

Simulation of fractional Brownian motion

Ton Dieker

ton@cwi.nl

CWI

P.O. Box 94079
1090 GB Amsterdam
The Netherlands

and

University of Twente
Department of Mathematical Sciences
P.O. Box 217
7500 AE Enschede
The Netherlands

Preface

In recent years, there has been great interest in the simulation of long-range dependent processes, in particular fractional Brownian motion. Motivated by applications in communications engineering, I wrote my master's thesis on the subject in 2002. Since many people turned out to be interested in various aspects of fractional Brownian motion, I decided to update my thesis and make it publicly available. Some references are added and the section on spectral simulation is rewritten according to the paper [22].

Fractional Brownian motion is not only of interest for communications engineers. Its properties have been investigated by researchers in theoretical physics, probability, statistics, hydrology, biology, and many others. As a result, the techniques that have been used to study this Gaussian process are quite diverse, and it may take some effort to study them. Undoubtedly, this also makes the field more interesting.

This report gives an introduction to generation and estimation of fractional Brownian motion. However, as the literature on the subject is quite extensive (see, for instance, [24]), it has not been my goal to write a complete introduction. Running the risk of satisfying nobody, it is my hope that this report provides some help to find a way through the literature. Since it is written on the level of a master's student, limited background is required.

I would like to take this opportunity to thank my thesis advisor, Michel Mandjes, for many discussions and for his help to prepare this manuscript.

Finally, I refer to my homepage <http://www.cwi.nl/~ton> for the C code that was used to write this report.

Amsterdam, February 2004
Ton Dieker

It is my great pleasure to thank Antoine Ayache for valuable discussions on wavelet methods for simulating fractional Brownian motion. He pointed me out that the section on wavelets contained a serious error, which led me to revise it.

Amsterdam, September 2004
TD

Contents

1	Introduction and motivation	1
1.1	A closer look at network traffic	1
1.1.1	Analysis of traffic traces	1
1.1.2	Self-similarity and long-range dependence	3
1.2	Fractional Brownian motion and fractional Gaussian noise	4
1.2.1	Definitions and basic properties	5
1.2.2	Spectral densities	7
1.2.3	Another long-range dependent process	9
1.3	Implications for network models	10
1.4	Outline and scientific contribution	11
2	Simulation methods	13
2.1	Exact methods	13
2.1.1	The Hosking method	13
2.1.2	The Cholesky method	14
2.1.3	The Davies and Harte method	15
2.2	Approximate methods	17
2.2.1	Stochastic representation method	17
2.2.2	Aggregating packet processes	17
2.2.3	(Conditionalized) Random Midpoint Displacement	18
2.2.4	Spectral simulation, the Paxson method and the approximate circulant method	19
2.2.5	Wavelet-based simulation	24
2.2.6	Other methods	27
	Appendix: Proof of $\xi_n^{(t)} \rightarrow \xi_n$ in L^2	29
3	Estimation and testing	31
3.1	Estimation	31
3.1.1	Aggregated variance method	31
3.1.2	Absolute moments method	33
3.1.3	Discrete variations	33
3.1.4	The Higuchi method	34
3.1.5	Periodogram method	34
3.1.6	Variance of the regression residuals	36
3.1.7	R/S analysis	36
3.1.8	Whittle's method	38
3.1.9	Wavelet method	38
3.2	Testing	40
3.2.1	A goodness-of-fit test for the spectral density	40
3.2.2	A chi-square test for fractional Gaussian noise	41
4	Evaluation of approximate simulation methods	43
4.1	Error analysis	43
4.1.1	The approximate circulant method	44
4.1.2	The RMD _{l,r} method	48
4.1.3	The wavelet method	51
4.2	Stationarity and covariance analysis	52
4.3	Testing the approximate samples	57
4.3.1	The goodness-of-fit test for the spectral density	57

4.3.2	The chi-square test	58
4.4	Long-range dependence in approximate samples	58
4.4.1	The Wilcoxon test	60
4.4.2	The Freund-Ansari-Bradley test	60
4.4.3	Estimation of H for the approximate methods	61
4.5	Time complexity analysis	64
5	Conclusions	67
Bibliography		69

CHAPTER 1

Introduction and motivation

Queueing theory plays an important role in the design of telecommunication networks. Classical queueing models often give insightful results, for example on (long-term) delays, packet losses, and buffer content distributions. However, research has shown that the assumptions in traditional models are not valid in some cases. Unfortunately, these cases are particularly important for data communication networks. Those assumptions concern mainly the arrival process of data packets.

In this report, we will study a random process, fractional Brownian motion, that models the arrivals of network packets better than classical models (at least for some types of traffic), but is parsimonious in the sense that only few parameters describe its statistical behavior. Research has shown that the use of fractional Brownian motion instead of a traditional model has impact on queueing behavior; it affects several aspects of queueing theory (e.g., buffer sizing, admission control and congestion control). The use of conventional models (e.g., Poisson-type models) results in optimistic performance predictions and an inadequate network design.

Since traditional models have extensively been studied in the past decades, relatively much is known about their properties, although there are still many open problems. In contrast, few theoretical results exist for queueing models based on fractional Brownian motion. Simulation studies that make use of generated fractional Brownian motion traces are therefore of crucial importance, especially for complex queueing systems.

It is the aim of this report to evaluate several simulation methods for fractional Brownian motion. Although some methods that simulate fractional Brownian motion are known, methods that simulate this process ‘approximately’ have been proposed to reduce the computation time. Therefore, we will focus on the question in what respect the approximate differ from exact samples.

In this first chapter, we start with the analysis of real network traffic traces to clarify that conventional assumptions do not hold. Thereafter, we define fractional Brownian motion and study its basic properties. In Section 1.3, we get some feeling for the impact of the use of fractional Brownian motion in a teletraffic framework by studying a simple but insightful queueing model. Section 1.4 outlines this report, and addresses its scientific contribution.

1.1 A closer look at network traffic

1.1.1 Analysis of traffic traces

Willinger *et al.* [57] have analyzed high-quality Local Area Network (LAN) Ethernet traces. During more than one day, every 10 milliseconds was measured how many packets and bytes passed the monitoring system at Bellcore Research. (Data is sent in packets of bytes across the network.)

By studying the data, the authors find arguments for the presence of so-called self-similarity and long-range dependence. Before giving definitions, we will make this intuitively clear by the plots in Figure 1.1, which are taken from [57]. Starting with a time unit of 100 seconds (the upper row), each subsequent row is obtained from the previous one by increasing the time resolution by a factor of 10 and by concentrating on a randomly chosen subinterval, indicated by a shade. The lower row corresponds to the finest time scale (10 milliseconds). The plots are made for actual Ethernet traffic (left column), a synthetic trace generated from a Poisson-type traffic model (middle column), and a synthetic trace generated from an appropriately chosen self-similar traffic model with a single parameter (right column). For a given time resolution (or, equivalently, interval length), the number of data packets arriving in each interval is computed and depicted as tiny bins.

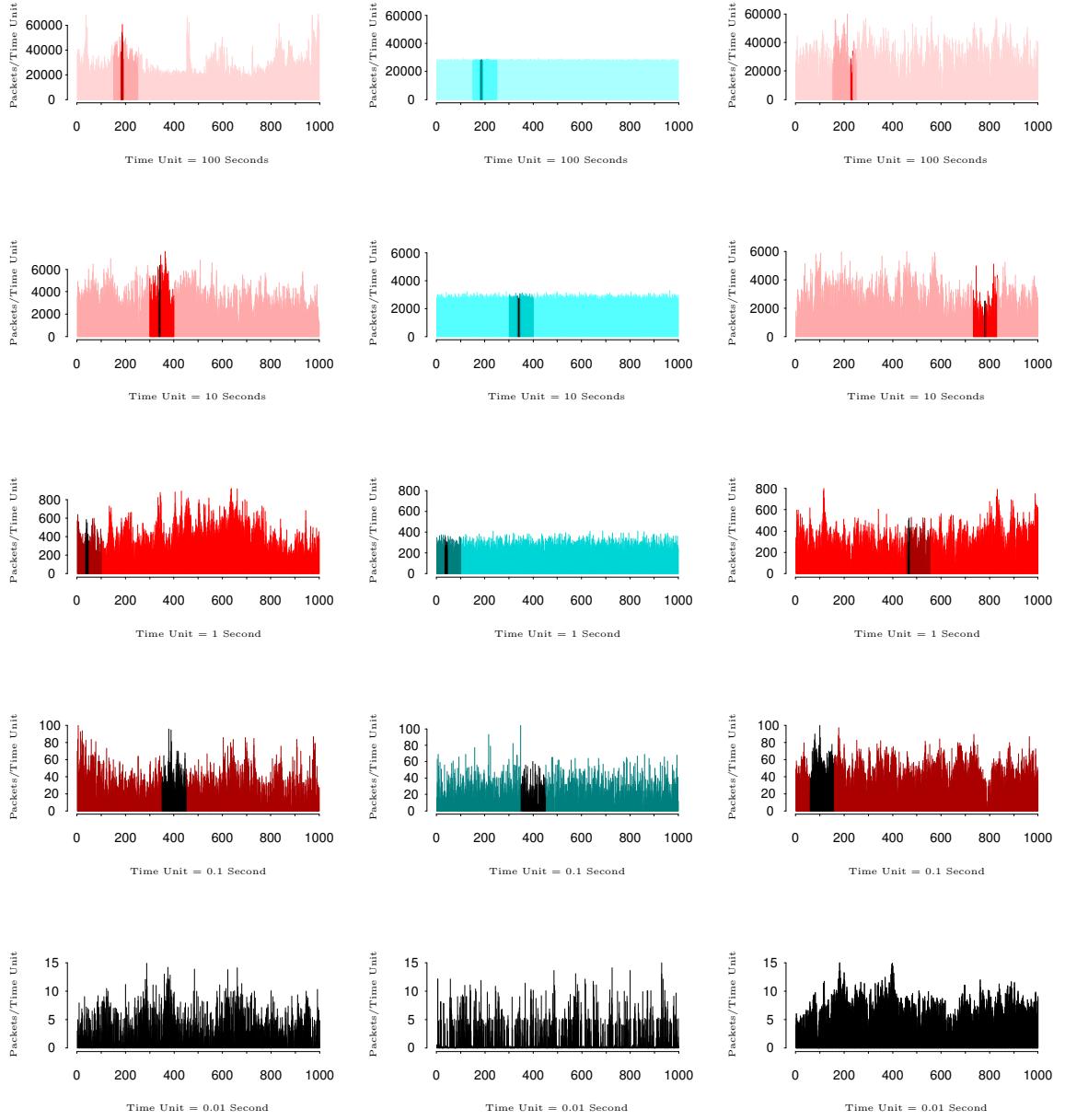


Figure 1.1: Actual Ethernet traffic (left column), a trace generated from a traditional traffic model (middle column), and a trace generated from a self-similar model (right column) on different time scales.

