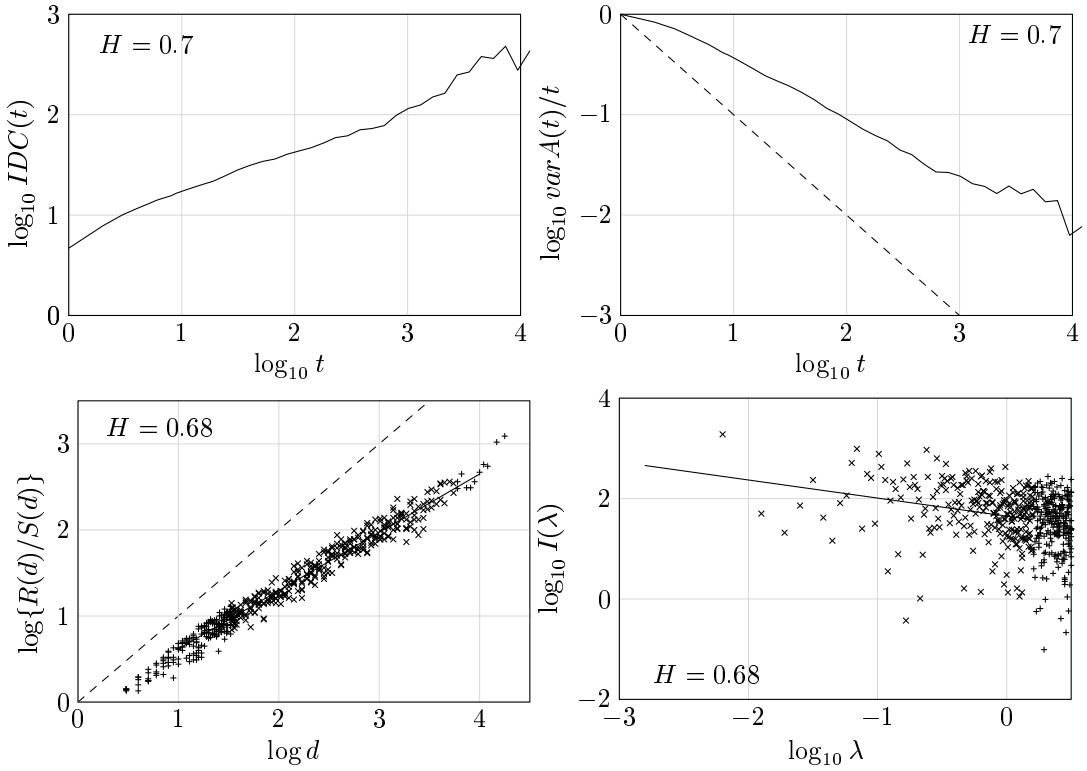

Figure 3.4: (a) IDC plot, (b) variance-time plot, (c) R/S diagram, and (d) periodogram plot for FUNET1 data.

points of the blocks which satisfy $(t_i - 1) + d \leq N$.

$$R(t_i, d) = \max\{0, W(t_i, 1), \ldots, W(t_i, d)\} - \min\{0, W(t_i, 1), \ldots, W(t_i, d)\}, \tag{3.4}$$

where

$$W(t_i, k) = \sum_{j=1}^{k} X_{t_i+j-1} - k \cdot \Big(\frac{1}{d}\sum_{j=1}^{d} X_{t_i+j-1}\Big), \quad k = 1, \ldots, d. \tag{3.5}$$

$S^2(t_i, d)$ denotes the sample variance of $X_{t_i}, \ldots, X_{t_i+d-1}$. For each value of d one obtains a number of R/S samples, which decreases from K for larger values of d. One computes these samples for logarithmically spaced values of d.

For a long range dependent data the R/S statistics is characterized by

$$\mathrm{E}\,[R(d)/S(d)] \sim \mathrm{const} \cdot d^{\alpha}, \tag{3.6}$$

as $d \to \infty$ and $0.5 < \alpha < 1$.

Plotting $\log R(t_i, d)/S(t_i, d)$ vs. $\log d$ results in the R/S plot. Next, a least squares line is fitted to the points, where both the R/S samples of the smallest and largest values of d are omitted. The slope of the regression line is an estimate for α [45].

In our case the analyzed data set is the increment process of $A(t)$ which is long-range dependent. Hence the slope of the regression line is the estimate of H, namely $\hat{H} = \hat{\alpha}$. Figure 3.4(c) shows the R/S plot for the FUNET1 data. The analyzed process was the sequence of cell counts in every 100μs. The estimated value of H for this data set is 0.68, which is nearly the same as the values calculated by the two previous methods.

The same analysis was made for all the FUNET measurement data sets (see Table 3.2).

Periodogram-based analysis

This method is used to identify the manifestation of self-similarity by frequency domain analysis of the measured data. Let $I(\cdot)$ denote the sample periodogram (i.e., power spectrum as estimated using a Fourier transform) defined by

$$I(\lambda) = \frac{1}{2\pi N}\left|\sum_{j=1}^{N} X_j e^{ij\lambda}\right|^2, \qquad \lambda \in [0, \pi). \tag{3.7}$$

As mentioned in Section 1.2, the spectral density of the increments of self-similar processes obeys a power law near the origin (see Eq.(1.21) in Section 1.2),i.e.,

$$I(\lambda) \sim \mathrm{const} \cdot |\lambda|^{\beta} \tag{3.8}$$

when $\lambda \to 0$ with $\beta = 1 - 2H$. Thus, the first idea to determine H is simply to plot the periodogram in a log-log grid, and to compute the slope of a regression line which is fitted

Table 3.2: The values of Hurst-parameter H calculated from different statistical methods

Filename	$\hat{H}_{mle}$	$\hat{H}_{idc}$	$\hat{H}_{var}$	$\hat{H}_{rs}$	$\hat{H}_{per}$
FUNET1	0.74, 0.75	0.7	0.7	0.68	0.68
FUNET2	0.72, 0.7	0.67	0.67	0.67	0.73
FUNET3	0.63, 0.76	0.66	0.66	0.68	0.68
FUNET4	0.77, 0.83	0.72	0.72	0.74	0.78
FUNETSTA.T3	-	0.70	0.70	0.82	0.94
FUNETSTA.T4	-	0.67	0.67	0.79	0.90

to a number of low frequencies. This should be an estimate of $1 - 2H$. In most of the cases this will lead to a wrong estimate of H since the periodogram estimation method is unbiased and inconsistent. However, this method can reveal the power spectrum near the origin. The periodogram plot is obtained by plotting $\log(I(\lambda))$ against $\log \lambda$.

Figure 3.4(d) presents the periodogram plot for the FUNET1 data set, where the analyzed time series was the number of cells in every 1 msec. The slope of the low frequency part—in the present context, the regression line was fitted to the lowest 50% of all frequencies—is clearly different from zero, the slope estimate is about -0.36 which yields $H = 0.68$. This result corresponds to the previously calculated values of H.

The analysis was made for all the data sets, and the results are listed in Table 3.2.

Summary

To summarize the results listed in Table 3.2, it can be concluded that:

- The estimated values of the parameter H are definitely greater than 0.5 for all cases.
- The values of H are nearly the same for all of the four analysis methods and for all the data sets. The common value of it is about 0.7. (Apart from the last two values for the FUNETSTA data sets.)

In spite of this, it would be too early to say that it follows from the results above that the measured traffic is self-similar with self-similarity parameter 0.7. To establish such a statement, one would need to carefully examine the structure of the analyzed data sets in more details.

In the next section, I investigate the problems arising during the estimation of the parameter H and determine those effects which can influence the results considerably.

3.2 Impacts on the Hurst parameter

In practice, using measured data sets the estimated values of H obtained from different analysis methods are influenced by the following circumstances:

- dependence on estimating technique;
- dependence on sample size;
- dependence on time scales;
- dependence on data structure.

Next, these points are investigated to see how the calculated self-similarity parameter is affected. The focus is on the robustness of $\hat{H}$ against these effects.

3.2.1 Estimating technique

In Section 3.1.2 four different statistical methods were presented testing for and estimating the degree of self-similarity, but this list is still far from being complete.

The absence of any results for the limit laws of the previously mentioned statistics make them inappropriate when a more refined data analysis is required. In contrast, a more refined data analysis is possible using periodogram-based methods in the frequency domain. Several periodogram-based estimators can be found in the literature, such as maximum likelihood type estimates (MLE) and related methods [55]. In particular, for Gaussian processes Whittle's estimate MLE has been studied extensively. Using these approaches, more information can be collected on the H-estimate, such as confidence intervals.

In practice, when the required preliminary conditions for the statistical tests are not fully satisfied, the different methods can give slightly different estimates of H. (As can be seen in Table 3.2.) To give a fully detailed analysis for all four statistical methods is beyond the scope of this study. Next, I choose one of them, the IDC plot for further investigation. This method, being simple and effective, requires small computation power and the resulting plot is descriptive enough to visualize the changes when the analyzed data set is modified.

3.2.2 Sample size

For a given time interval t the IDC value is defined by Eq.(3.2) in Section 3.1.2. For a self-similar process, this value is increasing without limit as t tends to infinity. In practice, we only have finite data sets and the value of t can not exceed the sample size. Furthermore, to get a reliable estimate of IDC(t) the maximum window size is limited approximately to the 10% of the sample size. (This is because using nonoverlapping windows of length t we need at least about 10 values to calculate the variance with acceptable confidence.) Thus, the calculated IDC(t) value is getting more and more inaccurate as t increases. As a result, the IDC plot becomes more and more noisy as t increases.

Table 3.3: Calculated values of Hurst-parameter H estimated from different subsets of the FUNET1 data.

FUNET1 (100%)	50%	25%
$H = 0.70$	$H_1 = 0.69$	$H_1 = 0.61$, $H_2 = 0.57$
	$H_2 = 0.70$	$H_3 = 0.71$, $H_4 = 0.62$

Example 1.1 In our analysis the critical issue is to examine the robustness of the estimated values of H against the sample size. To do this, the Hurst-parameter was estimated for different subsets of the whole FUNET1 data set. First the data was divided into two and four equal parts and the Hurst parameter was estimated in each case. Table 3.3 shows the calculated values.

Discussion 1.1 When the full data set was split into two parts, the estimates of H remained the same in both cases. However, when the data was split further, the estimated values for three subsets out of four differ from the original $\hat{H}$.

3.2.3 Time scales

In our case the analyzed process is the number of cells arriving in a fixed time interval of length t. The t represents the *time scale* at which the process is examined. Our task now is to find the appropriate time scales at which the IDC curve of the process increases monotonically. (As for the upper limit of t, the same holds as in the previous section, namely, the IDC is getting more and more inaccurate as t increases.)

Example 2.1 This terminology makes it possible to take advantage of the longer but not so detailed FUNETSTA measured data sets. Calculating the IDC values for FUNETSTA.T3 and plotting the result together with the IDC curve of the FUNET1 data set—taking into consideration the different time scales(!)—we get Figure 3.5. (Note, that there was a time gap between the FUNET1 and FUNETSTA.T3 measurements (about 30 min). Here we assume again, that the measured traffic was stationary.) Another drawback of bringing together the two plots is that the middle part of the curve is noisy and uncertain.

Discussion 2.1 Using a simple linear regression is a rough guess for this plot and estimating $\hat{H}$ gives the rather approximate value of 0.81. But the main result is that the curve increases monotonically throughout a time span that covers six orders of magnitude.

Example 2.2 In the previous example the plot tried to cover as many time scales as possible. Here I examine the Hurst parameter[1] when estimated at shorter but different time scales. To do this, the regression line was fitted to different parts of the IDC curve using a sliding window which covers only one decade of time. The resulted values of H estimated from $(t, 10t)$ interval and plotted at the middle of each interval can be seen on Figure 3.6.

[1]Strictly speaking, the term "Hurst parameter" should not be used when only a specific time scale is

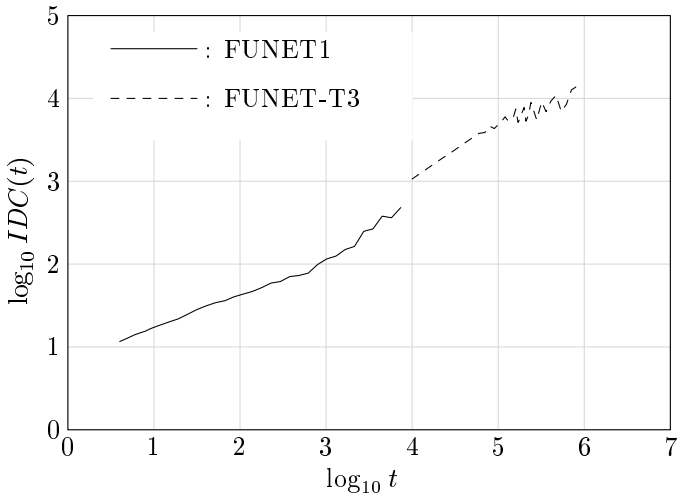

Figure 3.5: IDC plot for FUNET1 and FUNETSTA.T3 data sets.

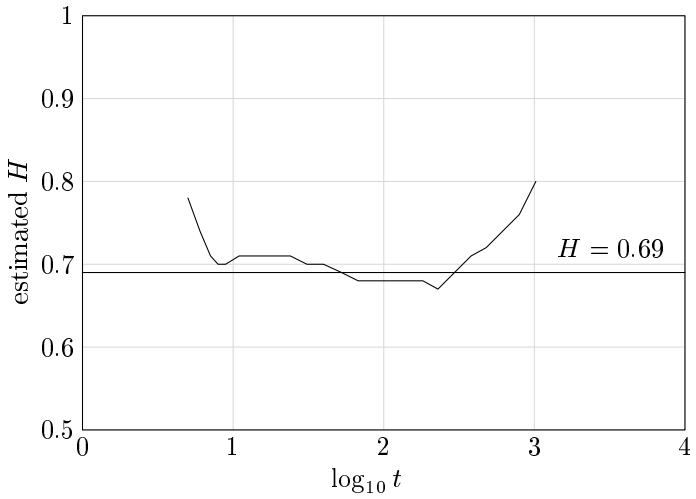

Figure 3.6: Hurst parameter estimates for different time scales.

Discussion 2.2 The results show that estimating the Hurst parameter at different time scales yields to values of $\hat{H}$ from 0.68 to 0.8. This is a clear indication of the importance of the problem of "what time scale to choose for Hurst parameter estimation" in practice. Moreover, the example also demonstrates that it is possible to have scaling traffic with different scaling parameters which should be distinguished from the Hurst parameter.

3.2.4 Level shifting

During the analysis I always assumed that the statistical properties of the measured cell stream are independent from time. This assumption is questionable, but—since the greatest part of statistical analysis methods require stationarity as a basic preliminary condition—it cannot be avoided. Strictly speaking, we assume that the measured process is stationary in the wide sense which means that its mean is finite and independent of time, and its autocorrelation function is finite and is invariant of time shift. (If we decided to treat our measured data sets as nonstationary sequences it would be almost impossible to make a comprehensive analytical study with meaningful results general enough to use elsewhere. Furthermore, in the case of finite data sets it is not possible to discriminate a stationary long-range dependent sequence from a nonstationary one.)

In this section the case when the assumption of stationarity does not hold was investigated (i.e., there is a level shift present in the measured traffic traces.) I examine how robust our statistical test is in case of a nonstationary cell sequence with a change in the mean as a function of time.

Example 3.1 In this first simple model nonstationarity is introduced by adding a CBR traffic to the second half of the measured data. (Note, that this example represents not just a theoretical problem but a possible event in practice: while measuring the network traffic suddenly a new source may start to emit cells with constant cell rate.) Figure 3.7 shows the calculated IDC plots for these new multiplexed data sets. 2.8 Mbps (CBR20—20% of the load of FUNET1) and 7 Mbps (CBR50—50% of the load of FUNET1) CBR rates were applied.