Apart from the upper left plot, which suggests the presence of a daily cycle (the LAN is observed for 27 hours), the plots in the left column look ‘similar’ in distributional sense: it is not possible to determine what the time unit is used by just looking at the shape of the plotted bins.

However, for a conventional model as in the middle column of Figure 1.1 (a batch Poisson model was simulated to generate the traffic), the number of packets per bin has a very low standard deviation at the coarsest time scale: the plot looks like a solid bar. By increasing the resolution, more ‘randomness’ shows up. This is caused by the fact that (independent) interarrival and service times have finite variance in traditional queueing models, e.g., exponential or Erlang distributions. The amount of traffic in disjoint time intervals is then mutual independent if the size of the intervals is large enough. We will make the limiting behavior more precise for an (ordinary) Poisson arrival process in Section 1.3. Note that the plots in the middle column look significantly different from those in the left column.

On the contrary, the plots in the right column (generated by a self-similar model) are visually almost indistinguishable from those in the left column. The only clear difference is that in the upper plot no cyclical behavior is present. However, when desirable, this daily behavior can be incorporated by adding some deterministic or stochastic trend.

We cannot just rely on one specific network trace of 27 hours to state that all LAN network traffic is self-similar. Therefore, Willinger *et al.* also study traces that are observed in other months and in a different network layout, making their conclusions more robust.

In another paper, Willinger *et al.* [58] tried to explain the self-similar nature of Ethernet LAN traffic by studying the source level, i.e., looking at the activity of all possible source-destination pairs in the LAN. They assume that each such pair is either sending bytes (‘ON’) or inactive (‘OFF’), where the *ON*- and *OFF*-periods strictly alternate. It is assumed that the lengths of the *ON*- and *OFF*-periods are independent and identical distributed (i.i.d.) and that all *ON*- and *OFF*-periods are also independent from one another. The authors find that the so-called ‘Noah Effect’ (high variability) can explain in some sense the self-similar phenomena that are observed on the aggregate level. The Noah effect is synonymous with infinite variance; it was empirically observed that many naturally occurring phenomena can well be described by distributions with infinite variance. The authors study the *cumulative* amount of packets (or bytes) when many sources are present. When this Noah effect is present in the *ON*- or *OFF*-periods, the cumulative amount of packets (after rescaling time) is described by a simple transformation of fractional Brownian motion.

The same self-similar behavior as in LAN Ethernet traffic is also discovered in wide area networks (WANs) [40], variable bit rate (VBR) video [10], etc. Many work is going on to show its presence in different kinds of data traffic.

1.1.2 Self-similarity and long-range dependence

We will now make mathematically precise what we mean by self-similarity and long-range dependence. It will be no surprise that these two concepts are related to fractional Brownian motion, but we will first define them in the framework of general stationary stochastic processes.

Let $X = \{X_k : k = 0, 1, 2, \dots\}$ be a stationary discrete-time stochastic process, meaning that the vectors $(X_{k_1}, \dots, X_{k_d})$ and $(X_{k_1+n}, \dots, X_{k_d+n})$ have the same distribution for all integers $d, n \geq 1$ and $k_1, \dots, k_d \geq 0$. For Gaussian processes, it is equivalent to require that $\gamma(k) := \text{Cov}(X_n, X_{n+k})$ does not depend on n . These two notions are sometimes referred to as strict stationarity and second-order stationarity, respectively. The function $\gamma(\cdot)$ is called the autocovariance function.

In Figure 1.1, we saw that summing up the amount of packets in the first 10 bins gives (up to scale) the same plot in distributional sense. This is also valid for the next 10 bins and so on. Suppose that this reasoning applies to a general factor $m \geq 1$, not necessarily a power of 10. Mathematically, we then have for every $k \geq 0$,

$$X_{km} + \dots + X_{(k+1)m-1} = a_m X_k, \quad (1.1)$$

where the equality is in the sense of equality in distribution. Although the scaling factor $a_m > 0$ is still not defined by this intuitive reasoning, we first define a new process $X^{(m)} = \{X_k^{(m)} : k = 0, 1, 2, \dots\}$ for every $m \geq 1$:

$$X_k^{(m)} = \frac{1}{m} (X_{km} + \dots + X_{(k+1)m-1}).$$

Following Cox [16], we call a discrete-time stochastic process X *self-similar* with Hurst parameter $0 < H < 1$ if X and $m^{1-H} X^{(m)}$ have the same finite-dimensional distributions for all $m \geq 1$. This means that for every $d \geq 1$ and $0 \leq k_1 < \dots < k_d$ the vector

$$(X_{k_1}, \dots, X_{k_d})$$

has the same distribution as the vector

$$(m^{1-H} X_{k_1}^{(m)}, \dots, m^{1-H} X_{k_d}^{(m)}),$$

implying that the correlation function $r(\cdot) = \gamma(\cdot)/\text{Var}(X_1)$ of X equals de correlation function $r^{(m)}(\cdot)$ of $X^{(m)}$ for all m . Thus, (1.1) holds with $a_m = m^H$ if X is self-similar with Hurst parameter H .

The above definitions can easily be extended to continuous-time stochastic processes. A continuous-time stochastic process $Y = \{Y(t) : 0 \leq t < \infty\}$ is called self-similar with Hurst parameter $0 < H < 1$ if $\{Y(at) : 0 \leq t < \infty\}$ and $\{a^H Y(t) : 0 \leq t < \infty\}$ have identical finite-dimensional distributions for all $a > 0$.

Weaker notions than self-similarity also exist. A process is *second-order self-similar* if the finite dimensional distributions of X and $m^{1-H} X^{(m)}$ have equal mean and covariance structure. If this is only true as $m \rightarrow \infty$, the process is called *asymptotically second-order self-similar*, see [16].

Next we introduce the definition of long-range dependence. A stationary discrete-time processes X is said to be a process with *long-range dependence*, long memory, or strong dependence when its autocovariance function $\gamma(\cdot)$ decays so slowly that $\sum_{k=0}^{\infty} \gamma(k) = \infty$ (in contrast to processes with summable covariances, which are called processes with short-range dependence, short memory, or weak dependence). Intuitively, when long-range dependence is present, high-lag correlations may be individually small, but their cumulative effect is significant.

As pointed out by Beran [9], such a covariance structure has an important impact on usual statistical inference. As an example, assume we have n observations of some random variable X with finite variance. The standard deviation of the mean is then proportional to $n^{1/2}$ if the observations are uncorrelated. If the covariances decay exponentially (as is the case with a so-called AR(1) process), the covariances are summable and similar behavior is observed: the standard deviation of the mean is proportional to $n^{1/2}$ for sufficiently large n , although the proportionality constant is different. Suppose, on the other hand, that the covariance function $\gamma(\cdot)$ decays hyperbolically:

$$\gamma(k) \sim c_\gamma |k|^{-\alpha} \quad (1.2)$$

as $|k|$ tends to infinity for some $0 < \alpha < 1$ and a finite positive constant c_γ . [In this report, we follow the notational convention that $f(x) \sim g(x)$ as $x \rightarrow X \in [-\infty, \infty]$ stands for $\lim_{x \rightarrow X} f(x)/g(x) = 1$.] It is readily checked that the process is long-range dependent if (1.2) holds. Under this long-range dependence, the standard deviation of the mean is proportional to $n^{-\alpha/2}$! Of course, this affects confidence intervals for the mean of X and all related test statistics. Moreover, the standard estimator for the variance of X becomes biased. This bias does not disappear when the sample size increases, as is the case for short-range dependent processes.

It is important to note that Equation (1.2) determines only the *decay* of the correlations. It is very well possible that there are some specific lags for which $\gamma(k)$ is particularly large, which makes the detecting of long-range dependence more difficult. In fact, it is theoretically impossible to conclude that long-range dependence is present in a finite sample.

In general, self-similarity and long-range dependence are not equivalent. As an example, the increments of a standard Brownian motion are self-similar with Hurst parameter $H = 1/2$, but clearly not long-range dependent (the increments are even independent). However, under the restriction $1/2 < H < 1$, long-range dependence is equivalent to asymptotic second-order self-similarity for stationary processes.

Before investigating the influence of fractional Brownian motion in a queueing framework, we will first define this process and derive its basic properties.

1.2 Fractional Brownian motion and fractional Gaussian noise

Since we will shortly see that fractional Brownian motion is a Gaussian process, its covariance structure is one of its most important features. Therefore, the key subjects of this section are the covariance structure of fractional Brownian motion and its incremental process, fractional Gaussian noise. Moreover, the properties of fractional Gaussian noise in the so-called frequency domain are described. Finally, another long-range dependent process is presented for later use.

1.2.1 Definitions and basic properties

Fractional Brownian motion

In the pioneering work by Mandelbrot and van Ness [42], fractional Brownian motion is defined by its *stochastic representation*

$$B_H(t) := \frac{1}{\Gamma(H + \frac{1}{2})} \left(\int_{-\infty}^0 [(t-s)^{H-1/2} - (-s)^{H-1/2}] dB(s) + \int_0^t (t-s)^{H-1/2} dB(s) \right), \quad (1.3)$$

where Γ represents the Gamma function $\Gamma(\alpha) := \int_0^\infty x^{\alpha-1} \exp(-x) dx$ and $0 < H < 1$ is called the Hurst parameter (we soon see the connection with the Hurst parameter for self-similar processes). The integrator B is a stochastic process, ordinary Brownian motion. Note that B is recovered by taking $H = 1/2$ in (1.3). For later use, we assume that B is defined on some probability space (Ω, \mathcal{F}, P) . We remark that this representation in terms of an integral with respect to Brownian motion is non-unique; see, e.g., p. 140 of [52] for a different representation. One point should be clarified in more detail: how to interpret an integral with random integrator B .

The notation suggests that this integral can be seen as a Lebesgue-Stieltjes integral. One could therefore think that the integral can pathwise be computed in a Lebesgue-Stieltjes sense (recall that B can be regarded as a random function). However, this is impossible, since the paths of Brownian motion are highly irregular. For example, though the trajectories are continuous, they are almost surely non-differentiable and, more importantly in this case, the paths do not have bounded variation with probability 1. This property prevents us from calculating the integral as a pathwise Lebesgue-Stieltjes integral.

Having observed this, we have still not solved the interpretation problem. In fact, the integral is a so-called stochastic integral with respect to usual Brownian motion. We will briefly explain its definition, but we will omit many details. For details on stochastic integration, the reader is referred to one of the many textbooks on stochastic calculus, e.g., Karatzas and Shreve [37].

Avoiding technicalities caused by an unbounded integration interval, we will only discuss the definition of the stochastic integral

$$\int_a^b \phi(s) dB(s)$$

for finite a and b . This integral has a natural definition when the integrand ϕ is a so-called simple function, which means that there exists an integer $\ell > 0$ and a strictly increasing sequence of real numbers $(t_j)_{j=0}^\ell$ with $t_0 = a$ and $t_\ell = b$, as well as a sequence of real numbers $(\phi_j)_{j=0}^{\ell-1}$ such that $\phi(s)$ can be written as $\phi(s) = \phi_j$ for $s \in (t_j, t_{j+1}]$ (the value at a turns out to have no influence). For such a simple function ϕ , the stochastic integral has the following natural definition:

$$\int_a^b \phi(s) dB(s) = \sum_{j=0}^{\ell-1} \phi_j (B(t_{j+1}) - B(t_j)). \quad (1.4)$$

A sequence of (Lebesgue) square integrable functions (ψ_n) is said to converge in L^2 -norm to a square integrable function ψ if

$$\lim_{n \rightarrow \infty} \int (\psi(s) - \psi_n(s))^2 ds = 0.$$

The integration interval is again $[a, b]$, which is sometimes made explicit by calling this convergence in $L^2([a, b])$ -norm. It is known that every square integrable function can be written as a limit in L^2 -norm of a sequence of simple functions. Assume therefore that ψ_n is a simple function for every n . A possible definition for $\int \psi(s) dB(s)$ would be $\lim_{n \rightarrow \infty} \int \psi_n(s) dB(s)$. Unfortunately, this limit only exists almost surely when ψ satisfies very restrictive conditions.

Since $\int_a^b \psi_n(s) dB(s)$ has a finite second moment, it is known that there exists a square integrable random variable Z on our probability space (Ω, \mathcal{F}, P) with the property

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[\left(Z - \int_a^b \psi_n(s) dB(s) \right)^2 \right] = 0, \quad (1.5)$$

where \mathbb{E} denotes the expectation operator with respect to P . This random variable Z is sometimes referred to as the \mathcal{L}^2 limit of approximating stochastic integrals of simple functions, and the type of convergence is called convergence in \mathcal{L}^2 -norm (note the subtle difference in notation with the previous L^2 -norm). If ψ is square integrable, this random variable is unique in the sense that another random variable that satisfies (1.5)

is almost sure equal to Z . This makes it possible to take this Z as the definition of the stochastic integral $\int \psi(s)dB(s)$. Moreover, it can be shown that Z is independent of the approximating sequence $(\psi_n)_n$, implying that if a given sequence of simple functions $(\psi_n)_n$ converges in L^2 -norm to ψ , the corresponding sequence of stochastic integrals $(\int \psi_n(s)dB(s))_n$ converges in L^2 -norm to $\int \psi(s)dB(s)$. We will use this fact in Chapter 2.

Note that $\int \psi(s)dB(s)$ is only defined for square integrable ψ by this reasoning. The integral Z is written as $\int \psi(s)dB(s)$, but it has nothing to do with other types of integrals with the same notation, although some of the usual properties of integrals are shared, e.g., linearity. Also note that some specific properties of Brownian motion are used in the way we defined the stochastic integral. Extensions to other processes are possible using the same arguments, although the definition of the stochastic integral is still a topic of research for interesting processes that do not fit in this framework.

Omitting details, the second moments of the stochastic integral in (1.3) can be computed by a standard formula, which leads to the fact that the variance of $B_H(t)$ is $V_H t^{2H}$ for some constant V_H . From now on, we will always assume that we deal with normalized (or standardized) fractional Brownian motion, which has *exactly* variance t^{2H} . We will use the same notation for the normalized process as before.

A normalized fractional Brownian motion $B_H = \{B_H(t) : 0 \leq t < \infty\}$ with $0 < H < 1$ is uniquely characterized by the following properties, cf., e.g., [45]:

- $B_H(t)$ has stationary increments;
- $B_H(0) = 0$, and $\mathbb{E}B_H(t) = 0$ for $t \geq 0$;
- $\mathbb{E}B_H^2(t) = t^{2H}$ for $t \geq 0$;
- $B_H(t)$ has a Gaussian distribution for $t > 0$.

We always suppose that we deal with a version of B_H with continuous sample paths (such a version exists by the Kolmogorov criterion for continuity of sample paths). From the first three properties it follows that the covariance function is given by

$$\rho(s, t) = \mathbb{E}B_H(s)B_H(t) = \frac{1}{2} \{t^{2H} + s^{2H} - (t-s)^{2H}\} \quad (1.6)$$

for $0 < s \leq t$. For Gaussian processes, the mean and covariance structure determine the finite-dimensional distributions uniquely. Therefore, we conclude from (1.6) that $\{B_H(at) : 0 \leq t < \infty\}$ and $\{a^H B_H(t) : 0 \leq t < \infty\}$ have the same finite-dimensional distributions: fractional Brownian motion with Hurst parameter H is self-similar with Hurst parameter H . In fact, fractional Brownian motion is the only Gaussian process with stationary increments that is self-similar [16].

Besides communications engineering, fractional Brownian motion has applications in other areas, such as finance, physics, and bioengineering, to name but a few. In bioengineering for instance, fractional Brownian motion is used to model regional flow distributions in the heart, the lung and the kidney [7].

Figure 1.2 consists of simulated sample paths for three different values of H . The differences between the graphs are more easily explained after having studied the incremental process, fractional Gaussian noise.

Fractional Gaussian noise

Define the incremental process $X = \{X_k : k = 0, 1, \dots\}$ of fractional Brownian motion, which is called fractional Gaussian noise, by

$$X_k = B_H(k+1) - B_H(k).$$

It is clear that X_k has a standard normal distribution for every k , but that there is (in general) no independence. To be precise, the corresponding autocovariance function $\gamma(\cdot)$ is of the form

$$\gamma(k) = \frac{1}{2} [|k-1|^{2H} - 2|k|^{2H} + |k+1|^{2H}] \quad (1.7)$$

for $k \in \mathbb{Z}$. If $H = 1/2$, all the covariances are 0 (except, of course, for $k = 0$). Since fractional Gaussian noise is a Gaussian process, this implies independence. This agrees with the properties of ordinary Brownian motion, which has independent increments.

By writing down the Taylor expansion at the origin of the function $h(x) = (1-x)^{2H} - 2 + (1+x)^{2H}$ and noting that $\gamma(k) = \frac{1}{2}k^{2H}h(1/k)$ for $k \geq 1$, it is easily seen from (1.7) that

$$\gamma(k) \sim H(2H-1)k^{2H-2} \quad (1.8)$$

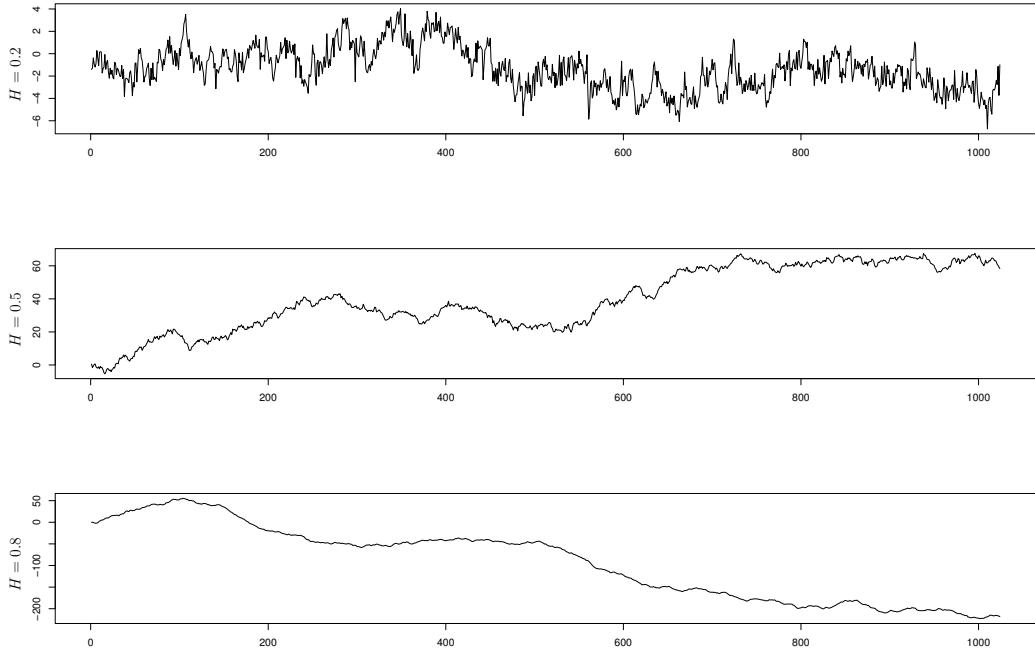


Figure 1.2: Samples of fractional Brownian motion for $H = 0.2$, $H = 0.5$ and $H = 0.8$.

as $k \rightarrow \infty$. This implies long-range dependence for $1/2 < H < 1$: $\sum_{k=0}^{\infty} \gamma(k) = \infty$. It is important to know that the measurements on real traffic traces indeed suggest that $1/2 < H < 1$, see [57].

Since we have noted that long-range dependence is closely related to self-similarity for $1/2 < H < 1$, it is natural to study whether fractional Gaussian noise is also a self-similar process. To make sure that this is indeed the case, it suffices to show that $mX^{(m)}$ and $m^H X$ have the same finite-dimensional distributions. Since we have already seen that fractional Brownian motion is self-similar, the following holds:

$$\begin{aligned} & \text{Cov}(X_{km} + \dots + X_{(k+1)m-1}, X_{lm} + \dots + X_{(l+1)m-1}) \\ &= \text{Cov}(B_H((k+1)m) - B_H(km), B_H((l+1)m) - B_H(lm)) \\ &\stackrel{*}{=} \text{Cov}(m^H(B_H(k+1) - B_H(k)), m^H(B_H(l+1) - B_H(l))) \\ &= \text{Cov}(m^H X_k, m^H X_l). \end{aligned}$$

At the equality $\stackrel{*}{=}$, we used the fact that the vectors $(B_H(mk), B_H(m(k+1)), B_H(ml), B_H(m(l+1)))$ and $(m^H B_H(k), m^H B_H(k+1), m^H B_H(l), m^H B_H(l+1))$ have the same distribution for $k \neq l$, because all finite-dimensional (particularly four-dimensional) distributions are equal (since fractional Brownian motion is self-similar). A similar argument shows that this is also the case for $k = l$. Because of the normality, it suffices to have equal means and covariances to conclude that the finite-dimensional distributions of $mX^{(m)}$ and $m^H X$ are equal: fractional Gaussian noise is self-similar.

It can be checked from (1.7) that the covariances are negative for $H < 1/2$ and positive for $H > 1/2$. This behavior is also recognized in Figure 1.3, in which samples of fractional Gaussian noise are depicted for the same values of the Hurst parameter H as in Figure 1.2. For $H = 0.2$, the negative correlation accounts for the high variability, whereas for $H = 0.8$ there are clearly periods in which the sample path increases and periods in which it decreases. Although the range on the vertical axis is not equal for every plot, it is plausible that the marginal distributions are independent of H (in fact, the standard deviation is 1).