Discussion 3.1 The effect on the IDC plot is clearly visible. For the FUNET+CBR20 plot the upper part of the curve is moved up a bit as well as the lower segment shifted down slightly. As a result, the estimated Hurst parameter is greater, about 0.72. For the FUNET+CBR50 case the effect is sharper, $\hat{H}$ being 0.8.

Example 3.2 To understand the effects of level shift on the IDC plot more deeply, I investigated a simple CBR model in this example. As a starting point a CBR traffic trace is chosen with the same rate as the mean rate of the FUNET1 traffic. The nonstationarity was introduced by increasing the CBR rate by 10, 20 and 50 percent abruptly at half time of the investigated time period. The IDC(t) value for an ideal CBR source without jitter is zero for all t which cannot be plotted on a logarithmical scale. The calculated IDC plots for the CBR traces with level shift can be seen on Figure 3.8.

---

considered. Instead, the "scaling parameter of a given time scale" would be appropriate here.

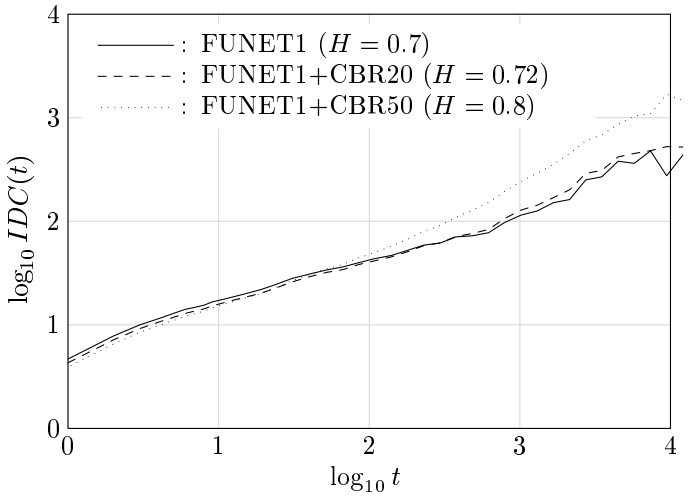

Figure 3.7: IDC plot for FUNET1 multiplexed with nonstationary CBR traffic.

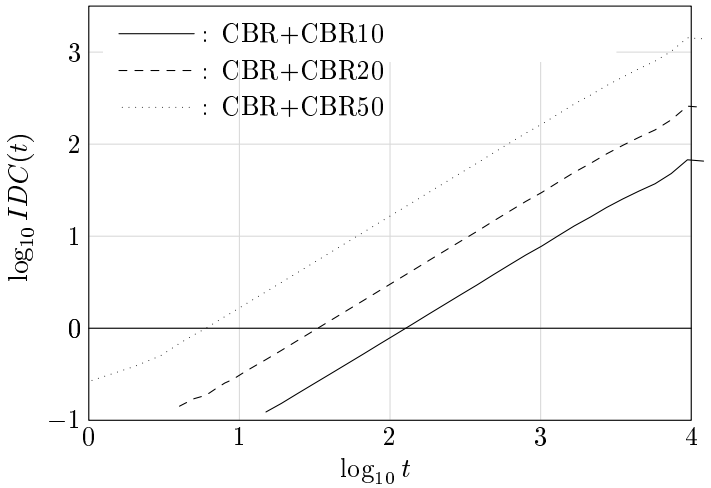

Figure 3.8: IDC plot for a CBR cell stream with level shift.

Discussion 3.2 As can be seen from the figure, all the IDC curves are straight lines with slope 1. The only difference is that the IDC values are higher when the level shift is stronger.

The simplicity of the examined CBR model makes it possible to calculate the IDC(t) values analytically. (In practice, calculating the IDC plot for a finite data set means evaluating a double sum to estimate the mean and the variance. The following results are derived from this IDC estimator.) The IDC(t) for the above data sets is of the form:

$$IDC(t) = \frac{(a_1 - a_2)^2}{2(a_1 + a_2)} t, \tag{3.9}$$

where a_1 and a_2 are the cell rates for the first and second half of the data respectively. For $a_1 \neq a_2$ the IDC plot is given by:

$$\log IDC(t) \simeq const + \log t, \quad \text{where} \quad const = \log \frac{(a_1 - a_2)^2}{2(a_1 + a_2)} \tag{3.10}$$

which gives us a straight line with slope 1.

The main result here is the fact that although the CBR data with level shift has nothing to do with self-similarity, the estimated IDC is a monotonically increasing straight line with slope 1.

Example 3.3 The first example is generalized here by replacing the CBR traffic with a Poisson process. Again, the FUNET1 data was modified by adding a Poisson traffic to the second half of the measured data to increase the mean rate by 20 and 50 percent. The calculated IDC plots can be seen in Figure 3.9.

Discussion 3.3 The effect of nonstationarity in the plots is the same as in Example 1. The upper part of the curves moved up and the lower-left segments are shifted down simultaneously, resulting in higher Hurst parameter estimates.

Example 3.4 To make the effect of level shifts on the IDC plot clearer, in this example a simple but inhomogeneous Poisson process is examined which changes its intensity in time. Here I consider the case when the Poisson source emits cells with rate λ_1 and suddenly changes its intensity to λ_2. Figure 3.10 presents the analysis result for these data sets. (For every process λ_1 was set to 1 and λ_2 changes as noted in the figure.)

Discussion 3.4 For such simple inhomogeneous Poisson processes the IDC estimate can be derived analytically. Let λ_1 and λ_2 denote the intensity parameters of the process for the two halves. Then, the IDC(t) value can be calculated as follows:

$$IDC(t) = 1 + \frac{(\lambda_1 - \lambda_2)^2}{2(\lambda_1 + \lambda_2)} t. \tag{3.11}$$

For the appropriate IDC plot for $\lambda_1 \neq \lambda_2$ and $t \to \infty$ we get:

$$\log IDC(t) \simeq const + \log t, \quad \text{where} \quad const = \log \frac{(\lambda_1 - \lambda_2)^2}{2(\lambda_1 + \lambda_2)}. \tag{3.12}$$

This equation gives a straight line with slope 1 as an asymptote.

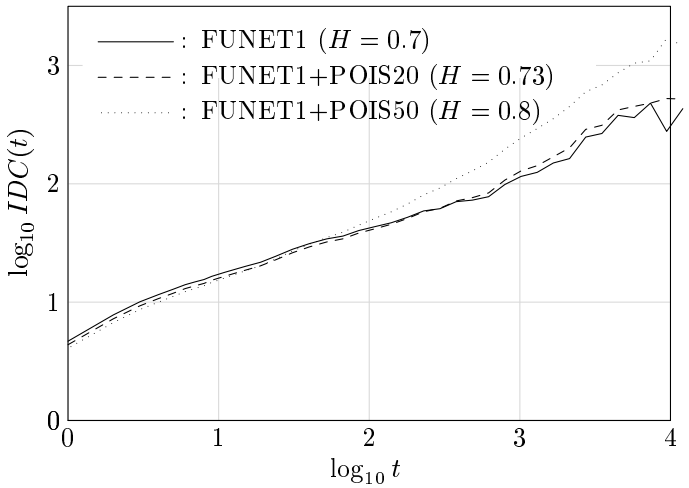

Figure 3.9: IDC plot for FUNET1 multiplexed with a nonstationary Poisson traffic.

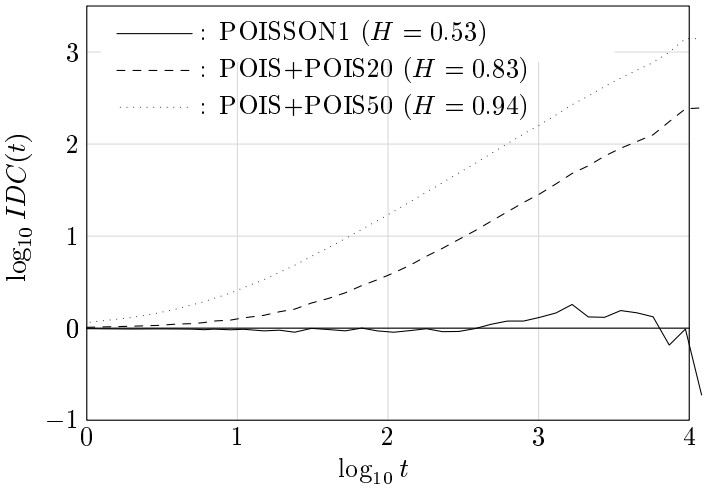

Figure 3.10: IDC plot for inhomogeneous Poisson processes

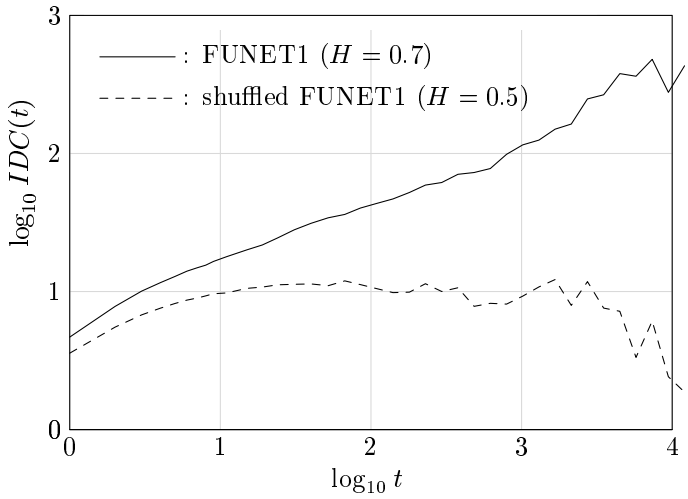

Figure 3.11: IDC plot for shuffled FUNET1 data.

The main message from this example is again that a monotonically increasing IDC does not necessarily come from the self-similar nature of the analyzed data. Instead, the reason behind can be the nonstationarity present in the sample trace. A linearly increasing IDC curve over many time scales (see Eq.(3.12)) can also be created even with a simple stationary Markovian model (e.g., with an Inhomogeneous Poisson Process in our example). In this case the increasing IDC curve again has nothing to do with self-similarity.

3.3 Impacts of network mechanisms on $\hat{H}$

In this section I examine how the estimated Hurst-parameter is affected when the structure of the measured data set is modified by shuffling, shaping, policing, and disturbing the cell stream.

Example 4.1 This first example is just a theoretical one without any practical meaning but gives us useful information about the correlation structure of the analyzed traffic. In order to show the difference from a short term correlated data a new data set was generated from the original FUNET1 data by mixing the sequence of cell interarrival times randomly thus building a new cell stream. The long-term correlations were obviously removed by this random shuffling which can be investigated in Figure 3.11.

Discussion 4.1 As for the shuffled FUNET1 data the IDC curve starts as in the case of the original plot but soon it stops increasing and remains constant for values of $\log_{10} t$

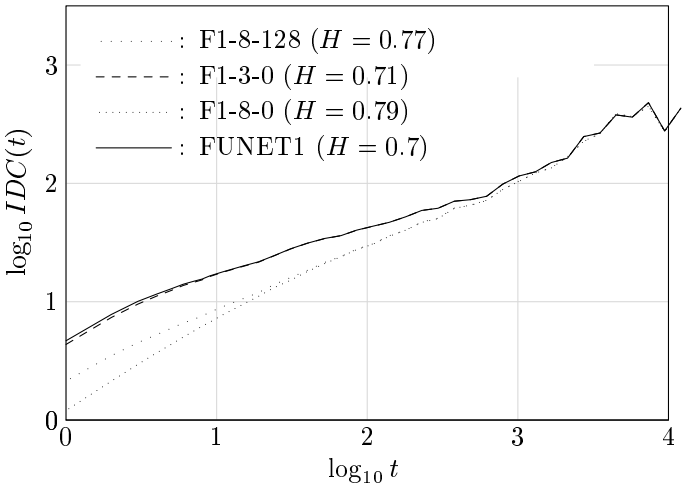

Figure 3.12: IDC plot for shaped FUNET1 data.

greater than 1. The self-similar feature disappeared but the resulting process is still bursty and possesses short-range dependence.

Example 4.2 This example investigates the effects of traffic shaping. Our shaping algorithm was the *leaky bucket shaping* which forces nonconforming cells to be delayed. Consider a leaky bucket with leak rate r and bucket size M. Cells which find the bucket content smaller than M are directly admitted to the network; otherwise, they are queued with FIFO discipline and admitted to the network with rate r.

The FUNET1 data (with average rate 33 072 cell/s which is about 14 Mbps) was shaped with parameters $M = 0$ (measured in cells), $r = 50$ Mbps (F1-3-0); $M = 0$, $r = 20$ Mbps (F1-8-0) and $M = 128$, $r = 20$ Mbps (F1-8-128). The analysis results are illustrated in Figure 3.12.

Discussion 4.2 The correlation of the shaped cell streams are slightly affected due to the shaping procedure. The IDC plot demonstrate the remained long-term correlation in the data. These results are consistent with the results reported in [39]. For the purpose of removing long-range dependence a drastic shaping is needed which means that shapers would have to use very large buffers which cannot be used in many applications due to the extreme introduced delay. The shaping effect, however, resulted in even higher values for the estimated Hurst parameter. The explanation and further discussion about this result is reported in Section 3.4.

Example 4.3 Instead of shaping, a policing algorithm was used which is nearly the same as our shaper but discards every nonconforming cell instead of delaying it. As a result, the policed cell stream contains less cells but will be conforming with the specified rate r and bucket size M. The parameters r and M were chosen as previously. The IDC

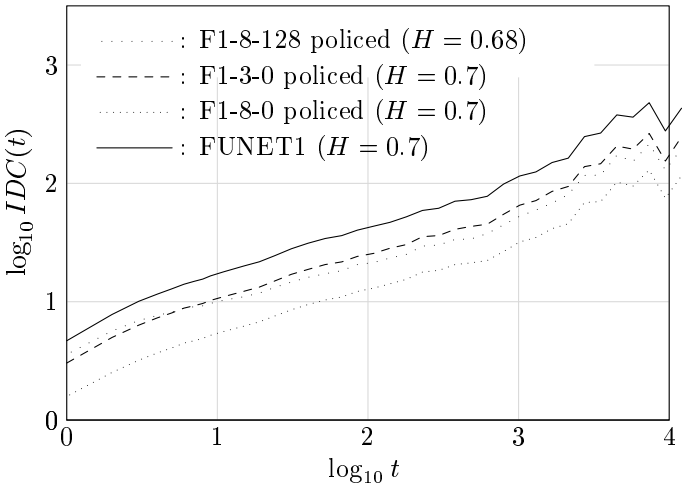

Figure 3.13: IDC plot for policed FUNET1 data.

plots for the policed cell streams can be seen in Figure 3.13.

Another interesting question is the structure of the discarded cell stream. Figure 3.14 represents the IDC curves for this kind of overflow process.

Discussion 4.3 As the Figures show, the IDC curves are almost the same in the case of both the policed cell streams and the overflow (dropped) cell streams with approximately the same Hurst parameter estimates. It can be concluded that the self-similarity feature is more robust for policing than for shaping. The results are consistent with the engineering intuition that FIFO queues behave like low-pass filters and the long-range correlations of the self-similar traffic (having power spectra divergence at low frequencies) are not affected.

Example 4.4 Finally, I examined how the traffic trace is modified when flowing through a multiplexing stage. The FUNET1 data (whose rate is about 9% of the link speed) was multiplexed with Poisson traffic with rate 90% of the link speed. Next, the IDC plot was calculated for those cells of the aggregated cell stream which cells belonged to the original FUNET1 data. The result can be seen on Figure 3.15.

Discussion 4.4 The results show that the IDC curve is only slightly modified by multiplexing a Poisson traffic with even a high load compared to our measured traffic. This also indicates the robust nature of the detected feature of self-similarity.