We now return to Figure 1.2, in which the cumulative sums of the same samples are plotted. The negative correlations for $H = 0.2$ are also observed in this plot, whereas the sample is more smooth for $H = 0.8$ due to the positive correlations.

1.2.2 Spectral densities

Instead of analyzing a stochastic process in the time domain as in the previous subsection, processes can also be analyzed in the so-called frequency or spectral domain. In this subsection, we only focus on stationary processes. This technique is called spectral analysis, which has many applications, e.g., in physics and time series analysis. A time-domain series can be transformed into a frequency-domain series without loss of information by the so-called Fourier transform. This means that the time-domain series is perfectly recovered from the frequency-domain series by the so-called inverse Fourier transform. We will see in Chapter 2 how

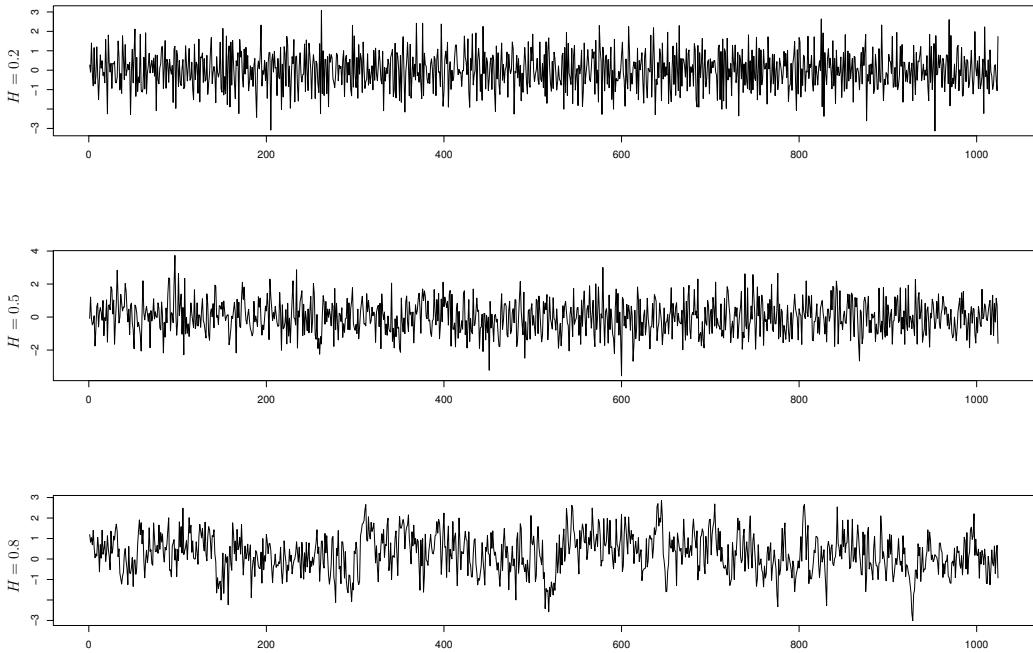


Figure 1.3: Samples of fractional Gaussian noise for $H = 0.2$, $H = 0.5$ and $H = 0.8$.

‘approximate’ fractional Brownian motion can be simulated by generating a frequency-domain sample and then transforming the resulting series. The required theoretical background is described in this subsection.

Fourier proved that a (deterministic) periodic function can be written as a unique linear combination of trigonometric functions with different frequencies, making it thus possible to describe this function completely by the amount in which each frequency is present. It is possible and useful to extend this idea to non-periodic functions; a deterministic function (or a realization of a stochastic process) can be thought to consist of trigonometric functions with different frequencies. The information to which extend each frequency is present in the signal is then summarized in the so-called *spectral density*, also called power spectrum density because of its interpretation in physics.

When analyzing stochastic processes in a frequency framework, it is impossible to study every realization individually. However, it turns out (see, e.g., [51]) that for stationary processes the *expected* frequency information is contained in the autocovariance function.

For stationary stochastic processes, the spectral density is computed as follows for frequencies $-\pi \leq \lambda \leq \pi$:

$$f(\lambda) = \sum_{j=-\infty}^{\infty} \gamma(j) \exp(ij\lambda), \quad (1.9)$$

where $\gamma(\cdot)$ again denotes the autocovariance function. The autocovariance function is recovered by the inversion formula

$$\gamma(j) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\lambda) \exp(-ij\lambda) d\lambda. \quad (1.10)$$

In this report, we are particularly interested in the spectral density of fractional Gaussian noise. It can be seen [9] that this density is given by

$$f(\lambda) = 2 \sin(\pi H) \Gamma(2H+1) (1 - \cos \lambda) [|\lambda|^{-2H-1} + B(\lambda, H)], \quad (1.11)$$

with

$$B(\lambda, H) = \sum_{j=1}^{\infty} \{(2\pi j + \lambda)^{-2H-1} + (2\pi j - \lambda)^{-2H-1}\}$$

for $-\pi \leq \lambda \leq \pi$. Since very high frequencies correspond to ‘waves’ *between* the sample points $\{0, 1, \dots\}$, it is only useful to define the density for a range of frequencies of length 2π (see [51] for details).

Unfortunately, no closed form of the spectral density for fractional Gaussian noise is known, but a quite useful result is the proportionality of this spectral density to $|\lambda|^{1-2H}$ near $\lambda = 0$, checked by noting that $1 - \cos(\lambda) = \frac{1}{2}\lambda^2 + O(\lambda^4)$ as $|\lambda| \rightarrow 0$. Therefore, f has a pole at zero for $H > 1/2$. For $H < 1/2$, f is non-differentiable at zero. The spectral density is plotted for several values of $H > 1/2$ in Figure 1.4. Note that a pole of the spectral density at zero is equivalent to long-range dependence (see Equation (1.9) and Section 1.1.2).

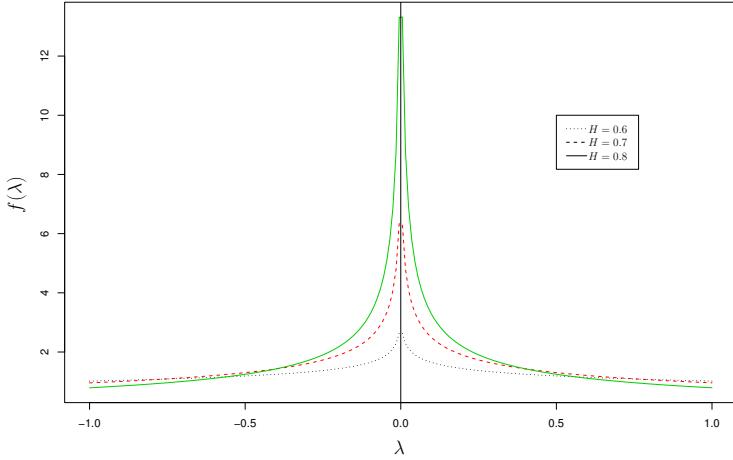


Figure 1.4: The spectral density for $H = 0.6$, $H = 0.7$ and $H = 0.8$.

A low frequency (i.e., small λ) corresponds to a long wavelength. Intuitively, a frequency λ has more impact when $f(\lambda)$ is larger. Thus, the larger H , the larger the low frequency component in the spectral density, as seen in Figure 1.4. We recognize this in Figure 1.3, where bigger ‘waves’ are present in the plot for $H = 0.8$.

To evaluate the spectral density numerically, the infinite sum in Equation (1.11) must be truncated. When the truncation parameter is chosen quite large, the function evaluation becomes computationally very demanding. Paxson [47] suggests and tests a useful approximation to overcome this problem. The key observation is that

$$\sum_{j=1}^{\infty} h(j) \approx \sum_{j=1}^k h(j) + \frac{1}{2} \int_k^{k+1} h(x) dx + \int_{k+1}^{\infty} h(x) dx,$$

for every increasing integrable function h and $k = 0, 1, \dots$. With ‘ \approx ’ is meant that the right hand side is a reasonable approximation for the left hand side. It is shown [47] that this is already a good approximation for $k = 3$. In that case, the approximation becomes

$$\tilde{B}_3(\lambda, H) = \sum_{j=1}^3 \{(a_j^+)^{-2H-1} + (a_j^-)^{-2H-1}\} + \frac{(a_3^+)^{-2H} + (a_3^-)^{-2H} + (a_4^+)^{-2H} + (a_4^-)^{-2H}}{8H\pi}, \quad (1.12)$$

where $a_j^{\pm} = 2\pi j \pm \lambda$. The author analyzes the relative error of this approximation, and then corrects the mean absolute error and linear variation in λ by fitting. As final approximation for $B(\lambda, H)$, he arrives at $\tilde{B}(\lambda, H)'$, where

$$\tilde{B}(\lambda, H)' = (1.0002 - 0.000134\lambda) \left\{ \tilde{B}_3(\lambda, H) - 2^{-7.65H-7.4} \right\}.$$

Some problems arise when computing a spectral density for fractional Brownian motion, being a non-stationary process: there exists no autocovariance function and Equation (1.9) is thus not applicable. Clearly, a time-independent density (as in the fractional Gaussian noise case) is unsuitable. We will not use the spectral density of fractional Brownian motion in this report, but the interested reader is referred to Flandrin [28] for a discussion.

1.2.3 Another long-range dependent process

Another widely used process with long-range dependence is fractional ARIMA. Since we will encounter this process only sideways, we discuss it briefly. The parameters of this model control the long-range dependence as well as the short term behavior. The fractional ARIMA model is based on the ARMA model.

An ARMA(p, q) process $W = \{W_k : k = 0, 1, \dots\}$ is a short memory process that is the solution of

$$\phi(L)W_k = \theta(L)\epsilon_k,$$

where ϕ and θ are polynomials of order p and q respectively and ϵ is a white noise process, i.e., the ϵ_k are i.i.d. standard normal random variables. The lag operator L is defined as $LW_k = W_{k-1}$.

A generalization of this model is the ARIMA(p, d, q) process for $d = 0, 1, \dots$, defined by the property that $(1 - L)^d W_k$ is an ARMA(p, q) process. As implied by its name, the fractional ARIMA model admits a

fractional value for the parameter d . For this, we have to know how $(1 - L)^d W_k$ is defined for fractional d . This is formally done by the binomial expansion:

$$(1 - L)^d W_k = \sum_{n=0}^{\infty} \binom{d}{n} (-L)^n W_k,$$

where the binomial coefficient is defined as

$$\binom{d}{k} = \frac{\Gamma(d+1)}{\Gamma(d-k+1)\Gamma(k+1)}.$$

Since the case $d > 1/2$ can be reduced to the case $-1/2 < d \leq 1/2$ by taking appropriate differences, the latter case is particularly interesting. W is a stationary process for $-1/2 < d < 1/2$. Long-range dependence occurs for $0 < d < 1/2$, implying that the process is also asymptotically second order self-similar in this case. The corresponding Hurst parameter is $H = 1/2 + d$.