3.4 Discussion on the Hurst parameter estimates

In this section the problem of interpretation of estimated Hurst parameter is discussed based on the previous results.

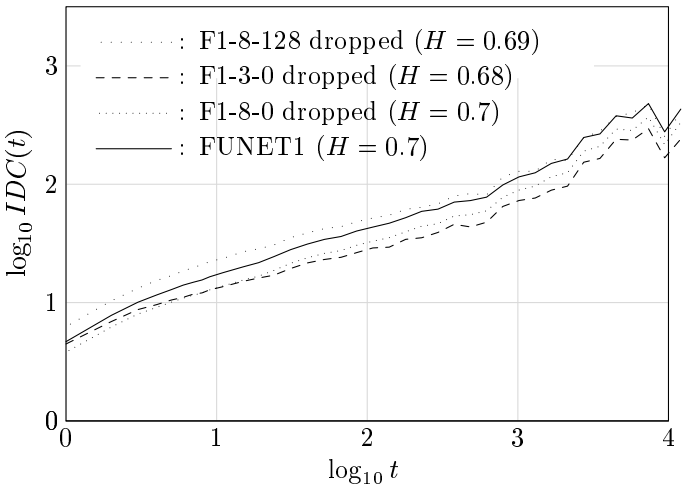

Figure 3.14: IDC plot for the overflow process of policed FUNET1 data.

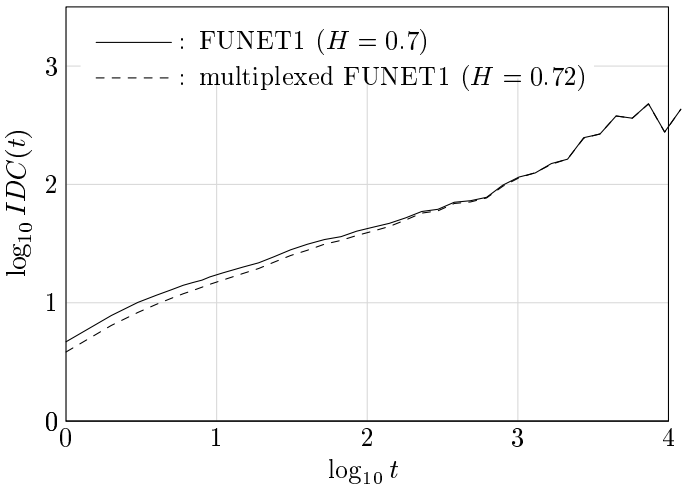

Figure 3.15: IDC plot for the multiplexed FUNET1 data.

The shaping example showed that the long-range dependence feature persisted but the estimated Hurst parameter was different and the direction of change in the value was very interesting. *The estimated Hurst parameter is increased due to shaping.* This can be explained as follows. On short time scales the shaper is effective and it smoothes the cell stream. That is why the variance of the number of cells in a given window is decreasing on short time scales which results in a shifting IDC curve to smaller values. However, on large time scales there is no significant effect of the shaping so there is no change in the IDC curve as can be seen in Figure 3.12. Therefore, it is obvious that the estimated Hurst parameter will be higher. This result is contrary to the usual interpretation of estimated Hurst parameter because the Hurst parameter is believed to be a measure of burstiness. For Poisson traffic, which is a smooth process, it is 0.5 and when increasing burstiness the Hurst parameter is increasing. However, our example shows that if we are smoothing the traffic the estimated Hurst parameter is increasing! If the process is a pure self-similar process there is a good interpretation of the estimated Hurst parameter, see e.g., [42]. However, it is not obvious what the interpretation of the Hurst parameter is in practice where the traffic structure is modified by several mechanisms (shaping, queueing, multiplexing, etc.) and the process is not a pure self-similar process. Can it be used as a burstiness measure? The investigations suggest the answer to be negative. Can we gain any information from the estimated Hurst parameter? Can we use it for dimensioning purposes?

The Hurst parameter could represent an important and compressed information (the degree of self-similarity) about a pure self-similar process but in practice when several effects modify the structure of the traffic they may distort the Hurst parameter estimates so much that its original meaning is hidden. *It could also happen that there is no useful information that we could gain from the estimated and distorted Hurst parameter.* This result raises the question how to avoid these distorting effects, e.g., how to check level shifts present in the data, or how to identify misleading cases like an inhomogeneous Poisson process when trying to capture long-range dependence. To answer these questions it is not a trivial task and is beyond the scope of this dissertation left for further studies. (Some ideas how to check level shifts in the data was published in [43].) In certain cases more characteristics are needed to describe the complex traffic and the appealing fractal characterization with only a few parameters will not be appropriate. In this case we can use more scaling parameters on different time scales and the framework of multifractal characterization should be introduced (see, e.g., [34]).

The statements above highlight the difference between theory and practice. In theory, it is possible to speak about self-similar processes with infinite structure, and long-range dependence which refers to the correlation structure at infinite lags. In practice, we only have finite measures and sample traces that show certain features which are closely related to the mathematical concepts above. To avoid any confusion caused by the improper use of the concepts of self-similarity, I will call a sample trace which shows self-similar features "scaling data" or "traffic with scaling behavior". From an engineering point of view, I call a (necessarily finite) sample path long range dependent, if there are significant correlations present for (arbitrarily) large time scales. Similarly, the "estimated Hurst parameter" characterizes the scaling behavior of the data at the examined time scales.

3.5 Impact of LRD on cell loss

3.5.1 ATM measurements

LAN interconnection is one of the most popular services provided by Telia, the Swedish network operator, on its ATM wide area network. Apart from business customers, different parts of the Swedish University Network (SUNET) are also attached to the Swedish ATM WAN. The aggregated traffic on the SUNET were analyzed during summer 1996 in the

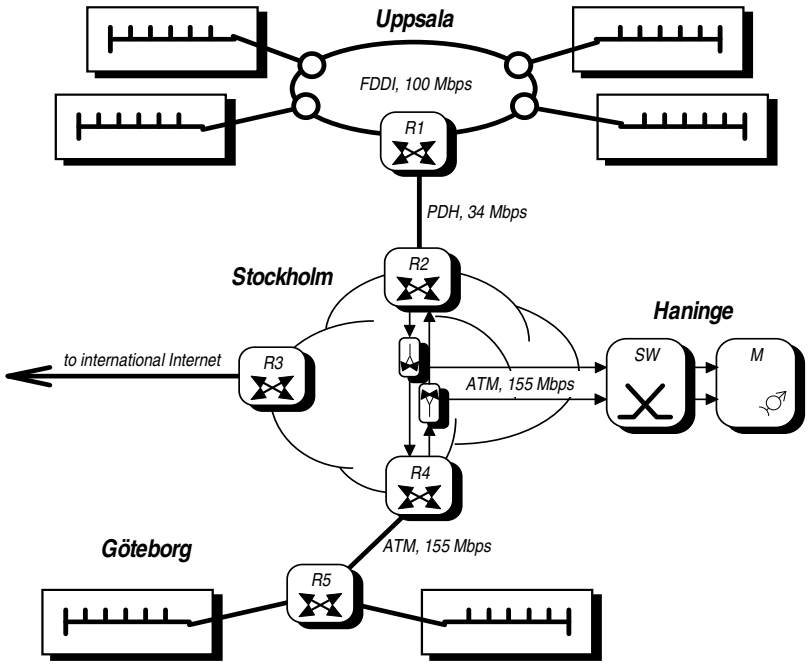

Figure 3.16: The configuration of measurements on the SUNET.

framework of a common trial between the SUNET community and Telia Research. The LAN traffic of universities in the northern region, around Uppsala are connected to an FDDI backbone which is connected via R1, R2 routers and a 34 Mbps PDH link to the ATM backbone in Stockholm (Figure 3.16). This network joins the northern LANs of SUNET to the international Internet backbone and to the southern university networks around Göteborg. A CBR connection with 90 000 cps (38.16 Mbps) cell rate was established on the SDH link between the routers R4 and R5 for the trial. The measurements reported here were performed on the connections between Uppsala and Göteborg. ATM traffic streams were duplicated by means of optical splitters avoiding impacts on original traffic flows. The duplicated traffic streams were routed on dedicated links to Telia Research in Haninge, where almost one hundred traffic traces were collected with more than 8 million cell arrivals

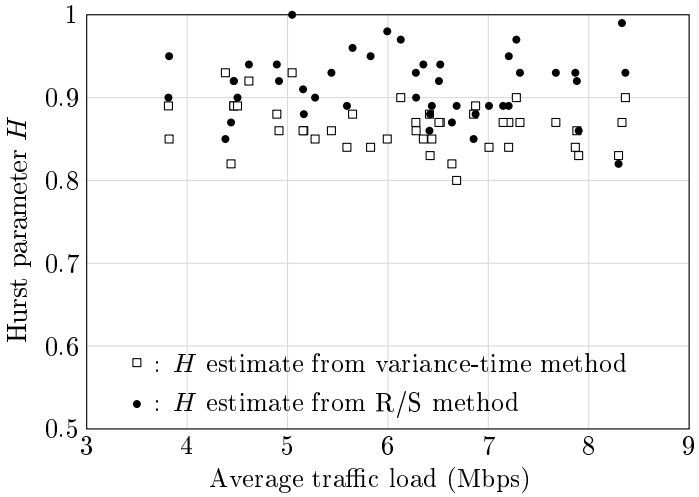

Figure 3.17: Estimated values of H as a function of traffic load.

in each trace using a non-commercial custom built measurement instrument developed in the RACE Parasol project [36].

These connections used Telia's Guaranteed Traffic Class thus the influence from other traffic in the network was negligible. A good assumption is that the traffic was an ordinary mix of common Internet traffic types such as HTTP, FTP, telnet, chat, IPphone etc.

3.5.2 Scaling analysis

To estimate the Hurst-parameter H, the R/S and variance-time analysis (see Section 3.1.2 and also in [4]) were performed for 45 data sets.

The obtained values of $\hat{H}$ are plotted on Figure 3.17. To get more information about the estimated Hurst-parameter of the measured traffic, $\hat{H}$ is plotted against the average load of the traces. As can be seen on the figure, $\hat{H}$ varies within the range $(0.8, 1)$ and does not depend significantly on the load. Figure 3.18 reveals an interesting phenomenon, namely, that there is a knee-point in the R/S diagram that separates two linear regions of sample points. Since LRD is an asymptotic property, the linear region to the right (sample points marked with an 'x') was used to determine the estimate of H. The same knee-point was found in all the data sets. Since the analyzed process was the number of cell arrivals in consecutive 0.1 msec time intervals, the knee-point is approximately at 0.3-0.4 seconds. The origin of this behaviour could be revealed by examining the burstiness structure of the data more deeply. (The authors of [37] showed that somewhat similar behavior can be experienced when there is a level shift in the data.)

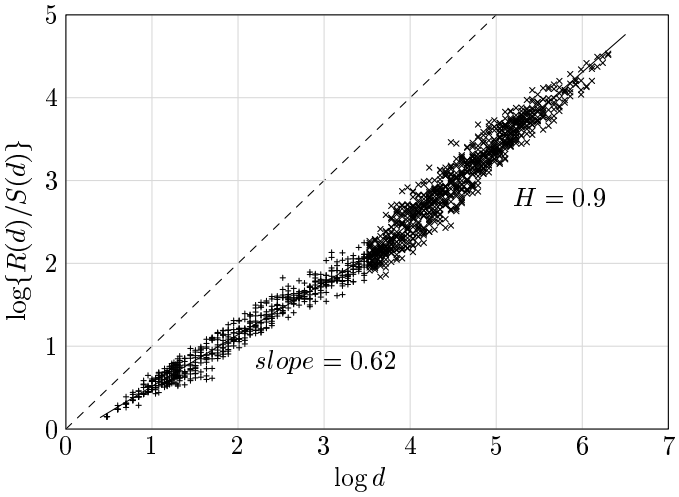

Figure 3.18: R/S values plotted against the logarithm of block size.

3.5.3 Relevance of time scales in queueing

An important question is what is the impact of LRD on queueing. Several engineering issues, such as buffer dimensioning and traffic control, are related to this question which makes it extremely important. There are two opposing viewpoints on this problem. One claim is that the queueing performance is determined by the time scale of busy periods of the queues and there is no practical impact of correlations above this time scale [26, 27, 28, 47]. The contradicting claim is based on several studies [22, 31] and states that LRD is one of the main characteristics of the traffic with significant effect on queueing behaviour. In this section we present some experimental results in order to clarify this question.

Queueing setup

In the performance analysis the queueing set-up shown in Figure 3.19 was used. The pre-recorded traces were taken from our SUNET database. The correlation structure of the original data was artificially modified by a 'shuffler' (see next section). The functionality of this shuffler was investigated by computer simulation. The performance measures under investigation were the complementary queue length distribution and the cell loss ratio (CLR).

External shuffling

In the analysis, external shuffling was used as a tool to modify the correlation structure of the measured traffic traces. Using the terminology of [2, 22], I call 'external shuffling' the

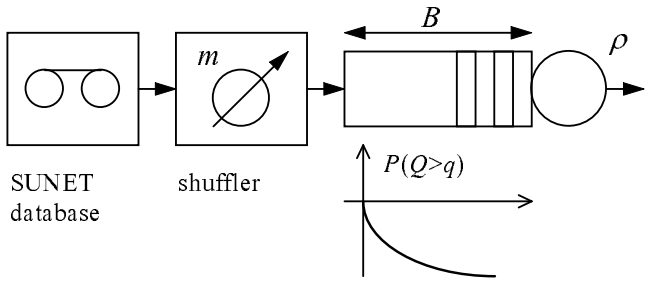

Figure 3.19: Queueing set-up.

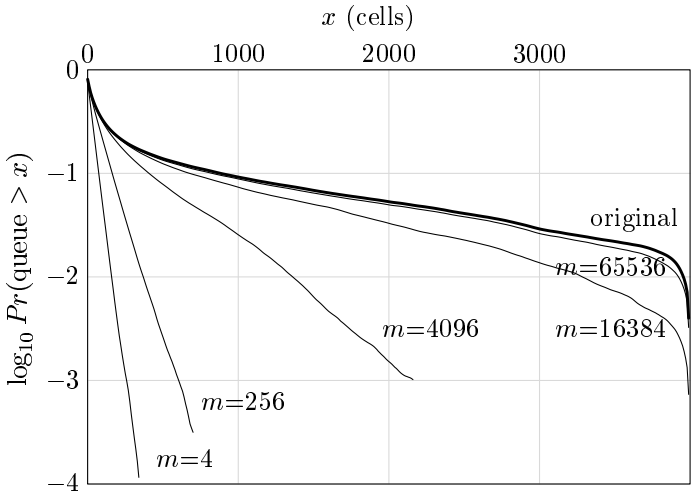

Figure 3.20: Complementary queue length distributions.

following method. First, divide the sequence of interarrival times[2] into blocks of size m. For a measured traffic trace containing N cell arrivals, there are N/m such blocks. Then the order of the blocks is shuffled, while preserving the cell sequence inside each block. Thus, for different values of m the short-range correlations (up to lag m) are preserved while eliminating the long-range correlations (beyond lag m).

Therefore by plotting the complementary queue length distributions for the original and the shuffled traces with different block sizes we have a simple tool to investigate the effect of short-term and long-term correlations in queueing (see Figure 3.20).

[2]Beside the permutation of a sequence of interarrival times, one could perform the same shuffling on the sequence of the number of arrivals in consecutive time slots [2].