For simulation issues, we refer to [32].

1.3 Implications for network models

Now that it is clear that fractional Brownian motion is a self-similar process with a long-range dependent incremental process (under the restriction $H > 1/2$), we base a network traffic model on fractional Brownian motion.

Norros [45] defines *fractional Brownian traffic* as

$$A(t) = Mt + \sqrt{aM}B_H(t). \quad (1.13)$$

Here, $A(t)$ represents the number of bits (or data packets) that is offered in the time interval $[0, t]$. This process has three parameters M , a and H . $M > 0$ is interpreted as the mean input rate, $a > 0$ as a variance coefficient, and $1/2 \leq H < 1$ as the Hurst parameter of de fractional Brownian motion B_H . Note that ordinary Brownian traffic (i.e., fractional Brownian traffic with $H = 1/2$) is a special case, but that no long-range dependence is present in the incremental process for $H = 1/2$.

It may seem a bit surprising that the square root of the mean input rate \sqrt{M} is present in (1.13) as a scaling factor for the fractional Brownian motion $B_H(t)$. Norros motivates this by a superposition property: the sum $A(t) = \sum_{i=1}^K A^{(i)}(t)$ of K independent fractional Brownian traffics with common parameters a and H but individual mean rates M_i can then be written as $A(t) = \sum_{i=1}^K M_i + \sqrt{a \sum_{i=1}^K M_i} B_H(t)$, where B_H is again a fractional Brownian motion with parameter H .

An obvious weakness of this traffic model is that a trajectory of the process may *decrease*. Since $A(t)$ represents the cumulative ‘traffic’ in $[0, t]$, the amount of traffic arriving in an interval $[t, t+T]$ is $A(t+T) - A(t)$, which can be negative. However, this problem is considered of secondary importance. The probability of decay in an interval $[t, t+T]$ is time-independent and may be quite small, depending on the choice of a and M . To make this more precise, we compute this probability explicitly using the stationarity of the increments of fractional Brownian motion and the self-similarity:

$$\begin{aligned} P(A(t+T) - A(t) < 0) &= P(\sqrt{aM}(B_H(t+T) - B_H(t)) + MT < 0) \\ &= P(\sqrt{aM}B_H(T) + MT < 0) \\ &= P(B_H(T) < -\sqrt{M/a}T) \\ &= P(B_H(1) < -\sqrt{M/a}T^{1-H}) \\ &= \Phi(-\sqrt{M/a}T^{1-H}), \end{aligned}$$

where Φ denotes the cumulative distribution function of the standard normal distribution. It is now clear that this probability rapidly decreases in the interval length T and the factor M/a . This means that negative traffic is no big issue as long as the timescale (sampling interval) and mean arrival rate (relative to the variance coefficient) are not too small.

Besides the possibility of negative traffic, the number of bits arriving in a time interval may not be integer-valued. This is also no big problem as long as the number of arrivals in each interval is not too small; potential fractional values are small compared to this number, which implies that only a small ‘error’ is made. Again, this is a result of the choice of the interval length, the variance coefficient a and arrival rate per time unit M .

Norros [44] studies a storage model with fractional Brownian traffic input. The work that is generated by the traffic process is offered to a server with a fixed capacity of C bits (or data packets) in one time unit. The work in the system $V(t)$ at time t is called *fractional Brownian storage* and is defined as

$$V(t) = \sup_{s \leq t} (A(t) - A(s) - C(t-s))$$

for $t \geq 0$. Note that this so-called workload process is non-negative, and negative traffic (i.e., a decreasing A) causes no problems for defining the workload process. The following logarithmic asymptotics for the tail behavior of the distribution of $V(t)$ are obtained:

$$\log P(V(t) > v) \sim -\frac{(C-M)^{2H}}{2H^{2H}(1-H)^{2-2H}aM} v^{2-2H} \quad (1.14)$$

as $v \rightarrow \infty$. Thus, the (tail) distribution can be approximated by a Weibull-type distribution. For exact asymptotics of $P(V(0) > v)$, the reader is referred to [35]; see also [50].

For the Brownian case $H = 1/2$, $a = 1$ and when $C = 1$, (1.14) reduces to an exponential tail behavior: $\frac{1}{v} \log P(V(t) > v) \sim -2(1-M)/M$. As pointed out by the author [44], this coincides with the heavy traffic asymptotics of the $M/D/1$ system.

This may be illustrated by the fact that fractional Brownian traffic (1.13) with $H = 1/2$ is (in some sense) the limit of a Poisson process. We will make this precise by studying a Poisson process $P = \{P(t) : t \geq 0\}$ with parameter M . For such a Poisson process,

$$Q^{(\alpha)}(t) = \frac{1}{\sqrt{\alpha M}} [P(\alpha t) - M\alpha t]$$

converges in distribution to a Gaussian variable with variance t as $\alpha \rightarrow \infty$. This is a result of the Central Limit Theorem, the fact that $P(\alpha t) = \sum_{j=1}^{\alpha} [P(jt) - P((j-1)t)]$ and the independence of Poisson process increments. We choose $d \geq 1$ time epochs $0 \leq t_1 < \dots < t_d$ and set $t_0 = 0$ for notational convenience. Using the stationarity of the Poisson process, we define $Q_j^{(\alpha)} = Q^{(\alpha)}(t_j) - Q^{(\alpha)}(t_{j-1})$ and deduce that $Q_j^{(\alpha)}$ converges in distribution to a Gaussian random variable with variance $t_j - t_{j-1}$ for $j = 1, \dots, d$. In addition, $(Q_1^{(\alpha)}, \dots, Q_d^{(\alpha)})$ has independent elements for fixed α , again by the independence of Poisson process increments. This independence is asymptotically retained, as can most easily be seen by considering cumulative distribution functions. For every $x_1, \dots, x_d \in \mathbb{R}$ it holds that

$$\begin{aligned} F_{(Q_1^{(\alpha)}, \dots, Q_d^{(\alpha)})}(x_1, \dots, x_d) &= F_{Q_1^{(\alpha)}}(x_1) \cdots F_{Q_d^{(\alpha)}}(x_d) \\ &\rightarrow \Phi_{t_1-t_0}(x_1) \cdots \Phi_{t_d-t_{d-1}}(x_d) = \Phi_{\text{BM}}(x_1, \dots, x_d), \end{aligned} \quad (1.15)$$

where $\Phi_{\sigma^2}(\cdot)$ denotes the cumulative distribution function of a Gaussian variable with variance σ^2 . The last expression in (1.15) is the cumulative distribution function of the increments of Brownian motion. We conclude that $(P(\alpha t) - \alpha M t)/\sqrt{\alpha M}$ converges in finite-dimensional distribution to ordinary Brownian motion. This suggests the approximation $A(t) \approx Mt + \sqrt{M}B(t)$, which is fractional Brownian traffic with parameters $H = 1/2$ and $a = 1$.

Obviously, different techniques (and underlying arrival processes) are required to view fractional Brownian motion with Hurst parameter $H \neq 1/2$ as a limiting process. We will briefly come back to this in Chapter 2.

The example of a simple queueing system with fractional Brownian traffic input is very important, because completely different behavior of quantities is observed when a long-range dependent arrival process is modeled instead of a traditional (e.g., Poisson) process. Much work is done nowadays to address this issue for different models.

For more complicated and realistic systems, simulation becomes an attractive approach. Especially when investigating the tail behavior of random variables (e.g., the number of packets in a buffer), very small probabilities have to be simulated. For this, importance sampling or quick simulation (often based on a large deviation result) may be useful. Fast simulation for long-range dependent processes is a relatively new area, but some work has been done by Huang *et al.* [33].

1.4 Outline and scientific contribution

As pointed out before, the evaluation of several simulation methods for fractional Brownian motion is the main issue of this report. However, it is often more convenient to work with the (stationary) incremental

process of fractional Brownian motion, fractional Gaussian noise (see Section 1.2). Some algorithms to simulate fractional Brownian motion (or fractional Gaussian noise) are described in Chapter 2. The first part addresses the available exact methods (i.e., the output of the method is a sampled realization of fractional Brownian motion). Several approximate methods are presented in the second part. A number of methods based on the so-called Fast Fourier Transformation (FFT) are included in this survey. By the uniform presentation, some insight is gained into the connections between these methods. The Paxson method, which was originally proposed by Paxson [47] using purely intuitive and asymptotic arguments, thus lacking a real theoretical foundation, is shown to fit in a general theoretical framework. This finding does not only clarify many empirical observations, it also leads to an important improvement of the method. In addition, it is shown that this FFT-based Paxson method and a generalization of the method, the approximate circulant method, are asymptotically exact in the sense that the variance of the approximation error decays to zero for every sample point as the sample size increases. It also turns out that this approximate circulant method is closely related to an exact method, the Davies and Harte method. From this observation is benefited to study the performance of this method.

To evaluate approximate simulation methods, we should know if generated approximate samples have the same properties as exact samples. Statistical estimation and testing techniques that can be used to investigate this are discussed in Chapter 3. Although most of the estimation techniques are standard tools in the analysis of long-range dependent processes, a number of recent estimation methods, based on discrete variations and wavelets, are also reviewed. In addition, a test is proposed that is perfectly suitable for our evaluation purposes. However, the test is time- and memory-consuming and its power is unknown.

The final chapter, Chapter 4, consists of the actual analysis of the most promising approximate simulation methods. First, an error analysis is performed using the insights of Chapter 2. When possible, the covariance structure of approximate samples is compared to the exact covariance structure. For the so-called Conditionalized Random Midpoint Displacement method, it is described how the covariances of the sample can numerically be computed. In the original paper, this was only done for the two simplest cases.

Then, the approximate methods are evaluated by an approach that is different from the approaches taken in previous studies. The accuracy of new simulation methods is mostly studied by estimating the Hurst parameter using a number of estimation methods. Note that the Hurst parameter determines the covariance structure completely, see (1.6). However, these estimation methods are in general based on only one property of an exact fractional Gaussian noise sample, thus limiting the conclusive power of the estimates. We can partially overcome this problem by estimating the Hurst parameter with several estimators. However, we still cannot be sure that the sample is fractional Gaussian noise when all available estimators produce ‘satisfactory’ estimates.

An additional problem is that most estimators are biased. We deal with this in Chapter 4 by using nonparametric inference techniques.

CHAPTER 2

Simulation methods

In the previous chapter, we have seen that the fractional Brownian motion, a Gaussian self-similar process, models some types of data traffic better and more realistically than conventional models. Since few theoretical results are known about queueing systems with a fractional Brownian motion-type arrival process, it is important to know how this process can be simulated. Therefore, an important role is played by simulation studies, especially for complex queueing systems. It is the aim of this chapter to describe some simulation methods for fractional Brownian motion, exact as well as approximate.