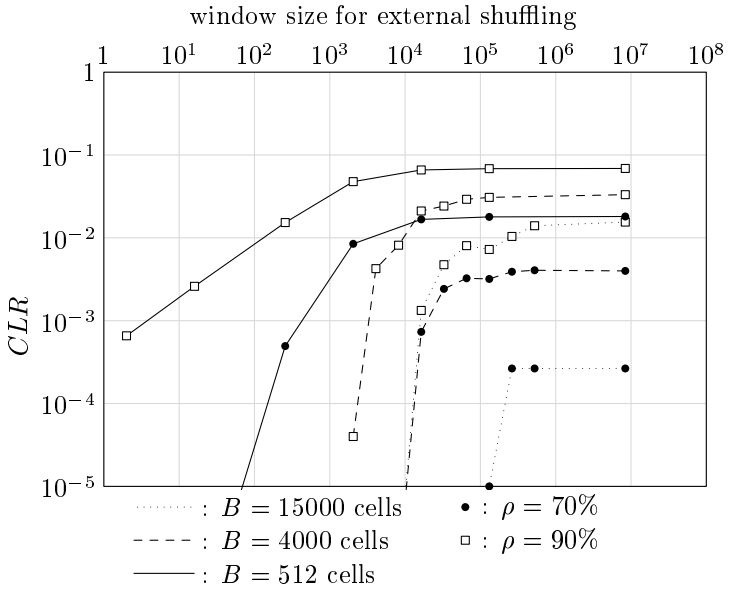

Figure 3.21: Cell loss ratio with external shuffling.

3.5.4 Queueing properties of LRD input

In case of LRD traffic input, the tail of the complementary queue length distribution decays slower than exponentially. The uppermost solid line related to the original traffic trace in Figure 3.20 shows this phenomenon. (The buffer size was set to 4000 cells and the service rate was chosen so that the utilization ρ was 0.7. The shuffling block size m is shown on the figure for each simulated curve.) By shuffling the input traffic, it can be observed that if the block size is increased the curve approaches the distribution curve of the original trace. It means that we can find the time scale of the relevant correlations, i.e., the block size of shuffling where the distribution is approximately the same as for the original case, which are important for queueing. It can also be concluded that beyond that time scale there is no significant effect of correlations. In our case this time scale is in the range of 4–5 seconds (block size of 10^5) for load 0.7. This finding is in accordance with the results reported in [26].

3.5.5 Impacts of LRD on cell loss

The queue length distribution in the infinite buffer case gives an upper bound on the cell loss ratio (CLR) for the finite buffer case. In practice, the latter is of interest. To investigate the effects of scaling behavior on CLR, simulation studies were performed with the measured traces as input for three different buffer sizes (512, 4000 and 15000 cells). The rate of service

was set to obtain utilizations 0.7 and 0.9. The results are plotted on Figure 3.21.

It can be observed in Figure 3.21 that *there is an upper time scale determined by the buffer size and the load where there is no effect of correlations on cell loss if we go beyond that time scale.* For example, in the experiments above cell lag 10^5 (approximately 5 seconds) the cell loss curves are practically constant even for the large (15000 cells) buffer case. It supports the finding derived from Figure 3.20, namely, that there is a time scale which determines the biggest lag of correlations which has effect on the cell loss. However, this upper time scale seems to be dependent on many parameters. Firstly, it depends on the buffer size, i.e., the bigger the buffer the bigger the upper time scale. For moderate and large buffers there is a sharp cut-off in cell loss as a function of the shuffling block size but for small buffers the appearance of this upper time scale is not very pronounced. Secondly, this upper time scale is also dependent on the load, the curves are shifting left as the load increases although the shifting of cut-off lag is not significant. In the experiments it can be observed that the cut-off lag is always above the buffer size but typically not more than one decade above the buffer size. These findings give us simple practical engineering rule of thumb for estimating the range of relevant correlation time scale.

3.6 Conclusion

In this chapter, practical questions and problems were investigated regarding the detection of self-similarity and the estimation of its scaling parameter H. The goal was to investigate the effects of traffic control mechanisms on the Hurst parameter estimates. Furthermore, the effects of scaling property present in the input traffic on CLR in queueing was also examined. The results were obtained by empirical studies using measured data from ATM wide area networks.

In practice, using measured data sets the estimated values of H obtained from different analysis methods are influenced by several factors. Four different H estimation method were investigated: the variance-time analysis, the Index of Dispersion for Counts (IDC) analysis, the R/S analysis and the periodogram-based method. According to the results the analyzed data sets showed self-similar behavior with Hurst parameter $H \approx 0.7$.

It can be also concluded that the traffic can have different scaling parameters at different time scales. Therefore the relevant time scale has to be specified prior to H estimation. The choice of the relevant time scale should be based on traffic engineering considerations.

It was also shown that the presence of level shifts in the data can have disastrous effect on the Hurst parameter estimates. The estimate is seriously distorted, thus level shift caused by nonstationarity must be removed prior to H estimation. This result emphasizes that it could also happen that there is no useful information we could gain from the estimated and distorted Hurst parameter.

The effects of the leaky-bucket shaping and policing mechanisms on the estimated Hurst parameter were also investigated. The results show that the estimated parameter $\hat{H}$ is increased due to shaping, and was practically unaffected by the policing mechanism. Thus, it can be concluded that the self-similar feature is a robust property. (Note, that it is not

obvious how to interpret the parameter H in practice, where the traffic structure is modified by several mechanisms and the process is not a pure self-similar process.)

To investigate the effects of scaling traffic on the cell loss ratio in queueing, simulation studies with measured traces as input were investigated. The obtained results show, that there is an upper time scale determined by the buffer size and the load where there is no effect of correlations on CLR if we go beyond that time scale. This time scale also depends on the buffer size (i.e., the bigger the buffer the bigger the time scale of interest) and on the load (i.e., the higher the load the smaller the time scale). In the experiments one can observe that this cut-off lag is always above the buffer size but typically not more than one decade above the buffer size. The findings above give practical engineering guidelines for estimating the range of relevant correlation time scales.

Chapter 4

Effective Bandwidth Formula Based on Queue Length Monitoring

If the actual cell loss performance of an ATM output buffer could be determined in real time, the rate of the server (that is, the VP bandwidth) could be adjusted such that the cell loss would be smaller than a pre-determined threshold (CLR_{obj}). Shioda and Saito presented a method for estimating the cell loss ratio in real time, and applied it to VP bandwidth estimation and call admission control [49]. They utilized the large deviation result of Glynn and Whitt [25], when the input is short-range dependent and the cell loss ratio (CLR) decays exponentially as the buffer size increases.

For long-range dependent traffic the asymptotics of the queue length distributions are no longer exponential. Recent results from the literature show that for self-similar traffic input the tail of a stationary queue length distribution is Weibullian [6].

The aim is to estimate the effective bandwidth of the traffic even if it shows long-range dependent properties. To do this, a simplified effective bandwidth formula can be derived that only requires the knowledge of the actual CLR and the utilization. I have proposed an algorithm to estimate the CLR in real time based on buffer measurements, which works for both the long-range and the short-range dependent case. I have also given an improved approximate effective bandwidth formula which takes into account long-range dependence.

4.1 Approximate effective bandwidth equation

Consider a single server queue that has an infinite buffer with stationary and ergodic cell arrivals from a single source. Let $A(t)$ denote the amount of traffic arrived in $(-t, 0]$, and the workload process is defined by

$$W(t) = A(t) - Ct, \tag{4.1}$$

where C is the amount of traffic that can be served in unit time. The virtual waiting time in the queue in the stationary state is given by [50]

$$Q = \sup_{t \geq 0} W(t). \tag{4.2}$$

Results from queueing analysis based on the large deviation principle reveal that for a wide range of traffic input (that does not have LRD!), the asymptotic tail distribution of the queue length decays exponentially [17]. The probability that the queue contents exceed a given (large) threshold k is thus characterized by two parameters, the asymptotic constant β and the asymptotic decay rate η, such that

$$p_k \stackrel{\text{def}}{=} P(Q \geq k) \approx \beta e^{-\eta k}. \tag{4.3}$$

It is known that the CLR of a single server queue with a finite buffer size K is less than p_K [17]. Hence, we have

$$CLR \leq \beta e^{-\eta K}. \tag{4.4}$$

From Eq.(4.4) we get that the QoS objective of the CLR, CLR_{obj}, is satisfied if

$$\eta \geq \eta_{obj} \stackrel{\text{def}}{=} \frac{\log \beta - \log CLR_{obj}}{K}. \tag{4.5}$$

Results from large deviation theory [25] yield the following necessary and sufficient condition for meeting the QoS objective of the CLR:

$$\alpha(\eta_{obj}) \stackrel{\text{def}}{=} \frac{M_A(\eta_{obj})}{\eta_{obj}} \leq C, \tag{4.6}$$

where $M_A(\cdot)$ is the cumulant generating function of $A(t)$. The function $\alpha(\eta)$ is called the effective bandwidth function of the source subjective to the condition that the tail distribution of the queue length has the decay rate η. In particular, $\alpha(\eta_{obj})$ is simply called the *effective bandwidth*, given the decay rate objective η_{obj}.

Shioda and Saito gave an approximation to calculate Eq.(4.6) using cumulant expansion and heavy traffic expansion techniques [49] (see Appendix C). Here I propose a not so rigorous but a more intuitive and lighweight method to get an approximate formula for the effective bandwidth.

A simple approximation formula has been given for the *M/D/1* system [45] that is accurate when the load is high (heavy-traffic assumption). This formula suggests that the asymptotic decay rate in Eq.(4.3) can be approximated as

$$\eta \approx 2\frac{1-\rho}{\rho} \tag{4.7}$$

with $\rho = A/C$ where A denotes the mean arrival rate. (Note that the results in [45] show that this approximation is very accurate even for small values of k.) Assuming further that $\beta \approx 1$ (which is not true in certain cases), for a given K from Eq.(4.3) and Eq.(4.7) we get

$$\log p_K \approx -2K\frac{1-\rho}{\rho}. \tag{4.8}$$

Although Eq.(4.8) was derived under the assumption that the input process is Poissonian, I expect that a similar result of the form

$$\log p_K \approx \frac{1}{d}\frac{1-\rho}{\rho} \tag{4.9}$$

holds approximately for a wider class of input processes with some constant d, because many similar formulae have been developed. Shioda and Saito [49] showed that Eq.(4.9) can be derived with

$$d = -\frac{1}{2K}\lim_{t\to\infty}\frac{Var\{A(t)\}}{E\{A(t)\}}. \tag{4.10}$$

Norros [41] studied a storage model with fractional Brownian input. In the limiting case when $H = 0.5$, the result reduces to Eq.(4.8) (or equivalently, $d = -1/(2K)$ in Eq.(4.9)).

Rearranging Eq.(4.9) we get

$$C = A(1 + d\log p_K). \tag{4.11}$$

Since the CLR of a single-server queue with a finite buffer of size K is less than p_K [17], from Eq.(4.11) the effective bandwidth of the source can be approximated by

$$\alpha(\eta_{obj}) \approx A(1 + d\log CLR_{obj}). \tag{4.12}$$

(Eq.(4.12) is the same as Eq.(C.12) on page 90.) Note the main difference between CLR_{obj} and p_K. While CLR_{obj} is the *desired* (or objective) cell loss ratio, p_K can be considered as the *actual* experienced loss rate.

By combining Eq.(4.11) and Eq.(4.12) we finally have

$$\alpha(\eta_{obj}) = A\left(1 - \frac{\log CLR_{obj}}{\log p_K}\right) + C\frac{\log CLR_{obj}}{\log p_K}. \tag{4.13}$$

To calculate the effective bandwidth, according to Eq.(4.13) we only need to measure the average arrival rate A and the probability p_K. (Note, that the finer details of the complementary queue length distribution are not important for us. We only need to know one specific point of it.)

4.2 Three-point buffer measurement method

When the traffic does not possess long-range dependence, the queue length distribution is asymptotically exponential (see Eq.(4.3)). Shioda and Saito showed [49] that for this case the parameters β and η can be estimated by monitoring the buffer occupancy levels at two different thresholds.

For long-range dependent traffic input the queue length asymptotics are rather characterized by a Weibullian distribution [6]:

$$p_k = \beta e^{-\eta k^\gamma}, \quad \gamma \in (0, 1]. \tag{4.14}$$

It should be noted, that not in all the cases of LRD input is the queue length distribution Weibullian. For example, certain on-off type models can generate algebraic decays [52]. However, compared to the exponential approximation, having the additional parameter γ the Weibullian distribution ensures a better fit than the exponential one. Furthermore, for some types of LRD traffic models, theoretical investigations show that the tail in the asymptotic region is Weibullian [6, 18, 41].

Because $\log p_K$ is a function of K^γ instead of K, three thresholds are needed instead of two to monitor the queue length. Suppose that at thresholds k_1, k_2, and k_3 the buffer occupancy levels p_{k_1}, p_{k_2} and p_{k_3} are observed. To make the calculations easier, assume that the ratios of the thresholds satisfy $k_2/k_1 = k_3/k_2 = c$, with $c > 1$ (for example, for $c = 2$, one possible setting is $k_1 = 5$, $k_2 = 10$, and $k_3 = 20$). The asymptotics of the queue length distribution (Eq.(4.14)) yield

$$\log p_K = \frac{\log p_{k_2} - \log p_{k_1}}{k_2^\gamma - k_1^\gamma} K^\gamma + \frac{k_2^\gamma \log p_{k_1} - k_1^\gamma \log p_{k_2}}{k_2^\gamma - k_1^\gamma}, \tag{4.15}$$

where

$$\gamma = \frac{\log(\log p_{k_2} - \log p_{k_3}) - \log(\log p_{k_1} - \log p_{k_2})}{\log c}. \tag{4.16}$$

Figure 4.1 helps to explain the seemingly complicated formulae Eq.(4.15) and Eq.(4.16). Having the unknown parameters β, η and γ in Eq.(4.14) we need to solve the three equations

$$\log p_{k_i} = \log \beta - \eta k_i^\gamma, \quad i = 1, 2, 3. \tag{4.17}$$

Once we know the three parameters, p_K can be calculated using Eq.(4.14).

4.2.1 Buffer thresholds, length of monitoring interval

The actual buffer-monitoring thresholds depend on the buffer length, the CLR objective, and the length of the monitoring interval, as well as on the rate and burstiness of the traffic. Here I propose some rules-of-thumb and simple formulas for calculating them.

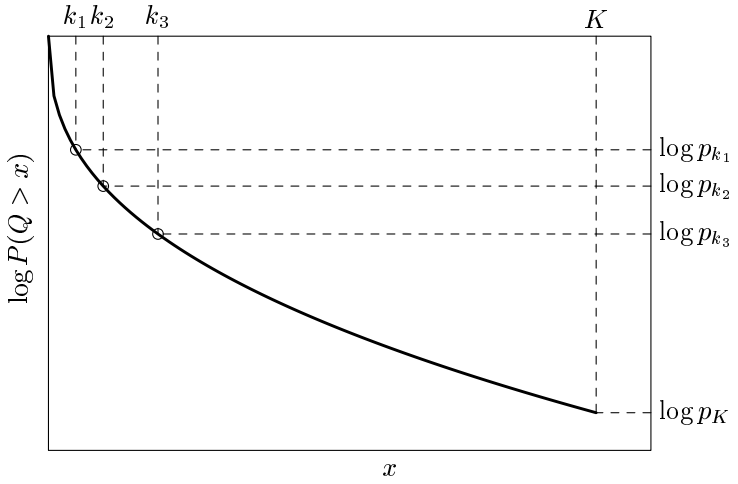

Figure 4.1: Three-point buffer monitoring method

The number of cell arrivals per monitoring interval (τA) can be calculated by fixing the length of the monitoring interval (τ) and taking into account the average arrival rate (A) of the incoming traffic (which can be determined based on measurement). Because we want to measure the probability p_{k_3} of crossing the threshold k_3, to obtain a statistically reliable estimate, we need an adequate number of crossing events as the queue fluctuates in the buffer. The smallest probability we can monitor is $(\tau A)^{-1}$ which corresponds to the case when only one cell is counted. Smaller probabilities cannot be monitored at all (see the dark-shaded area in Figure 4.2). Since cell arrivals are bursty in nature, we need at least in the order of 10^2 crossing events to get reliable probability estimates (see light-shaded area in Figure 4.2). Therefore the largest monitoring threshold k_3 has to be set so that the probability estimate p_{k_3} satisfy:

$$\log p_{k_3} \geq 2 - \log(\tau A). \tag{4.18}$$

Assuming that the VP bandwidth is set so that the CLR objective is satisfied ($p_K \approx CLR_{obj}$), an upper bound $\overline{k_3}$ for the largest threshold k_3 can be calculated using Eq.(4.14) with $\beta \approx 1$, and Eq.(4.18):

$$k_3 \leq \overline{k}_3 = K \left(\frac{\log CLR_{obj}}{2 - \log(\tau A)} \right)^{1/\gamma}. \tag{4.19}$$

To proceed further γ has to be specified somehow in Eq.(4.19). It can be assumed that the traffic does not possess long-range dependence, and thus $\gamma = 1$. However, this assumption would overestimate the upper bound for setting the monitoring thresholds, and this is because of the convexity of the Weibullian curve on the logarithmic plot. A more

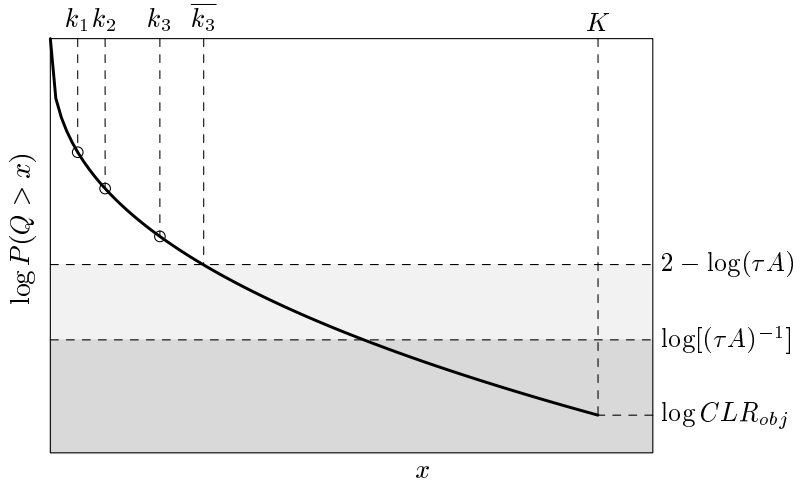

Figure 4.2: Bounds for setting the monitoring thresholds k_1, k_2 and k_3.

realistic approach is when we try to calculate γ assuming that the traffic is well modeled with a self-similar process with index H. Assume that H is known, or can be estimated based on previous measurements. (Note, that for our purpose a rough estimate is adequate.) Results from the literature show that for self-similar traffic input Eq.(4.14) holds with $\gamma = 2 - 2H$. Based on this result the upper bound $\overline{k_3}$ can be calculated now.

The last step is to determine the actual threshold values. Since they must be integers, calculate k_1 first as

$$k_1 = \left\lfloor \frac{\overline{k}_3}{c^2} \right\rfloor, \tag{4.20}$$

where $\lfloor \cdot \rfloor$ denotes the integer part. Finally, set

$$k_2 = c \cdot k_1, \quad \text{and} \quad k_3 = c \cdot k_2. \tag{4.21}$$

In setting the lower thresholds, k_1 and k_2, we should note that our queue length asymptotics are valid only for the tails of the distribution. However, as stated earlier, the thresholds cannot be set arbitrarily far from the origin. This sometimes results in the fact that the thresholds are set to the "pre-asymptotic" region. In this case the Weibullian fit is, so to say, used for the entire queue length distribution. Fortunately, this is usually not a problem but an approximation that works reasonably well (see, for example, Figure 4.6 later). Nonetheless, the three thresholds should cover a span wide enough to yield a reliable estimate for $\log p_K$. The value of this tradeoff can be evaluated only for the individual application.

4.2.2 Validation with simulations

Simulation experiments are described in this section to study the Weibullian approximation for the queue length distribution compared to the exponential one. Three cases are examined. First, the aggregated traffic is short-range dependent with low burstiness. In this case, the queue length distribution can be well approximated by an exponential distribution with asymptotic constant β close to one [9]. Second, the offered traffic is short-range dependent, but more bursty, implying $\beta << 1$. Finally, the traffic is long-range dependent.

Short-range dependent traffic

Consider N identical but independent on-off sources with alternating activity periods (T_{on}) and silence periods (T_{off}). In the on-states, cells were offered at a constant rate, while in the off-states the source remained silent. The duration of each on-state and each off-state were exponentially distributed. (This model is referred as the 'on-off source'.) The cells from the multiplexed on-off sources arrived at the monitored buffer. This buffer could hold 128 cells.

For the short-range dependent input, ten sources were multiplexed, and cells were offered from each source at a rate of 2 Mbps during exponentially distributed on periods with mean 0.543 ms in the 'non-bursty' case, and 5.43 ms in the 'bursty' case. The average rate of each source was set to 0.2 Mbps. At three different thresholds (5, 10, and 20), the frequencies in which the number of cells in the buffer were greater than or equal to the thresholds were measured every 20 minutes. The service was set so that the estimated CLR was approximately 10^{-4} and it was 2.35 Mbps for the 'non-bursty' case and 6.05 Mbps for the 'bursty' case.

The simulated CLR is plotted together with the estimated values using Weibullian and exponential approximations for the queue length distributions in Figure 4.3 and Figure 4.5. (The 'theoretical CLR' is the probability that the queue length exceeds K in the case of an infinite buffer, as calculated from the queue length asymptotics. Actually, this should be an upper bound on the actual CLR for the finite-buffer case.)

In the case of 'non-bursty sources', the parameter γ in the Weibullian distribution (Eq.(4.14)) becomes $\simeq 1$, and the fitted distribution is close to the real one (Figure 4.4). The proposed method works well, and gives only slightly higher CLR estimates as the exponential one.

In the case of 'bursty sources' (Figure 4.6)—although the tail asymptotics for the distribution *are* exponential,— monitoring the queue at two points and estimating the queue length distribution by an exponential distribution is not appropriate here, the CLR is underestimated. The three-point measurement method works better in this case (Figure 4.5).

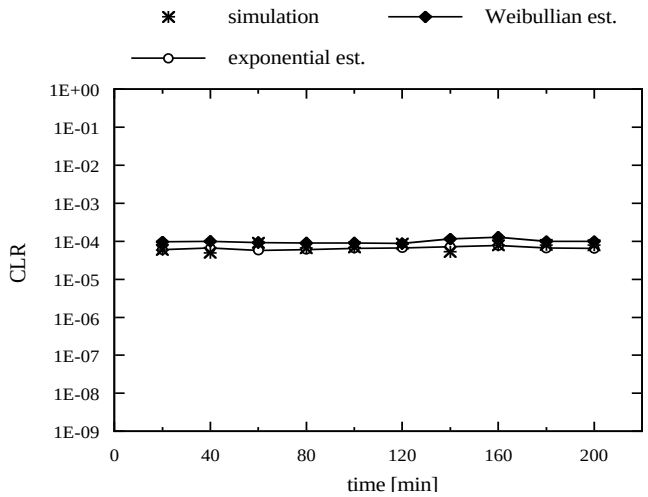

Figure 4.3: CLR estimates using the proposed measurement method (non-bursty case).

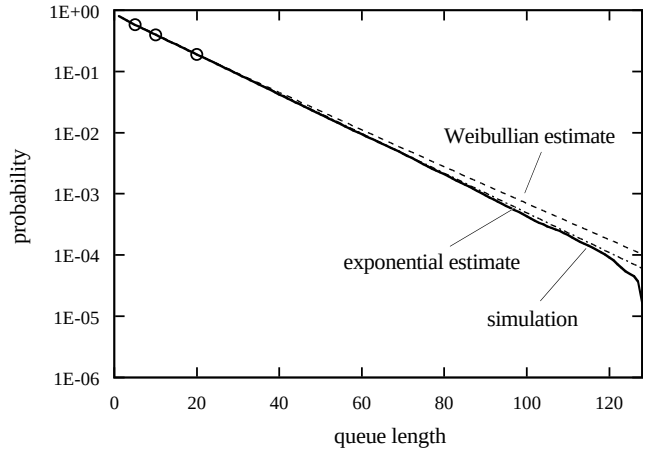

Figure 4.4: Queue length distribution for non-bursty on-off sources.

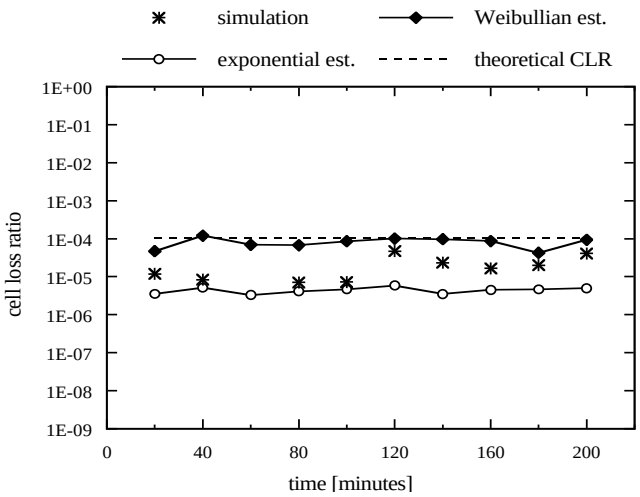

Figure 4.5: CLR estimates using the proposed measurement method (on-off sources).

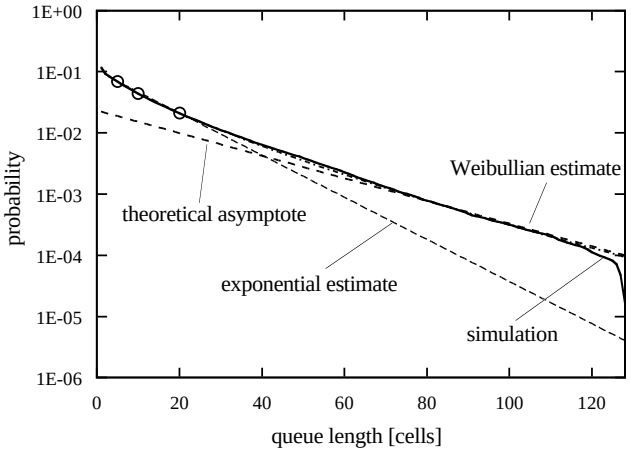

Figure 4.6: Queue length distribution for on-off sources.

Long-range dependent traffic

Usual (finite state space) Markovian source models are capable of generating only short-range dependent traffic. Therefore, to generate long-range dependent traffic, we need to find an appropriate method. Paxson discussed six different such methods [44]. For the simulator, the on-off model originally proposed by Leland *et al.* [31] is suitable.

To introduce LRD into the on-off model, assume a heavy tail for the distribution of the activity periods T_{on} while the off periods remain exponentially distributed [6]. The (transformed) Pareto distribution is used for T_{on}, where

$$P(T_{\text{on}} \leq t) = 1 - \left(\frac{\theta}{t+\theta}\right)^{3-2H} \tag{4.22}$$

for some $\theta > 0$, $t > 0$, and $0.5 < H < 1$; therefore, periods of heavy-tailed activity generate LRD with Hurst parameter H. (This source model is referred as a 'Pareto-type source'.) It has been shown that, under the constraints of a large number of such sources and a high load, the aggregated traffic, properly normalized, converges to an exactly self-similar Gaussian process [31, 7].

In the simulation the number of sources was set to 100. The mean duration of the on period was 0.1 ms, and the parameters of the Pareto distribution were set using $H = 0.7$. The service rate was 26 Mbps, and all the other parameters were retained unchanged. As shown in Figure 4.7, the difference between the exponential and Weibullian estimates is about four orders of magnitude. The measured CLR values are bounded by the two estimates, and are of the order of 10^{-5}. The reason that the exponential fitting underestimates the CLR is shown in Figure 4.8. *Any* exponential estimate would underestimate the CLR. In comparison, the Weibullian approximation fits the queue length distribution relatively well, and the estimate of p_K again turns out to be an upper bound on the actual CLR.

4.3 Improved effective bandwidth formula

The greatest advantage of the effective bandwidth formula of Eq.(4.12) comes from its simplicity. It is rather simple in the sense that only one parameter, d, describes the relevant characteristics of the traffic. However, in certain cases this simplified equation is not fully adequate.

4.3.1 Scaling properties of the effective bandwidth

The effective bandwidth satisfies the *additive property*, that is, the effective bandwidth of the aggregation of independent sources is equal to the sum of the effective bandwidths of aggregated sources.

On the other hand, when independent and bursty sources are multiplexed on a single link, the independence in the statistical variations of the individual sources offers the potential for a reduction in the bandwidth required for the aggregated stream. There is no

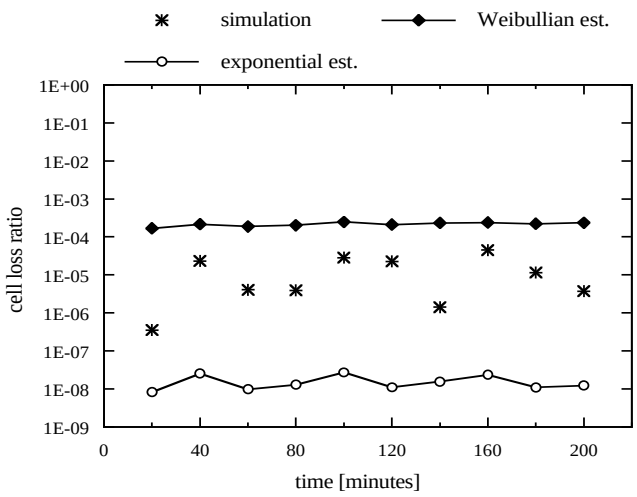

Figure 4.7: CLR estimates using the proposed method for Pareto-type sources.

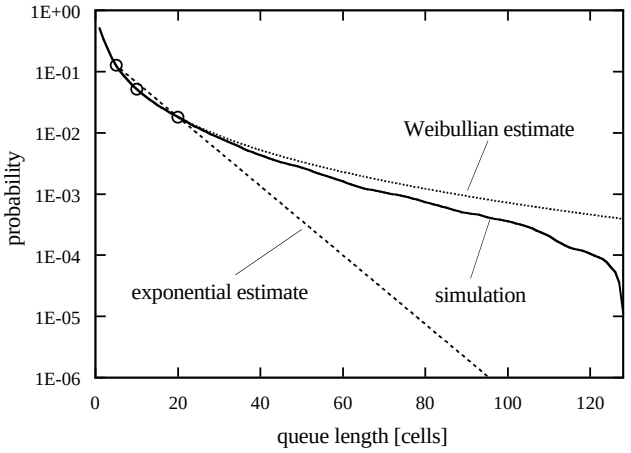

Figure 4.8: Queue length distribution for Pareto-type sources.

need to increase the bandwidth proportionally as the number of sources increase if we want to keep the desired cell loss ratio. This achievable multiplexing gain is missing from the effective bandwidth approach.

Next, I try to describe this effect quantitatively. This section addresses the effects of the Hurst parameter H, and asymptotic constant β to the effective bandwidth formula. For the parameter H, an idea is presented for estimating and incorporating it into the algorithm.

4.3.2 The asymptotic constant β

As for the asymptotic constant β (see Eq.(4.3), when deriving the very simple and effective bandwidth formula we assumed that $\beta \approx 1$. This simple approximation is convenient because β is hard to obtain; it is also consistent with the notion of effective bandwidth. However, in certain cases its value can be very different from one when the number of component arrival processes is large [9]. (Note, that Elvalid *et al.* made an intuitive and quite clear argument about the asymptotic constant [21]: they gave an approximate formula for multiplexed on-off sources with their peak and average rate given.)