Since it is only possible to simulate in discrete time, we adapt a discrete time notation Y_0, Y_1, \dots for the values of the fractional Brownian motion at the discrete time epochs $0, 1, \dots$. Once this fractional Brownian motion sample is simulated, a realization on another equispaced grid is obtained by using the self-similarity property.

2.1 Exact methods

2.1.1 The Hosking method

The Hosking method (also known as the Durbin or Levinson method) is an algorithm to simulate a general stationary Gaussian process; therefore, we will focus on the simulation of fractional Gaussian noise X_0, X_1, \dots . Recall that a fractional Brownian motion sample Y_0, Y_1, \dots is obtained from a fractional Gaussian noise sample by taking cumulative sums. The method generates X_{n+1} given X_n, \dots, X_0 recursively. It does not use specific properties of fractional Brownian motion nor fractional Gaussian noise; the algorithm can be applied to any stationary Gaussian process. The key observation is that the distribution of X_{n+1} given the past can explicitly be computed.

Write $\gamma(\cdot)$ for the covariance function of the (zero-mean) process, that is

$$\gamma(k) := \mathbb{E}X_n X_{n+k},$$

for $n, k = 0, 1, 2, \dots$. Assume for convenience that $\gamma(0) = 1$. Furthermore, let $\Gamma(n) = (\gamma(i-j))_{i,j=0,\dots,n}$ be the covariance matrix and $c(n)$ be the $(n+1)$ -column vector with elements $c(n)_k = \gamma(k+1)$, $k = 0, \dots, n$. Define the $(n+1) \times (n+1)$ matrix $F(n) = (\mathbf{1}(i=n-j))_{i,j=0,\dots,n}$, where $\mathbf{1}$ denotes the indicator function. Note that premultiplying this matrix with a column vector or postmultiplying with a row vector ‘flips’ this vector.

The matrix $\Gamma(n+1)$ can be splitted as follows:

$$\Gamma(n+1) = \begin{pmatrix} 1 & c(n)' \\ c(n) & \Gamma(n) \end{pmatrix} \quad (2.1)$$

$$= \begin{pmatrix} \Gamma(n) & F(n)c(n) \\ c(n)'F(n) & 1 \end{pmatrix}, \quad (2.2)$$

where the prime denotes vector transpose.

We next compute the conditional distribution of X_{n+1} given X_n, \dots, X_0 . In fact, we show that this distribution is Gaussian with expectation μ_n and variance σ_n^2 given by

$$\mu_n := c(n)'\Gamma(n)^{-1} \begin{pmatrix} X_n \\ \vdots \\ X_1 \\ X_0 \end{pmatrix}, \quad \sigma_n^2 := 1 - c(n)'\Gamma(n)^{-1}c(n). \quad (2.3)$$

Define $d(n) = \Gamma(n)^{-1}c(n)$ for convenience. It is left to the reader to check with (2.1) and (2.2) that the following two expressions for the inverse of $\Gamma(n+1)$ hold:

$$\Gamma(n+1)^{-1} = \frac{1}{\sigma_n^2} \begin{pmatrix} 1 & -d(n)' \\ -d(n) & \sigma_n^2 \Gamma(n)^{-1} + d(n)d(n)' \end{pmatrix} \quad (2.4)$$

$$= \frac{1}{\sigma_n^2} \begin{pmatrix} \sigma_n^2 \Gamma(n)^{-1} + F(n)d(n)d(n)'F(n) & -F(n)d(n) \\ -d(n)'F(n) & 1 \end{pmatrix}, \quad (2.5)$$

From (2.4) is readily deduced that for each $x \in \mathbb{R}^{n+1}$ and $y \in \mathbb{R}$:

$$\begin{pmatrix} y & x' \end{pmatrix} \Gamma(n+1)^{-1} \begin{pmatrix} y \\ x \end{pmatrix} = \frac{(y - d(n)'x)^2}{\sigma_n^2} + x'\Gamma(n)^{-1}x. \quad (2.6)$$

This implies that the distribution of X_{n+1} given X_n, \dots, X_0 is indeed Gaussian with expectation μ_n and variance σ_n^2 .

Now that we know this distribution, the required sample is found by generating a standard normal random variable X_0 and simulate X_{n+1} recursively for $n = 0, 1, \dots$. Of course, the difficulty is to organize the calculations to avoid matrix inversion in each step. The algorithm proposed by Hosking [32] computes $d(n)$ recursively. The algorithm is presented here in a slightly different but faster way than in the original article.

Suppose we know μ_n , σ_n^2 and $\tau_n := d(n)'F(n)c(n) = c(n)'F(n)d(n)$. Using (2.5), it is easy to see that σ_n^2 satisfies the recursion

$$\sigma_{n+1}^2 = \sigma_n^2 - \frac{(\gamma(n+2) - \tau_n)^2}{\sigma_n^2}.$$

A recursion for $d(n+1) = \Gamma(n+1)^{-1}c(n+1)$ is also obtained from (2.5):

$$d(n+1) = \begin{pmatrix} d(n) - \phi_n F(n)d(n) \\ \phi_n \end{pmatrix},$$

with

$$\phi_n = \frac{\gamma(n+2) - \tau_n}{\sigma_n^2}.$$

The first n elements of the $(n+2)$ -row vector $d(n+1)$ can thus be computed from $d(n)$ and ϕ_n . Given $d(n+1)$, it is clear how to compute μ_{n+1} , σ_{n+1}^2 and τ_{n+1} . We start the recursion with $\mu_0 = \gamma(1)X_0$, $\sigma_0^2 = 1 - \gamma(1)^2$ and $\tau_0 = \gamma(1)^2$. Note that cumulative sums have to be computed when a sample fractional Brownian motion Y_0, Y_1, \dots is required.

An approach that is closely related to the Hosking algorithm is the innovations algorithm. Instead of recursively generating X_{n+1} given X_n, \dots, X_0 , this method simulates the *innovation* $X_{n+1} - \mu_n$ given $X_n - \mu_{n-1}, \dots, X_1 - \mu_0, X_0$. An advantage of this algorithm is that it can also be applied to non-stationary processes. For details we refer to [11].

Because of the computations in the recursions of $d(n)$, μ_n and τ_n , the algorithm's complexity is of order N^2 for a sample of size N . As an example, the computation of $d(n) - \phi_n F(n)d(n)$ for given ϕ_n and $d(n)$ requires order n computer flops. Doing this for $n = 1, 2, \dots, N$ requires order N^2 flops.

An advantage of the Hosking method is that traces can be generated on-the-fly. This means that the sample size does not need to be known in advance. This occurs for example when a simulation should stop at some random time.

Another strength of this method is its extreme simplicity. This is used in Huang *et al.* [33] to express some likelihoods in terms of by-products of the algorithm, like ϕ_n . Based on these likelihoods, an algorithm for fast simulation in a queueing framework is given.

When more than one trace is needed, as in simulation studies with more than one run, most of the computations need not be done again. However, because the calculation of μ_n depends on the generated points X_0, \dots, X_n , the number of computer flops is still of order N^2 ; the time complexity is thus not reduced.

2.1.2 The Cholesky method

Not surprisingly, the Cholesky method (e.g., [4]) makes use of the so-called Cholesky decomposition of the covariance matrix. This means that the covariance matrix $\Gamma(n)$ can be written as $L(n)L(n)'$, where $L(n)$ is an $(n+1) \times (n+1)$ lower triangular matrix. Denoting element (i, j) of $L(n)$ by l_{ij} for $i, j = 0, \dots, n$, $L(n)$ is said to be lower triangular if $l_{ij} = 0$ for $j > i$. It can be proven that such a decomposition exists when $\Gamma(n)$ is a symmetric positive definite matrix.

Unlike the Hosking method, the Cholesky method can be applied to non-stationary Gaussian processes, but we assume stationarity for notational reasons.

The elements of $L(n)$ can be computed by noting that element (i, j) of $L(n)L(n)'$ and $\Gamma(n)$ should be equal for $j \leq i$ (then also for $j > i$ because of the symmetry). That is,

$$\gamma(i-j) = \sum_{k=0}^j l_{ik} l_{jk}, \quad j \leq i. \quad (2.7)$$

This equation reduces to $\gamma(0) = l_{00}^2$ for $i = j = 0$. For $i = 1$ we get the two equations

$$\gamma(1) = l_{10} l_{00}, \quad \gamma(0) = l_{10}^2 + l_{11}^2$$

determining l_{10} and l_{11} . Since it is clear from this that l_{ij} cannot depend on n , $L(n+1)$ can be computed from $L(n)$ by adding a row (and some zeros to make $L(n+1)$ lower triangular) determined by

$$\begin{aligned} l_{n+1,0} &= \frac{\gamma(n+1)}{l_{00}} \\ l_{n+1,j} &= \frac{1}{l_{jj}} \left(\gamma(n+1-j) - \sum_{k=0}^{j-1} l_{n+1,k} l_{jk} \right), \quad 0 < j \leq n \\ l_{n+1,n+1}^2 &= \gamma(0) - \sum_{k=0}^n l_{n+1,k}^2. \end{aligned}$$

From these formulas follows that $L(n)$ is unique under the additional restriction that the elements on the main diagonal are strictly positive. When $\Gamma(n+1)$ is a positive definite matrix, the non-negativity of $l_{n+1,n+1}^2$ is guaranteed, so that the matrix $L(n+1)$ is real.

Denote by $V(n) = (V_i)_{i=0,\dots,n}$ an $(n+1)$ -column vector of i.i.d. standard normal random variables, and construct $V(n+1)$ from $V(n)$ by padding a standard normal random variable. The key idea is to simulate $X(n) = L(n)V(n)$ recursively. For every $n \geq 0$, $X(n)$ has zero mean and covariance matrix

$$\text{Cov}(L(n)V(n)) = L(n)\text{Cov}(V(n))L(n)' = L(n)L(n)' = \Gamma(n).$$

Summarizing, X_{n+1} can quite easily be simulated once $L(n+1)$ is computed:

$$X_{n+1} = \sum_{k=0}^{n+1} l_{n+1,k} V_k. \quad (2.8)$$

Like the Hosking method, one does not need to set the time horizon in advance. The drawback of the method is that it becomes slow (order N^3 for N points) and demanding in terms of storage, because the matrix $L(n)$, that grows in every step of the recursion, has to be kept in memory.

When more than one sample is needed, a lot of computing time can be saved by calculating $L(n)$ only once. In fact, every additional sample then requires order N^2 computer flops.