The asymptotic constant β is in fact incorporated in the proposed method. The three-point buffer monitoring algorithm was constructed from Eq.(4.14) which contains β.

4.3.3 The Hurst parameter H

Burstiness across different time scales is described by the Hurst parameter, hence the multiplexing gain is also affected by H [29]. Next I describe how the parameter H could be used explicitly in the algorithm. (Note that, similar to β, H is also already implicitly considered because the assumption of the Weibullian queue length distribution corresponds to the LRD phenomenon.) For this, results for the well-known fractional Brownian traffic model [41, 42] are used with the following notations and assumptions:

- The fractional Brownian traffic model with parameters $\{m, a', H\}$ is a good approximation for the aggregated cell stream, i.e.,

$$A(t) = mt + \sqrt{ma'}Z(t). \tag{4.23}$$

Note the different variance parameter a' in Eq.(4.23) (cf. Eq.(2.1) on page 12 where $a = ma'$). This assumption is validated by findings in [51], namely, that the superposition of many on-off sources whose on- and/or off-periods have heavy-tailed distribution can produce self-similar aggregate network traffic, i.e., converges to the fractional Brownian traffic model.

- When the number of sources is changed, a' and H of the aggregated traffic remain unchanged, while the mean rate m is scaled appropriately.

The reason behind this second assumption is the following: The parameter a' can usually be assumed to have a fixed value independently from m. Indeed, consider a superposition of N

independent and identically distributed cumulative traffic processes $A(t) = \sum_{i=1}^{N} A_i(t)$ such that $\mathrm{E}[A_i(1)] = m_0$ and $\mathrm{Var}[A_i(1)] = a' m_0$: we have $\mathrm{Var}[A(1)] = a' \mathrm{E}[A(1)]$ independent of N. Thus, H and a' characterize the type of the traffic mix while m gives its amount [41].

From the scaling property of the fractional Brownian storage system (see [42]) the following results can be derived:

$$C \approx m + \left(\kappa(H)\sqrt{-2\log p_K}\right)^{1/H} a'^{1/(2H)} K^{(H-1)/H} m^{1/(2H)} \tag{4.24}$$

with $\kappa(H) = H^H(1-H)^{1-H}$. (Note that the only approximate part of Eq.(4.24) within the framework of this model is the coefficient in front of the powers of a', K, and m.) If Eq.(4.24) is used as a substitute for the effective bandwidth formula, instead of Eqs.(4.9) and (4.12), with $A \equiv m$ we can write

$$d_H = \frac{A - C}{(-A\log p_K)^{1/(2H)}}, \tag{4.25}$$

$$\alpha(\eta_{obj}) = \hat{A} - \hat{d}_H(-\hat{A}\log CLR_{obj})^{1/(2H)}, \tag{4.26}$$

where

$$d_H = -H(1-H)^{(1-H)/H}(2a')^{1/(2H)} K^{(H-1)/H} \tag{4.27}$$

describes the connection between a and d_H. As can be seen from Eq.(4.26), the calculated effective bandwidth C is a nonlinear function of the traffic arrival rate A for $H > 0.5$. In this case, when the number of (bursty) sources is doubled ($A = 2A_0$), the bandwidth is increased only moderately ($\alpha(\eta_{obj}) = (2H+1)/(2H)C_0$). As for $H = 0.5$, it is easy to see that Eqs. (4.25) and (4.26) give the same results as Eqs.(4.9) and (4.12), with $d_H \equiv d = a'/(2K)$.

4.3.4 Hurst parameter estimation

To explicitly incorporate the Hurst parameter into our method, an on-line H estimator is needed that is unbiased under very general conditions and robust against the presence of deterministic trends, and that also can be implemented efficiently. Well-known simple estimators are not practical for our purposes because they cannot fulfill these three requirements [38].

In using a Weibullian approximation for the queue length distribution, we already assumed that the distribution has the form

$$P(Q > K) \approx \beta e^{-\eta K^{\gamma}}, \tag{4.28}$$

with $\gamma = 2 - 2H$. Based on queue length monitoring, a first approach could be to estimate H as $\hat{H} = 1 - \gamma/2$. However, this does not work in practice. Monitoring the system for a short period of time is insufficient for estimating the Hurst parameter. To estimate it, a new method based on a different approach must be applied that takes into consideration the traffic structure at longer time scales.

Recently, a wavelet-based tool for analyzing long-range dependence was introduced [1]. This tool is intended to be used in an on-line measuring environment; thus it seems to be appropriate for this purpose.

4.4 Conclusion

In this chapter, the goal was to estimate the effective bandwidth of the traffic using real-time measurements. To achieve this goal an approximate effective bandwidth formula was developed, that only requires the knowledge of the actual CLR and the utilization. An algorithm to estimate the CLR in real time using buffer measurements was also proposed which works also for the case when the input process is long-range dependent. The actual buffer monitoring thresholds depend on several factors. Guidelines and simple formulae were given for setting these thresholds, and the proposed method was validated by simulations using different types of input traffic. Finally, a modified version of the approximate effective bandwidth formula was proposed to incorporate the scaling effect captured by the Hurst parameter.

Chapter 5

Summary of the Dissertation

The mathematical background of self-similarity and long-range dependence was founded in Chapter 1. The three main goals of this dissertation were as follows. In Chapter 2 the goal was to develop an efficient parameter estimation method of the fractional Brownian traffic model. In Chapter 3 the performance implications of long-range dependence was investigated. Finally, in Chapter 4 the aim was to estimate the effective bandwidth of the traffic based on real-time measurements.

5.1 Parameter estimation of fractional Brownian traffic

I have studied the parameter estimation of a traffic model based on fractional Brownian motion in Chapter 2. Throughout my work I have applied the maximum likelihood estimation (MLE) method.

- I have given an explicit expression for the maximum likelihood (ML) estimates $\hat{m}$ and $\hat{a}$ in terms of H, together with their variance. I have also given a simplified expression for the log-likelihood function from which the estimate $\hat{H}$ is obtained as the maximizing argument. The maximization can be done in an alternative way that we do not need to calculate the determinant of the autocovariance matrix, nor the derivative of its inverse whose parametric form is unknown.

A major difficulty in the Gaussian MLE method is the calculation of the inverse and determinant of the covariance matrix appearing in the likelihood function. To further improve the estimation method (i.e., to calculate the inverse more efficiently), it is desirable to reduce the number of samples on one hand, and on the other hand, to find an efficient calculation method for the matrix inversion. The use of geometrical sampling helps in both cases.

- I have proposed an effective calculation method to provide the model parameter estimates using geometrical sampling scheme. To make it practically tractable, I have developed approximation methods to further reduce the computational power needed.

Computer simulations were used to validate the proposed estimation methods and approximations, as well as to demonstrate the advantage of geometrical sampling over the traditional linear sampling scheme.

The geometrical sampling scheme with the matrix-calculation formulae helps in two ways. First, the calculation complexity is decreased to $\mathcal{O}(n^2)$ and thus the method can be used up to a few hundred sample points. Second, the samples are distributed in a better way on different time scales, thus the estimation of H becomes more reliable. According to my experience, using geometrical sampling the estimate $\hat{H}$ is unbiased for sample sizes larger than 25 and its variance is also reduced to about one third, when compared to the linear sampling scheme.

For H the proposed estimation method was also compared to the wavelet-based method of Abry and Veitch which is known to be a fast and efficient estimator, and turned out to be very efficient. 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.

The results showed that the geometrical sampling scheme gives a great advantage. (Though the proposed approximations do not work in favor of a, the joint estimation of H and a performs much better.) The proposed approximations work well if applied appropriately and considerably reduce the computation power needed, making it possible to apply the ML method for larger sample sizes.

The work presented in Chapter 2 was made within the framework of the COM^2 project funded by the Academy of Finland. It is part of the Research Programme for Telecommunication Electronics initiated by the Academy of Finland in 1997 to further advance scientific research in the field of tele- and data communications. A program library of Matlab routines has resulted from the work and is now available freely on the Internet [T6]. The results were published in [C1, C2, T1] and [T6], and are also contributions to the European COST 257 project.

5.2 Characterization and control of scaling traffic

In Chapter 3, I have given engineering guidelines to be considered when using the self-similar modeling concept in case of real traffic. I have investigated the impacts of traffic control mechanisms on the Hurst parameter estimates, and concluded that the scaling property is quite robust. I also investigated the effects of scaling property on CLR in queueing.

To do this, empirical studies using data taken from ATM wide area networks were performed.

5.2.1 Hurst parameter estimation of real traffic

The estimation of the Hurst parameter is not easy in practice.

- I have shown that real traffic can have different scaling parameters at different time scales. Therefore the relevant time scale has to be specified prior to H estimation.

- I have investigated the effects of level shifts present in the data and showed that the presence of such level shift can have a disastrous effect on $\hat{H}$, i.e., the estimate is seriously distorted. Therefore level shifts caused by nonstationarity must be removed prior to H estimation.

This result emphasizes that it could also happen that there is no useful information we could gain from the estimated and distorted Hurst parameter estimates. This happens when the assumption of stationarity is violated, and the real problem is that it is not possible to discriminate a stationary long-range dependent sequence from a nonstationary one in the case of finite data sets.

5.2.2 Impacts of network mechanisms on H

I have investigated the impacts of traffic shaping and policing on the structure of the traffic and the Hurst parameter estimates.

The applied shaping algorithm was the leaky bucket shaping which forces nonconforming cells to be delayed.

- I have found that the estimated Hurst parameter was increased due to shaping, and was practically unaffected by the policing mechanism.

It can be concluded that the self-similar feature is more robust for policing than for shaping. If the process is a pure self-similar process there is a good interpretation of the Hurst parameter. However, it is not obvious how to interpret it in practice, where the traffic structure is modified by several mechanisms and the process is not a pure self-similar process.

5.2.3 Impact of scaling property on cell loss

There are different concerns about whether LRD is an important traffic descriptor for cell loss estimation or not. I have investigated this question and tried to identify the relevant correlation time scales of actual measured traffic. Queueing and shuffling analysis were performed in order to investigate the effect of scaling on cell loss. Simulation studies were performed with measured traces as input for different buffer sizes and different utilizations. I have shown, that:

- There is an upper time scale (also called as critical time scale) determined by the buffer size and the load where there is no effect of correlations on cell loss if we go beyond that time scale. The critical time scale depends on the buffer size, i.e., the bigger the buffer the bigger the upper time scale. The critical time scale also depends on the load, i.e., the higher the load the smaller the time scale of interest.

Strictly speaking, the existence of such cut-off lag excludes the mathematical notion of LRD (which is an asymptotic definition and only tells us something about the behavior of the correlations as the lag tends to infinity). In practice, however, the real question remains: "How long is long-range dependence?" The findings above give us simple practical engineering rules of thumb for estimating the range of relevant correlation time scale from cell loss point of view.

The results of Chapter 3 provide guidelines which are to be considered in many aspects of traffic characterization and network management when dealing with traffic with long-range dependence and self-similar features. The queueing analysis results can form the basis of buffer dimensioning. The results were published in [C4, C5, C8, C9, T2, T3].

5.3 Measurement-based effective bandwidth formula

In Chapter 4, the methodology used was to estimate the effective bandwidth of the traffic using real time traffic measurements. To do this, an approximate effective bandwidth formula was developed, which only requires the knowledge of the actual CLR and the utilization.

5.3.1 Three-point buffer monitoring method

When the traffic does not possess long-range dependence, the queue length distribution is asymptotically exponential. In this case Shioda and Saito showed [49] that the parameters β and η in Eq.(4.3) on page 62 can be estimated by monitoring the buffer occupancy levels at two different thresholds.

For a self-similar traffic input the queue length asymptotics are rather characterized by a Weibullian distribution, and in this case three thresholds are needed instead of two to monitor the queue length.

- I have proposed an algorithm to estimate the CLR in real time based on buffer measurements, which works for both the long-range and the short-range dependent case.

The actual buffer-monitoring thresholds depend on the buffer length, the CLR objective, and the length of the monitoring interval, as well as on the rate and burstiness of the traffic.

- I have proposed some guidelines and simple formulae for setting the buffer monitoring thresholds k_1, k_2 and k_3.

The proposed three-point buffer monitoring method was first validated by simulations using different types of traffic input.

5.3.2 Improved approximate effective bandwidth formula

When independent and bursty sources are multiplexed on a single link, the independence in the statistical variations of the individual sources offers the potential for a reduction in the bandwidth required for the combined stream. This achievable multiplexing gain is missing from the effective bandwidth approach.

Burstiness across different time scales is described by the Hurst parameter, hence the multiplexing gain is also affected by H. I have investigated how the parameter H could be used explicitly in the approximate effective bandwidth formula.

- I have given a modified version of the proposed approximate effective bandwidth formula to incorporate the scaling effect captured by H.

The results in Chapter 4 form the basis of a real time VP bandwidth control algorithm published in [J1] and [J2]. The proposed methods are subject to a Japanese patent application [T4] submitted.

Appendix A

Some Remarks on the Estimates

A.1 Variance of $\hat{m}(H)$

For the time being we assume that H is known exactly, $\hat{H} = H$. We have

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

where we have made use of the definition $\mathrm{E}\left[\mathbf{Z}\mathbf{Z}^{\mathrm{t}}\right] = \Gamma_H$.

A.2 Expectation of $\hat{a} = a(H)$

Next consider the estimator Eq.(2.23) for a, $\hat{a} = a(H)$. Again, for the time being we assume that H is known exactly and calculate the expectation of $\hat{a}$,

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

where we have used $\mathrm{E}\left[\mathbf{Z}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{Z}\right] = \mathrm{E}\left[\mathbf{N}^{\mathrm{t}}\mathbf{N}\right] = n$ since $\mathbf{Z} \sim \Gamma_H^{1/2}\mathbf{N}$ where $\mathbf{N}$ is a vector of independent standard Gaussian variables.