Although not obvious by comparing the algorithms, the Hosking method is essentially equivalent to the Cholesky method (but faster!) in the sense that the Hosking method computes implicitly the same matrix $L(n)$. Indeed, in the Hosking method, X_{n+1} is generated by $\mu_n + \sigma_n V_{n+1}$ and μ_n only depends on X_n, \dots, X_0 . Analogously, X_n is generated by $\mu_{n-1} + \sigma_{n-1} V_n$, where μ_{n-1} only depends on X_{n-1}, \dots, X_0 . Repeating this, X_{n+1} is generated by some linear combination of $n+1$ i.i.d. standard normal variables:

$$X_{n+1} = \sum_{k=0}^{n+1} h_{n+1,k} V_k,$$

where we know that $h_{n+1,n+1} = \sigma_n > 0$. The analogy with (2.8) is now clear: just like the Cholesky method, the sample is generated by multiplying a vector with i.i.d. standard normal components with a square lower triangular matrix. But the matrices must then be equal, as the Cholesky decomposition is unique if the elements on the main diagonal are positive!

2.1.3 The Davies and Harte method

This algorithm was originally proposed by Davies and Harte [19] and was later simultaneously generalized by Dietrich and Newsam [23] and Wood and Chan [59]. Like the Hosking and the Cholesky method, the method

tries to find a ‘square root’ of the covariance matrix, in the sense that $\Gamma = GG'$ for some square matrix G . Assume that a sample of size N is needed and that the size of the covariance matrix is a power of two, which means that $N = 2^g$ for some $g \in \mathbb{N}$. To be able to compute such a square root efficiently, the main idea is to embed Γ in a so-called circulant covariance matrix C of size $2N = 2^{g+1}$. More precisely, define C by

$$\left(\begin{array}{ccccccccc} \gamma(0) & \gamma(1) & \cdots & \gamma(N-1) & 0 & \gamma(N-1) & \gamma(N) & \cdots & \gamma(2) & \gamma(1) \\ \gamma(1) & \gamma(0) & \cdots & \gamma(N-2) & \gamma(N-1) & 0 & \gamma(N-1) & \cdots & \gamma(3) & \gamma(2) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \gamma(N-1) & \gamma(N-2) & \cdots & \gamma(0) & \gamma(1) & \gamma(2) & \gamma(3) & \cdots & \gamma(N-1) & 0 \\ 0 & \gamma(N-1) & \cdots & \gamma(1) & \gamma(0) & \gamma(1) & \gamma(2) & \cdots & \gamma(N-2) & \gamma(N-1) \\ \gamma(N-1) & 0 & \cdots & \gamma(2) & \gamma(1) & \gamma(0) & \gamma(1) & \cdots & \gamma(N-3) & \gamma(N-2) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \gamma(1) & \gamma(2) & \cdots & 0 & \gamma(N-1) & \gamma(N-2) & \gamma(N-3) & \cdots & \gamma(1) & \gamma(0) \end{array} \right). \quad (2.9)$$

Note that the i th row can be constructed by shifting the first row $i - 1$ ‘places’ to the right and padding the removed elements on the left. Also note that the matrix is symmetric and that the upper left corner (the shaded area) is Γ . When $\gamma(\cdot)$ is the covariance function of fractional Gaussian noise and the zeros in the matrix are replaced by $\gamma(N)$, the matrix is positive definite [17]. It becomes clear in Section 2.2.4 why we use zeros. This has no influence on the resulting fractional Gaussian noise sample as long as the λ_j are non-negative (in practice, this is satisfied for any reasonable sample size N).

It should be said that the circulant matrix is not necessarily positive definite for general autocovariance functions, but it is described in [23, 59] how to deal with that situation. For some autocovariance functions, it is even impossible to embed the covariance matrix in a positive definite matrix. The authors propose to make the method approximate in that case.

If the number of required points is no power of two, more zeros have to be padded in the first row to obtain a circulant matrix, but this does not change the essence of the method.

The algorithm makes use of the following theorem. Every circulant matrix C can be decomposed as $C = Q\Lambda Q^*$, where Λ is the diagonal matrix of eigenvalues of C , and Q is the unitary matrix defined by

$$(Q)_{jk} = \frac{1}{\sqrt{2N}} \exp\left(-2\pi i \frac{jk}{2N}\right), \quad \text{for } j, k = 0, \dots, 2N-1,$$

where $i = \sqrt{-1}$. Q^* denotes the complex conjugate of the transpose of Q . It is easy to check that Q is indeed unitary: $QQ^* = I$. The eigenvalues, which constitute the matrix Λ , are given by

$$\lambda_k = \sum_{j=0}^{2N-1} r_j \exp\left(2\pi i \frac{jk}{2N}\right) \quad \text{for } k = 0, \dots, 2N-1, \quad (2.10)$$

with r_j the $(j+1)$ th element of the first row of C . Because the matrix C is positive definite and symmetric, the resulting eigenvalues will be positive and real. Therefore, the matrix with eigenvalues $\sqrt{\lambda_1}, \dots, \sqrt{\lambda_{2N-1}}$ and the same eigenvectors as C is also positive definite and real. Note that since $C = Q\Lambda Q^*$ and Q is unitary, the matrix $S = Q\Lambda^{1/2}Q^*$ satisfies $SS^* = SS' = C$. Therefore, S has exactly the desired property. However, S is *not* the same matrix that was found in the Cholesky method, since S is not lower triangular.

To obtain a sample path of the process, we need to find a way to simulate $Q\Lambda^{1/2}Q^*V$, where V is a vector with i.i.d. standard normal elements. This is done in three successive steps:

1. Compute the eigenvalues with Equation (2.10). This can efficiently be done with the Fast Fourier Transform (FFT). For a given complex sequence $(\alpha_k)_{k=0}^{j-1}$, the FFT is an algorithm to compute the Fourier transform of that sequence, i.e.,

$$\sum_{k=0}^{j-1} \alpha_k \exp\left(2\pi i \frac{nk}{j}\right) \quad (2.11)$$

for $n = 0, \dots, j-1$. When j is a power of two, the number of calculations required by the FFT is of order $j \log(j)$; a considerable gain in speed compared to the straightforward calculation of order j^2 .

2. Calculate $W = Q^*V$. A derivation of the covariance structure of W leads to the following simulation scheme:

- Generate two standard normal random variables W_0 and W_N ;

- For $1 \leq j < N$, generate two independent standard normal random variables $V_j^{(1)}$ and $V_j^{(2)}$ and let

$$\begin{aligned} W_j &= \frac{1}{\sqrt{2}}(V_j^{(1)} + iV_j^{(2)}) \\ W_{2N-j} &= \frac{1}{\sqrt{2}}(V_j^{(1)} - iV_j^{(2)}). \end{aligned}$$

The resulting vector W has the same distribution as Q^*V .

3. Compute $Z = Q\Lambda^{1/2}W$:

$$Z_k = \frac{1}{\sqrt{2N}} \sum_{j=0}^{2N-1} \sqrt{\lambda_j} W_j \exp\left(-2i\pi \frac{jk}{2N}\right). \quad (2.12)$$

Again, this calculation is best done with the Fast Fourier Transform for maximum speed. In summary, the sequence $(Z_k)_{k=0}^{2N-1}$ is the Fourier transform of

$$w_k := \begin{cases} \sqrt{\frac{\lambda_k}{2N}} V_k^{(1)} & k = 0; \\ \sqrt{\frac{\lambda_k}{4N}} (V_k^{(1)} + iV_k^{(2)}) & k = 1, \dots, N-1; \\ \sqrt{\frac{\lambda_k}{2N}} V_k^{(1)} & k = N; \\ \sqrt{\frac{\lambda_k}{4N}} (V_{2N-k}^{(1)} - iV_{2N-k}^{(2)}) & k = N+1, \dots, 2N-1, \end{cases} \quad (2.13)$$

A sample of fractional Gaussian noise is obtained by taking the first N elements of Z . It is left to the reader to check that Z is real by construction.

Looking at (2.9), we see that the last N elements of Z also have the desired covariance structure. So in this setup, we get a second sample ‘for free’. However, these two samples may *not* be put together to obtain a double-sized sample, because the correlation structure *between* the two samples is not according to fractional Gaussian noise. Since the two samples are moreover not independent, this second sample is mostly useless.

The main advantage of this method is the speed. More precisely, the number of computations is of order $N \log(N)$ for N sample points. When more traces are needed, the eigenvalues need only to be calculated once. However, the calculations in the second step should be done separately for each trace. Because of this, the number of computations is still of order $N \log(N)$.

2.2 Approximate methods

2.2.1 Stochastic representation method

As we saw in Section 1.2.1, Mandelbrot and van Ness [42] defined fractional Brownian motion by a stochastic integral with respect to ordinary Brownian motion. A natural idea is to approximate this integral by Riemann-type sums to simulate the process.

When approximating (1.3) by sums, the first integral should be truncated, say at $-b$. The approximation $\tilde{B}_H(n)$ is for $n = 1, \dots, N$ given by

$$\tilde{B}_H(n) = C_H \left(\sum_{k=-b}^0 [(n-k)^{H-1/2} - (-k)^{H-1/2}] B_1(k) + \sum_{k=0}^n (n-k)^{H-1/2} B_2(k) \right), \quad (2.14)$$

where B_1 resp. B_2 are vectors of $b+1$ resp. $N+1$ i.i.d. standard normal variables, mutually independent. The constant C_H in (2.14) is not equal to the constant in (1.3), since we switched to study normalized fractional Brownian motion.

The approximation can be made better by increasing the truncation parameter b and choosing a finer grid for the Riemann sums. In [14], it is suggested to choose $b = N^{3/2}$. This method is only interesting from a historical point of view and is considered no good way to generate fractional Brownian motion.

2.2.2 Aggregating packet processes

A possible explanation for the self-similarity in LAN traffic is the presence of the Noah effect in the packet processes of individual source-destination pairs, cf. Section 1.1.1. We can therefore simulate processes on the micro-scale and aggregate them to obtain Brownian traffic.

For this, we first have to gain more insight into this aggregation result. This section is mainly based on Willinger *et al.* [58].

Suppose there are S i.i.d. sources. Each source s has active and inactive periods, which is modeled by a stationary binary time series $\{W^{(s)}(t) : t \geq 0\}$. $W^{(s)}(t) = 1$ means that the source is sending a packet at time t and this is not the case if $W^{(s)}(t) = 0$. The lengths of the active ('ON') periods are i.i.d., those of the inactive periods ('OFF') are i.i.d., and the lengths of the ON- and OFF-periods are independent. An OFF-period follows an ON-period and the ON- and OFF-period lengths may have different distributions. Rescaling time by a factor T , let

$$W_S(Tt) = \int_0^{Tt} \left(\sum_{s=1}^S W^{(s)}(u) \right) du,$$

be the aggregated cumulative packet counts in the interval $[0, Tt]$.

For simplicity, let the distributions of the ON- and OFF-periods both be Pareto with joint parameter $1 < \alpha < 2$. Recall that a random variable X has a Pareto distribution with parameter $\alpha > 0$ if $P(X > t) = t^{-\alpha}$ for $t \geq 1$. Note that the ON- and OFF-periods have infinite variance as a result of $1 < \alpha < 2$.