A.3 Variance of $\hat{a}(H)$

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

$$\begin{aligned} \mathrm{Var}\left[\hat{a}\right] &= \mathrm{E}\left[\hat{a}^2\right] - \mathrm{E}\left[\hat{a}\right]^2 \\ &= \mathrm{E}\left[\hat{a}^2\right] - \frac{(n-1)^2}{n^2}a^2. \end{aligned} \tag{A.3}$$

To calculate $\mathrm{E}\left[\hat{a}^2\right]$ first we rewrite Eq.(2.23) as follows:

$$\begin{aligned} \hat{a} &= \frac{1}{n(\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{t})}\left[(\mathbf{X}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{X})(\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{t}) - (\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{X})^2\right] \\ &= \frac{a}{n(\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{t})}\underbrace{\left[(\mathbf{Z}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{Z})(\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{t}) - (\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{Z})^2\right]}_{A(\mathbf{Z})} \end{aligned} \tag{A.4}$$

and now we have

$$\mathrm{E}\left[\hat{a}^2\right] = \frac{a^2}{n^2(\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{t})^2}\cdot\mathrm{E}\left[A(\mathbf{Z})^2\right]. \tag{A.5}$$

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

$$\begin{aligned}\mathrm{E}\left[A(\mathbf{Z})^2\right] &= \mathrm{E}\left[A\left(\frac{\partial}{\partial \mathbf{s}}\right)^2 e^{\mathbf{s}^{\mathrm{t}}\mathbf{Z}}\right]_{\mathbf{s}=0} \\ &= \left[A\left(\frac{\partial}{\partial \mathbf{s}}\right)^2 \mathrm{E}\left[e^{\mathbf{s}^{\mathrm{t}}\mathbf{Z}}\right]\right]_{\mathbf{s}=0} \\ &= \left[A\left(\frac{\partial}{\partial \mathbf{s}}\right)^2 M(\mathbf{s})\right]_{\mathbf{s}=0}\end{aligned} \tag{A.6}$$

with

$$M(\mathbf{s}) = \mathrm{E}\left[e^{\mathbf{s}^{\mathrm{t}}\mathbf{Z}}\right] = e^{\frac{1}{2}\mathbf{s}^{\mathrm{t}}\Gamma_H \mathbf{s}}. \tag{A.7}$$

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

$$\begin{aligned}(\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\nabla_{\mathbf{s}}) M(\mathbf{s}) &= (\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\Gamma_H\,\mathbf{s}) M(\mathbf{s}) \\ &= (\mathbf{t}^{\mathrm{t}}\,\mathbf{s}) M(\mathbf{s});\end{aligned} \tag{A.8}$$

$$(\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\nabla_{\mathbf{s}})^2 M(\mathbf{s}) = \left[(\mathbf{t}^{\mathrm{t}}\,\mathbf{s})^2 + (\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{t})\right] M(\mathbf{s}); \tag{A.9}$$

$$\begin{aligned}(\nabla_{\mathbf{s}}^{\mathrm{t}}\,\Gamma_H^{-1}\,\nabla_{\mathbf{s}}) M(\mathbf{s}) &= (\nabla_{\mathbf{s}}^{\mathrm{t}}\,\Gamma_H^{-1}\,\Gamma_H\,\mathbf{s}) M(\mathbf{s}) \\ &= (\nabla_{\mathbf{s}}^{\mathrm{t}}\,\mathbf{s}) M(\mathbf{s}) \\ &= \left[n + (\mathbf{s}^{\mathrm{t}}\,\Gamma_H\,\mathbf{s})\right] M(\mathbf{s}).\end{aligned} \tag{A.10}$$

Next we derive

$$\begin{aligned}A(\nabla_{\mathbf{s}}) M(\mathbf{s}) &= \left[(\nabla_{\mathbf{s}}^{\mathrm{t}}\,\Gamma_H^{-1}\,\nabla_{\mathbf{s}})(\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{t}) - (\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\nabla_{\mathbf{s}})^2\right] M(\mathbf{s}) \\ &= \left\{(\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{t})\left[n + (\mathbf{s}^{\mathrm{t}}\,\Gamma_H\,\mathbf{s})\right] - \left[(\mathbf{t}^{\mathrm{t}}\,\mathbf{s})^2 + (\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{t})\right]\right\} M(\mathbf{s}) \\ &= \left\{(\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{t})\left[(n-1) + (\mathbf{s}^{\mathrm{t}}\,\Gamma_H\,\mathbf{s})\right] - (\mathbf{t}^{\mathrm{t}}\,\mathbf{s})^2\right\} M(\mathbf{s}),\end{aligned} \tag{A.11}$$

and now we are ready to calculate

$$\begin{aligned}A(\nabla_{\mathbf{s}})^2 M(\mathbf{s}) &= \underbrace{(n-1)(\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{t}) A(\nabla_{\mathbf{s}}) M(\mathbf{s})}_{T_1(\mathbf{s})} \\ &+ \underbrace{A(\nabla_{\mathbf{s}})\left[(\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{t})(\mathbf{s}^{\mathrm{t}}\,\Gamma_H\,\mathbf{s}) - (\mathbf{t}^{\mathrm{t}}\,\mathbf{s})^2\right] M(\mathbf{s})}_{T_2(\mathbf{s})}\end{aligned} \tag{A.12}$$

According to Eq.(A.6) we need to get $T_1(0)$ and $T_2(0)$, As for the first term we have

$$T_1(0) = (n-1)^2(\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{t})^2. \tag{A.13}$$

However, for $T_2(0)$ we have

$$\begin{aligned} T_2(\mathbf{s}) &= A(\nabla_{\mathbf{s}})\underbrace{\left[(\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{t})(\mathbf{s}^{\mathrm{t}}\,\Gamma_H\,\mathbf{s}) - (\mathbf{t}^{\mathrm{t}}\,\mathbf{s})^2\right]}_{T_3(\mathbf{s})}M(\mathbf{s}) \\ &= \left[(\nabla_{\mathbf{s}}^{\mathrm{t}}\,\Gamma_H^{-1}\,\nabla_{\mathbf{s}})(\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{t}) - (\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\nabla_{\mathbf{s}})^2\right]T_3(\mathbf{s})M(\mathbf{s}). \end{aligned} \tag{A.14}$$

To solve Eq.(A.14) it is useful to calculate the following terms:

$$\begin{aligned} (\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\nabla_{\mathbf{s}})T_3(\mathbf{s})M(\mathbf{s}) &= (\mathbf{t}^{\mathrm{t}}\,\mathbf{s})T_3(\mathbf{s})M(\mathbf{s}) \\ &+ \underbrace{\left[2(\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{t})(\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\Gamma_H\,\mathbf{s}) - 2(\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{t})(\mathbf{t}^{\mathrm{t}}\,\mathbf{s})\right]}_{0}M(\mathbf{s}) \\ &= \underbrace{\left[(\mathbf{t}^{\mathrm{t}}\,\mathbf{s})(\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{t})(\mathbf{s}^{\mathrm{t}}\,\Gamma_H\,\mathbf{s}) - (\mathbf{t}^{\mathrm{t}}\,\mathbf{s})^3\right]}_{T_4(\mathbf{s})}M(\mathbf{s}); \end{aligned} \tag{A.15}$$

$$\begin{aligned} (\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\nabla_{\mathbf{s}})^2T_3(\mathbf{s})M(\mathbf{s}) &= (\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\nabla_{\mathbf{s}})T_4(\mathbf{s})M(\mathbf{s}) \\ &= (\mathbf{t}^{\mathrm{t}}\,\mathbf{s})T_4(\mathbf{s})M(\mathbf{s}) \\ &+ 3(\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{t})\left[(\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{t})(\mathbf{s}^{\mathrm{t}}\,\Gamma_H\,\mathbf{s}) - (\mathbf{t}^{\mathrm{t}}\,\mathbf{s})^2\right]M(\mathbf{s}); \end{aligned} \tag{A.16}$$

$$\begin{aligned} (\nabla_{\mathbf{s}}\,\Gamma_H^{-1}\,\nabla_{\mathbf{s}})T_3(\mathbf{s})M(\mathbf{s}) &= T_3(\mathbf{s})(\nabla_{\mathbf{s}}^{\mathrm{t}}\mathbf{s})M(\mathbf{s}) \\ &+ 2(\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{t})(\nabla_{\mathbf{s}}^{\mathrm{t}}\mathbf{s})M(\mathbf{s}) - 2(\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\nabla_{\mathbf{s}})(\mathbf{t}^{\mathrm{t}}\,\mathbf{s})M(\mathbf{s}) \\ &= T_3(\mathbf{s})\left[n + (\mathbf{s}^{\mathrm{t}}\,\Gamma_H\,\mathbf{s})\right]M(\mathbf{s}) \\ &+ 2(\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{t})\left[n + (\mathbf{s}^{\mathrm{t}}\,\Gamma_H\,\mathbf{s})\right]M(\mathbf{s}) - 2\left[(\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{t}) + (\mathbf{t}^{\mathrm{t}}\,\mathbf{s})^2\right]M(\mathbf{s}). \end{aligned} \tag{A.17}$$

Note that in Eq.(A.17) we used the results from Eq.(A.10). Since all terms in Eq.(A.15) are zero when $\mathbf{s} = 0$, from Eq.(A.14) using Eq.(A.17) and the fact that $T_3(0) = 0$ we get

$$T_2(0) = 2(n-1)(\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{t})^2. \tag{A.18}$$

Substituting this result and Eq.(A.13) into Eq.(A.6) using Eq.(A.12) we have

$$\mathrm{E}\left[A(\mathbf{Z})\right] = T_1(0) + T_2(0) = (n^2-1)(\mathbf{t}^{\mathrm{t}}\,\Gamma_H^{-1}\,\mathbf{t})^2, \tag{A.19}$$

and from Eq.(A.5) we get

$$\mathrm{E}\left[\hat{a}^2\right] = \frac{n^2-1}{n^2}a^2. \tag{A.20}$$

Finally, from Eq.(A.3) we get the variance of the estimate $\hat{a}$ as

$$\mathrm{Var}\left[\hat{a}\right] = \frac{2(n-1)}{n^2}a^2. \tag{A.21}$$

Appendix B

Inverse and Determinant of Band Matrices

B.1 General case

Consider the following symmetrically partitioned hypermatrix

$$\mathbf{H} = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix}, \tag{B.1}$$

where $\mathbf{B}(r \times r)$ and $\mathbf{C}(q \times q)$ matrices are quadratic. Assume now, that the inverse of $\mathbf{H}$ is known and is of the form

$$\mathbf{H}^{-1} = \begin{pmatrix} \mathbf{U} & \mathbf{L} \\ \mathbf{S} & \mathbf{V} \end{pmatrix}, \tag{B.2}$$

where $\mathbf{S}(r \times r)$ and $\mathbf{L}(q \times q)$ are quadratic. The determinant of submatrix $\mathbf{C}$ of matrix $\mathbf{H}$ can be calculated as [46]

$$|\mathbf{C}| = (-1)^{rq} \begin{vmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{vmatrix} \cdot |\mathbf{S}| \tag{B.3}$$

Assume further that matrix $\mathbf{C}$ is an n-by-n band matrix of the form

$$\mathbf{C} = \begin{pmatrix} c_1^{(1)} & c_1^{(2)} & \cdots & c_1^{(p)} & 0 & \cdots & 0 \\ c_1^{(2)} & c_2^{(1)} & \cdots & c_2^{(p-1)} & c_2^{(p)} & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & 0 \\ c_1^{(p)} & c_2^{(p-1)} & \cdots & c_p^{(1)} & c_p^{(2)} & \ddots & c_{n-p+1}^{(p)} \\ 0 & c_2^{(p)} & \cdots & c_p^{(2)} & c_{p+1}^{(1)} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & c_{n-1}^{(2)} \\ 0 & \cdots & 0 & c_{n-p+1}^{(p)} & \cdots & c_{n-1}^{(2)} & c_n^{(1)} \end{pmatrix}. \tag{B.4}$$

To calculate the determinant $|\mathbf{C}|$, Eq.(B.3) is suitable with properly chosen submatrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{D}$ suitable for our purpose. Let $\mathbf{A}$ and $\mathbf{D}$ be of the form

$$\mathbf{A}_{(p-1)\times n} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots & & \vdots \\ 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \end{pmatrix}, \quad \mathbf{D}_{n\times(p-1)} = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix}, \tag{B.5}$$

and $\mathbf{B}$ is a $(p-1)$-by-$(b-1)$ matrix full of zeros. Hence, the determinant of the hypermatrix $\mathbf{H}$ can be calculated as

$$\begin{vmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{vmatrix} = \prod_{i=1}^{n-p+1} c_i^{(p)}. \tag{B.6}$$

Thus, from Eq.(B.3) and Eq.(B.6) we get

$$|\mathbf{C}| = (-1)^{n(p-1)} \cdot |\mathbf{S}| \cdot \prod_{i=1}^{n-p+1} c_i^{(p)}. \tag{B.7}$$

Eq.(B.7) contains the determinant of the $(p-1)$-by-$(p-1)$ submatrix $\mathbf{S}$. The matrix $\mathbf{S}$ can be calculated by solving a recursive set of equations. The following section gives the algorithm for the simplest case when $p = 2$. The idea can easily be generalized for larger values of p.

B.2 $p = 2$ case

Let $\mathbf{C}$ be a symmetric continuant matrix of the form

$$\mathbf{C} = \begin{pmatrix} a_1 & b_1 & 0 & \cdots & 0 \\ b_1 & a_2 & b_2 & \ddots & \vdots \\ 0 & b_2 & a_3 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & b_{n-1} \\ 0 & \cdots & 0 & b_{n-1} & a_n \end{pmatrix} \tag{B.8}$$

with $b_i \neq 0$. According to Eq.(B.2) the inverse of the lower-triangular hypermatrix $\mathbf{H}$ is of the form

$$\left(\begin{array}{ccccc|c} 1 & 0 & 0 & \cdots & 0 & 0 \\ \hline a_1 & b_1 & 0 & \cdots & 0 & 0 \\ b_1 & a_2 & b_2 & \ddots & \vdots & \vdots \\ 0 & b_2 & a_3 & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & b_{n-1} & 0 \\ 0 & \cdots & 0 & b_{n-1} & a_n & 1 \end{array}\right)^{-1} = \left(\begin{array}{c|cccc} u_1 & 0 & \cdots & 0 & 0 \\ u_2 & l_{21} & \ddots & \vdots & \vdots \\ \vdots & \vdots & \ddots & 0 & 0 \\ u_n & l_{n1} & \cdots & l_{n(n-1)} & 0 \\ \hline s & v_1 & \cdots & v_{n-1} & v_n \end{array}\right), \tag{B.9}$$

or, equivalently, with supressed notations we have

$$\begin{pmatrix} \mathbf{e}_1^{\mathrm{t}} & 0 \\ \mathbf{C} & \mathbf{e}_n \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{u} & \mathbf{L} \\ s & s\mathbf{v}^{\mathrm{t}} \end{pmatrix}. \tag{B.10}$$

The submatrix $\mathbf{S}$ (the single value s in the present case) can be calculated using the following recursive equations:

$$\begin{aligned} u_1 &= 1, \\ u_2 &= -\frac{1}{b_1} a_1, \\ u_{i+1} &= -\frac{1}{b_i}(a_i u_i + b_{i-1} u_{i-1}), \quad i = 2, 3, \ldots, n-1, \\ s &= a_n u_n + b_{n-1} u_{n-1}. \end{aligned} \tag{B.11}$$

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

$$|\mathbf{C}| = (-1)^n \cdot s \cdot \prod_i^{n-1} b_i. \tag{B.12}$$

Note, that the following approximation works reasonably well in practice:

$$|\mathbf{C}| \approx \prod_i^{n-1} |b_i|. \tag{B.13}$$

Appendix C

Derivation of the Effective Bandwidth Formula

The following derivation was published in [49].

Provided that the input process satisfies some mild technical conditions, the effective bandwidth function $\alpha(\theta)$ (see Eq.(4.6) on page 62) strictly increases for θ [8]. Hence, η is given by the unique solution of the following equation:

$$\alpha(\eta) = C. \tag{C.1}$$

The effective bandwidth function $\alpha(\eta)$ can be expanded by η (Cumulant Expansion) as follows:

$$\begin{aligned} \alpha(\eta) &= \lim_{t\to\infty} \frac{1}{t\eta} \log \mathrm{E}\left[e^{A_t\eta}\right] \\ &= \lim_{t\to\infty} \frac{1}{t\eta} \sum_{k=1}^{\infty} \frac{\langle A_t^k \rangle_c}{k!} \eta^k, \end{aligned} \tag{C.2}$$

where $\langle A_t^k \rangle_c$ is called the k-th cumulant. It is easy to see the following:

$$\langle A_t^k \rangle_c = \mathrm{E}\left[A_t\right], \tag{C.3}$$

$$\langle A_t^1 \rangle_c = \mathrm{Var}\left[A_t\right]. \tag{C.4}$$

Eq.(C.2) is further simplified as follows:

$$\alpha(\eta) = A \sum_{k=0}^{\infty} d_k^{(2)} \eta^k, \tag{C.5}$$

where

$$d_k^{(2)} = \lim_{t\to\infty} \frac{\langle A_t^{k+1} \rangle_c}{(k+1)! \langle A_t^1 \rangle_c}. \tag{C.6}$$

Regarding the asymptotic constant β, we also have the following similar expansion by η [5]:

$$\log \beta = \sum_{k=1}^{\infty} d_k^{(1)} \eta^k, \tag{C.7}$$

where

$$d_k^{(1)} = -\lim_{t\to\infty} \left\{ \langle A_t^k \rangle_c - \left(\lim_{t\to\infty} \frac{\langle A_t^k \rangle_c}{t} \right) t \right\}. \tag{C.8}$$

Once we know the coefficients, $\{d_i^{(1)}\}$ and $\{d_i^{(2)}\}$, the effective bandwidth of the source can be estimated simply. For example, by taking only the first term in the expansion of η in Eq.(C.8), we have

$$\begin{aligned} \eta_{obj} &= \frac{\log \beta - \log CLR_{obj}}{K} \\ &\approx \frac{d_1^{(1)} \eta_{obj} - \log CLR_{obj}}{K}, \end{aligned} \tag{C.9}$$

and from this we have

$$\eta_{obj} = \frac{\log CLR_{obj}}{d_1^{(1)} - K}. \tag{C.10}$$

We therefore have

$$\alpha(\eta_{obj}) = A \left\{ 1 + d_1^{(2)} \frac{\log CLR_{obj}}{d_1^{(1)} - K} + d_2^{(2)} \left(\frac{\log CLR_{obj}}{d_1^{(1)} - K} \right)^2 + \cdots \right\}. \tag{C.11}$$

Approximating $\alpha(\eta)$ further by taking only the first two terms in Eq.(C.11), we get

$$\alpha(\eta) \approx A \{1 + d \log CLR_{obj}\} \tag{C.12}$$

with

$$d \stackrel{\text{def}}{=} \frac{d_1^{(2)}}{d_1^{(1)} - K}. \tag{C.13}$$

Note, that for Gaussian processes

$$d_1^{(1)} = 0, \tag{C.14}$$

$$d_1^{(2)} = \frac{1}{2} \cdot \text{IDC} \stackrel{\text{def}}{=} \frac{1}{2} \lim_{t\to\infty} \frac{\text{Var}\left[A(t)\right]}{\text{E}\left[A(t)\right]}, \tag{C.15}$$

$$d_i^{(2)} = 0 \quad \text{for all } i \geq 2. \tag{C.16}$$

(The first two are valid even for the case where $A(t)$ is not Gaussian.)

Bibliography

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

[2] A.T. Andersen and B.F. Nielsen, *On the implications of certain random permutations of interarrival times or counts in a point process*, technical document COST 257TD(020), Rome, Italy, January 1998.

[3] J. Beran, R. Sherman, M.S. Taqqu, and W. Willinger, *Long-range dependence in variable-bit-rate video traffic*, IEEE Trans. on Communications **43** (1995), no. 2/3/4, 1566–1579.

[4] Jan Beran, *Statistics for long-memory processes*, Chapman & Hall, One Penn Plaza, New York, NY 10119, 1995.

[5] D.D. Botvich and N.G. Duffield, *Large deviations, the shape of the loss curve, and economies of scale in large multiplexers*, ftp://stp01.stp.dias.ie/DAPG/dapg9412.ps, 1995.

[6] F. Brichet, J. Roberts, A Simonian, and D. Veitch, *Heavy traffic analysis of a storage model with long range dependent on/off sources*, Queueing Systems **23** (1996), 197–215.

[7] Traffic Cell by Cell: Experiences and Pleriminary Findings from BAGNet, *Measuring ATM*, Proc., PMCCN'97 (Tsukuba, Japan), 1997, pp. 91–110.

[8] C.S. Chang, *Stability, queue length, and delay of deterministic and stochastic queueing networks*, IEEE Trans. on Automatic Control **39** (1994), no. 5, 913–931.

[9] G.L. Choudhury, D.M. Lucantoni, and W. Whitt, *Squeezing the most out of ATM*, IEEE Trans. on Communications **44** (1996), no. 2, 203–217.

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

[11] ———, *A class of second-order stationary self-similar processes for 1/f phenomena*, IEEE Trans. on Signal Processing **45** (1997), no. 2, 396–410.

[12] D.R. Cox and P.A.W. Lewis, *The statistical analysis of series of events*, Menthuen, 1966.

[13] I. Daubechies, *Ten lectures on wavelets*, SIAM, Philadelphia (PA), 1992.

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

[15] M. Deriche and A.A. Tewfik, *Maximum likelihood estimation of the parameters of discrete fractionally differenced Gaussian noise process*, IEEE Trans. Signal Proc. **41** (1993), no. 10, 2977–2989.

[16] 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*, Proc., Computational Techniques and Applications Conference (CTAC93), World Scientific, July 1993, pp. 184–192.

[17] N.G. Duffield, *Exponential bounds for queues with Markovian arrivals*, Queueing Systems **17** (1994), 413–430.

[18] N.G. Duffield and N. O'Connell, *Large deviations and overflow probabilities for the general single-server queue, with applications*, Proc., Cam. Phil. Soc., vol. 118, 1994, pp. 363–374.

[19] D.E. Duffy, A.A. McIntosh, and W. Willinger, *Statistical analysis of CCSN/SS7 traffic data from working CCS subnetworks*, IEEE Journal on Selected Areas in Communications **12** (1994), no. 3, 544–551.

[20] Z. Dziong, M. Juda, and L.G. Mason, *A framework for bandwidth management in ATM networks—aggregate equivalent bandwidth estimation approach*, IEEE/ACM Trans. on Networking **5** (1997), no. 1, 134–147.

[21] A. Elwalid, D. Heyman, T.V. Lakshman, D. Mitra, and A. Weiss, *Fundamental bounds and approximations for ATM multiplexers with applications to video teleconferencing*, IEEE JSAC **13** (1995), no. 6, 1004–1016.

[22] A. Erramilli, O. Narayan, and W. Willinger, *Experimental queueing analysis with long-range dependent packet traffic*, IEEE/ACM Trans. on Networking **4** (1996), no. 2, 209–223.

[23] A. Erramilli, R.P. Singh, and P. Pruthi, *Chaotic maps as models of packet traffic*, Proc., 14th International Teletraffic Congress ITC'14, June 1994, pp. 329–338.

[24] K. Meier-Hellstern et al., *Traffic models for ISDN data users: Office automation application*, Proc., 13th Int. Teletraffic Congress (Copenhagen, Denmark), 1991, pp. 167–172.

[25] P.W. Glynn and W. Whitt, *Logarithmic asymptotics for steady-state tail probabilities in a single-server queue*, J. Appl. prob. **31** (1993), no. A, 131–159.

[26] M. Grossglauer and J-C. Bolot, *On the relevance of long-range dependence in network traffic*, Proc., SIGCOMM'96 (CA, USA), 1996, pp. 15–24.

[27] D.P. Heyman and T.V. Lakshman, *What are the implications of long-range dependence for VBR-video traffic engineering?*, IEEE/ACM Transactions on Networking **4** (1996), no. 3, 301–317.

[28] K. Kobayashi and Y. Takahashi, *The tail probability of a Gaussian fluid queue under finite measurement of input processes*, Proc., Int. Conference on the Performance and Management of Complex Communication Networks—PMCCN'97 (Tsukuba, Japan), November 1997, pp. 57–72.

[29] K.R. Krishnan, A.L. Neidhardt, and A. Erramilli, *Scaling analysis in traffic management of self-similar processes*, Proc., ITC'97 (Washington), 1997, pp. 1087–1096.

[30] J.W. Lamperti, *Semi-stable stochastic processes*, Trans. Am. Math. Soc. **104** (1962), 62–78.

[31] W.E. Leland, M.S. Taqqu, W. Willinger, and D.V. Wilson, *On the self-similar nature of Ethernet traffic (extended version)*, IEEE/ACM Transactions on Networking **2** (1993), no. 1, 1–15.

[32] K. Lindberg, *The FUNET network connection to foreign countries*, CSCNews **7** (1995), no. 4, 19.

[33] B.B. Mandelbrot, *The fractal geometry of nature*, W.H. Freeman and Co., San Francisco, 1982.

[34] P. Mannersalo and I. Norros, *Multifractal analysis of real ATM traffic: A first look*, COST257TD(97)19, January 1997.

[35] ———, *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.

[36] O. Melteig, *Introduction to the Parasol project*, Proc., The 9th Nordic Teletraffic Seminar (Norwegian Telecom, Research Department), August 1990.

[37] S. Molnár and T.D. Dang, *Pitfalls in long range dependence testing and estimation*, Proc., GLOBECOM'00 (San Francisco, CA, USA), November 2000.

[38] S. Molnár, A. Vidács, and A.A. Nilsson, *Bottlenecks on the way towards fractal characterization of network traffic: Estimation and interpretation of the hurst parameter*, Proc., PMCCN'97 (Tsukuba, Japan), 1997, pp. 125–144.

[39] A.L. Neidhardt and A. Erramilli, *Shaping and policing of fractal traffic*, Proc., ITC Specialists Seminar on Control in Communications (Lund, Sweden), 1996, pp. 253–264.

[40] B. Ninness, *Maximum likelihood estimation of the parameters of fractional Brownian motions*, Proc., 34th Conference on Decision & Control (New Orleans, LA), December 1995, pp. 4018–4023.

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

[42] ———, *On the use of fractional Brownian motion in the theory of connectionless networks*, IEEE Journal on Selected Areas in Communications **13** (1995), no. 6, 953–962.

[43] T. Éltető and S. Molnár, *On the distribution of round-trip delays in TCP/IP networks*, Proc., 24th Annual Conference on Local Computer Networks (LCN'99) (Lowell/Boston, Massachusetts, USA), October 1999.

[44] V. Paxson and S. Floyd, *Wide-area traffic: The failure of Poisson modeling*, IEEE/ACM Transactions on Networking **3** (1995), no. 3.

[45] 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.

[46] P. Rózsa, *Linear algebra and its applications*, 3 ed., University Press, Budapest, 1991, (in Hungarian).

[47] B.K. Ryu and A. Elwalid, *The importance of long-range dependence of VBR video traffic in ATM traffic engineering: Myths and realities*, Proc., SIGCOMM'96 (CA, USA), 1996, pp. 3–14.

[48] G. Samorodnitsky and M.S. Taqqu, *Stable non-Gaussian processes: Stochastic models with infinite variance*, Chapman & Hall, New York, One Penn Plaza, New York, NY 10119, 1994.

[49] S. Shioda and H. Saito, *Real-time cell loss ratio estimation and its application to ATM traffic controls*, Proc., IEEE INFOCOM'97 (Kobe, Japan), 1997, pp. 1074–1081.

[50] L. Takács, *Introduction to the theory of queues*, Oxford University Press, New York, 1962.

[51] M.S. Taqqu, W. Willinger, and R. Sherman, *Proof of a fundamental result in self-similar traffic modeling*, Computer Communication Review (1997), no. 27, 5–23.

[52] B. Tsybakov and N.D. Georganas, *On self-similar traffic in ATM queue: Definitions, overflow probability bound, and cell delay distribution*, IEEE/ACM Trans. on Networking **5** (1997), no. 3, 397–409.

[53] D. Veitch and P. Abry, *A wavelet based joint estimator of the parameters of long-range dependence*, submitted to Special Issue of the IEEE Trans. on Information Theory, Multiscale Statistical Signal Analysis and its Applications, 1998.

[54] C. Walsh and B. McGurk, *Investigations of the performance of a measurement-based connection admission control algorithm*, Proc., 5th IFIP Workshop on Performance Modeling and Evaluation of ATM Networks (Ilkley, West Yorkshire, UK), 1997.

[55] W. Willinger, M.S. Taqqu, W.E. Leland, and D.V. Wilson, *Self-similarity in high-speed packet traffic: Analysis and modeling of Ethernet traffic measurements*, Statistical Science **10** (1995), no. 1, 67–85.

Publications

Journal papers

[J1] **A. Vidács**, Sz. Malomsoky, and H. Saito. A simple adaptive bandwidth control for real traffic. *Advances in Performance Analysis*, 2(1):21–44, 1999.

[J2] Sz. Malomsoky, **A. Vidács**, and H. Saito. Real time VP bandwidth control for long-range dependent traffic. *Int. Journal of Communication Systems*, (12):229–247, 1999.

[J3] T.D. Dang, S. Molnár, and **A. Vidács**. Investigation of fractal properties in data traffic. *Journal on Communications*, XLIX:12–18, November-December 1998.

[J4] **A. Vidács**, Zs. Kenesi, A. Rétfalvi, P. Pozsgai, and S. Molnár. Traffic description for telecom networks. *Hungarian Telecommunication Periodicals, Selected Papers*, pages 18–25, 1995.

[J5] Zs. Kenesi, **A. Vidács**, and S. Molnár. Fractals in telecommunications. *Magyar Távközlés*, 6(7):3–9, July 1995. (in Hungarian)

Conference papers

[C1] **A. Vidács** and J.T. Virtamo. Parameter estimation of geometrically sampled fractional Brownian traffic. In *Proc., INFOCOM 2000*, Tel-Aviv, Israel, March 2000.

[C2] **A. Vidács** and J.T. Virtamo. Maximum likelihood estimation of the parameters of fractional Brownian traffic with geometrical sampling. In *Proc., IFIP TC6 5th Int. Conf. on Broadband Communications (BC'99)*, pages 51–62, Hong Kong, November 1999.

[C3] S. Molnár, T.D. Dang, and **A. Vidács**. Heavy tailedness, long-range dependence and self-similarity in data traffic. In *Proc., 7th Int. Conf. on Telecommunication Systems, Modelling and Analysis*, pages 10–21, Nashville, Tenessee, USA, March 1999.

[C4] **A. Vidács**, S. Molnár, G. Gordos, and I. Cselényi. The impact of long-range dependence on cell loss in an ATM wide area network. In *Proc., GLOBECOM'98*, Sydney, Australia, November 1998.

[C5] S. Molnár, **A. Vidács**, and A.A. Nilsson. Bottlenecks on the way towards fractal characterization of network traffic: Estimation and interpretation of the Hurst parameter. In *proc., PMCCN'97*, pages 125–144, Tsukuba, Japan, 1997.

[C6] **A. Vidács**, Sz. Malomsoky, and H. Saito. Real time cell loss ratio estimation for bursty and self-similar traffic. In *Proc., Int. Conference on the Performance and Management of Complex Communication Networks (PMCCN'97), Workshop 2*, pages 7–14, Tsukuba, Japan, November 1997.

[C7] Sz. Malomsoky, **A. Vidács**, and H. Saito. Bandwidth control and its applicability based on queue length monitoring. In *Proc., Int. Conference on the Performance and Management of Complex Communication Networks (PMCCN'97), Workshop 2*, pages 17–21, Tsukuba, Japan, November 1997.

[C8] S. Molnár and **A. Vidács**. On modeling and shaping self-similar ATM traffic. In *Proc., 15th International Teletraffic Congress (ITC15)*, Washington, USA, July 1997.

[C9] S. Molnár, **A. Vidács**, and A. Hegedűs. Meeting a challenge: Modeling self-similar LAN/MAN traffic. In *Proc., 8th IEEE Workshop on Local and Metropolitan Area Networks*, Berlin/Potsdam, Germany, August 1996.

Patents and Technical documents

[T1] **A. Vidács** and J.T Virtamo. Time domain MLE of the parameters of FBM traffic. Technical Document, Report 99/3, Helsinki University of Technology, Lab. of Telecommunications Technology, Espoo, 1999

[T2] S. Molnár, **A. Vidács**, and I. Cselényi. Queueing performance in the presence of long-range dependence. Technical Document TD(98)051, COST 257, Granada, Spain, September 1998.

[T3] S. Molnár and **A. Vidács**. How to characterize Hursty traffic? Technical Document TD(98), COST 257, Rome, Italy, January 1998.

[T4] Sz. Malomsoky, **A. Vidács** and H. Saito. Virtual path bandwidth control apparatus and virtual path bandwidth dimensioning method. Japanese patent application 9-318654, December 1997.

[T5] Sz. Malomsoky and **A. Vidács**. Further results on VP Bandwidth Control for Long-Range Dependent Traffic. Technical Report, November 1997.

[T6] **A. Vidács**. Matlab code for the ML estimation of the parameters of fractional Brownian traffic. http://keskus.hut.fi/tutkimus/com2/fbm/index.shtml