It is pointed out in [58] that the following limit result holds for $\{W_S(Tt) : 0 \leq t < \infty\}$:

$$\lim_{T \rightarrow \infty} \lim_{S \rightarrow \infty} T^{-H} S^{-1/2} \left(W_S(Tt) - \frac{1}{2} T St \right) = \sigma B_H(t) \quad (2.15)$$

for some $\sigma > 0$, where $H = (3 - \alpha)/2$ and $B_H(t)$ denotes fractional Brownian motion with Hurst parameter H . The limits are limits in the sense of finite-dimensional distributions. A similar result holds when other kinds of heavy-tailed distributions for the ON- and OFF-periods are chosen.

We should interpret this asymptotic result with care, because (2.15) is not valid when the order of the limits is reversed. Forgetting about this, we can intuitively say that $W_S(Tt)$ closely resembles

$$\frac{1}{2} T St + T^H \sqrt{S} \sigma B_H(t),$$

which is fractional Brownian traffic with parameters $M = \frac{1}{2} TS$ and $a = 2\sigma^2 T^{2H-1}$ in the notation of Section 1.3.

The above result can be used for simulation purposes, by aggregating a large number of sources with Pareto ON- and OFF-periods. This results in an order N algorithm when N sample points are needed. However, the number of sources that have to be aggregated to get proper output may be quite large, which may result in a serious slowdown. Moreover, there is no clue how many sources are required to produce reasonable results. In fact, because of the mentioned problem with the reversal of the limits, we have to be very suspicious about the obtained sample.

Another aggregation method for simulating long-range dependent processes uses the $M/G/\infty$ queueing model, where customers arrive according to a Poisson process and have service times drawn from a distribution with infinite variance. When X_t denotes the number of customers in the system at time t , $\{X_t : t \geq 0\}$ is asymptotically self-similar [40]. Again, one must trade off length of computation for degree of self-similarity.

2.2.3 (Conditionalized) Random Midpoint Displacement

The Hosking and Cholesky method generate a fractional Gaussian noise sample recursively. Instead of simulating X_{n+1} given the *whole* past, the generation may be sped up by simulating X_{n+1} given the last m generated values for some $m \geq 1$. However, the long-term correlations are then destroyed, which makes this approach not suitable for generating long-range dependent processes.

The idea behind the Random Midpoint Displacement (RMD) method is to generate the X_n in another order, so that conditioning on only one already generated sample point preserves some long-range dependence. The suitability of the RMD method in a teletraffic framework was first investigated in [39].

A generalization of the RMD method, the Conditionalized RMD method [46], generates the sample in the same order as the RMD method, but more sample points are used in the conditioning. The method has two integer-valued truncation parameters, $l \geq 0$ and $r \geq 1$. We construct a sample of fractional Brownian motion on $[0, 1]$ for notational reasons. Note that this sample can be scaled onto an interval of any desired length using the self-similarity property.

Denote $X_{ij} = B_H(j2^{-i}) - B_H((j-1)2^{-i})$ for $i = 0, 1, 2, \dots$, $j = 1, \dots, 2^i$. For fixed i , $\mathcal{X}_i = \{X_{ij} : j = 1, \dots, 2^i\}$ can be considered a scaled sample of fractional Gaussian noise of size 2^i . However, there is a strong correlation between these samples for different i :

$$X_{i-1,j} = X_{i,2j} + X_{i,2j+1}. \quad (2.16)$$

The Conditionalized RMD algorithm makes use of this recursive structure by simulating the processes $\mathcal{X}_0, \mathcal{X}_1, \dots$ until the desired resolution, say $N = 2^g$, is reached. A fractional Brownian motion sample is then found by taking cumulative sums of \mathcal{X}_g . We will refer to the Conditionalized RMD method with truncation parameters l and r as RMD $_{l,r}$.

Assume that we have obtained realizations of $\mathcal{X}_0, \mathcal{X}_1, \dots, \mathcal{X}_{i-1}$. To obtain a realization of \mathcal{X}_i as well, it suffices to generate $X_{i,j}$ for odd j because of (2.16). Note that the realization of \mathcal{X}_{i-1} contains all generated information so far. Let us proceed from left to right, and assume that $X_{i1}, \dots, X_{i,2k}$ have been generated ($k = 0, 1, \dots, 2^{i-1} - 1$). When we continue to generate $X_{i,2k+1}$ by conditioning on *all* past values $X_{i1}, \dots, X_{i,2k}$ in \mathcal{X}_i and *all* values of \mathcal{X}_{i-1} that are still relevant, $X_{i-1,k+1}, \dots, X_{i-1,2^{i-1}}$, no approximation is involved. The algorithm stays then exact but also becomes very slow. Therefore, Norros *et al.* [46] propose to use only the last l values of $X_{i1}, \dots, X_{i,2k}$ (if possible) and the first r values of $X_{i-1,k+1}, \dots, X_{i-1,2^{i-1}}$ (if possible). For $l = 0$ and $r = 1$, the algorithm reduces to the original RMD method described in [39].

Putting this mathematically, we set

$$X_{i,2k+1} = e(i, k) (X_{i,(2k-l+1)\vee 1}, \dots, X_{i,2k}, X_{i-1,k+1}, \dots, X_{i-1,(k+r)\wedge 2^{i-1}})' + \sqrt{v(i, k)} U_{ik},$$

where $\{U_{ik} : i = 0, 1, \dots; k = 0, \dots, 2^{i-1} - 1\}$ is a set of independent standard Gaussian variables, $e(i, k)$ is a row vector such that

$$\begin{aligned} & e(i, k) (X_{i,(2k-l+1)\vee 1}, \dots, X_{i,2k}, X_{i-1,k+1}, \dots, X_{i-1,(k+r)\wedge 2^{i-1}})' \\ &= \mathbb{E} [X_{i,2k+1} | X_{i,(2k-l+1)\vee 1}, \dots, X_{i,2k}, X_{i-1,k+1}, \dots, X_{i-1,(k+r)\wedge 2^{i-1}}], \end{aligned}$$

and $v(i, k)$ is the scalar given by

$$\text{Var} [X_{i,2k+1} | X_{i,(2k-l+1)\vee 1}, \dots, X_{i,2k}, X_{i-1,k+1}, \dots, X_{i-1,(k+r)\wedge 2^{i-1}}],$$

following the convention $a \vee b = \max(a, b)$ and $a \wedge b = \min(a, b)$. The quantities $e(i, k)$ and $v(i, k)$ are computed with a slight modification of the technique that was used in the derivation of the Hosking algorithm, Equation (2.6). The covariances that are needed for this are computed by straightforward substituting the definition of X_{ij} and applying (1.6).

By the stationarity of the increments of B_H and by self-similarity, $e(i, k)$ is independent of i and k when $2k \geq l$ and $k \leq 2^{i-1} - r$. Moreover, it does not depend on l and r when $2^i < l + 2r$. Thus, the number of vectors $e(i, k)$ that needs to be known up to stage j , where j is the smallest integer greater than $\log_2(l + 2r)$, is the total number of vectors that should be computed. This number is not larger than $1 + 2 + 2^2 + \dots + 2^{j-1} = 2^j - 1 < 2l + 4r$. The same holds for the scalars $v(i, k)$, except that ‘independence with respect to i ’ is replaced by scaling a constant factor 2^{-2H_i} .

Some optimizing may be done to calculate the needed vectors $e(i, k)$ and scalars $v(i, k)$ (e.g., in a recursive approach like Hosking’s method rather than explicit matrix inversion), but since $2l + 4r$ is typically quite small and since the sizes of the matrices are therefore also quite small, not much computational gain should be expected from this (if at all).

It is clear that the algorithm as described above is unsuitable for on-the-fly generation of the sample, because the time horizon should be known in advance. It is, however, not necessary to generate the splits in exactly the same order as above. That is, instead of generating each resolution completely before moving to the finer one, we can have several unfinished resolutions at the same time. For details we refer to [46].

The complexity of this algorithm is of order N when a sample of size N is needed. Note that this is faster than the fastest exact method. Therefore, the conditionalized RMD algorithm is extremely interesting for simulation purposes. In Chapter 4, we compare the output this method for several values of l and r to see if a reasonable trade-off between speed and precision is made.

2.2.4 Spectral simulation, the Paxson method and the approximate circulant method

Spectral simulation is a method to simulate stationary Gaussian processes using spectral analysis. When extracting an algorithm out of spectral considerations, the Fast Fourier Transform (FFT) plays a major role.

First, we describe the general principle of spectral simulation, which was already briefly touched in Section 1.2.2. The idea of spectral simulation is to simulate a process in the frequency domain and transform the resulting series to the time domain. Although it is not possible to obtain an exact fractional Brownian motion sample when this approach is followed, we will shortly see that the accuracy increases as the sample size grows.

In the course of the exposition, it becomes clear that spectral simulation is closely related to the Paxson method and the Davies and Harte method. The resulting modification of the Davies and Harte method (making the exact algorithm approximate to speed up the simulation) will be called the approximate circulant method.

Spectral simulation

We have already seen that unambiguous spectral analysis of stochastic processes is only possible for stationary processes. Throughout this subsection, we assume that the required sample size N is a power of two. We deal with a general real-valued stationary Gaussian process with spectral density $f(\lambda)$, possibly with a singularity at 0 (as in the case of fractional Gaussian noise, on which we will focus). See Section 1.2.2 for a review of spectral densities. The considerations in this subsection are mainly based on the theory of spectral analysis, see, e.g., [41, 51]. The results of this subsection have been published as [22].

The spectral analysis of a stationary discrete-time Gaussian process $X = \{X_n : n = 0, \dots, N - 1\}$ shows that it can be represented in terms of the spectral density as

$$X_n = \int_0^\pi \sqrt{\frac{f(\lambda)}{\pi}} \cos(n\lambda) dB_1(\lambda) - \int_0^\pi \sqrt{\frac{f(\lambda)}{\pi}} \sin(n\lambda) dB_2(\lambda), \quad (2.17)$$

where the equality is to be understood as equality in distribution. This result is called the spectral theorem. The two integrators are mutually independent (ordinary) Brownian motions; for the definition of such a stochastic integral, see Section 1.2.1. Using elementary stochastic calculus and (1.10), its validity is easily checked by showing that the right-hand side of (2.17) has covariance function (1.7).

Spectral simulation is based on approximating Equation (2.17); the integrand is replaced by a simple function. Define $\xi_n(\lambda) := \sqrt{f(\lambda)/\pi} \cos(n\lambda)$ and fix some integer ℓ . After setting $t_k = \pi k/\ell$ for $k = 0, \dots, \ell - 1$, we define a simple function $\xi_n^{(\ell)}$ on $[0, \pi]$ for $0 \leq n \leq N - 1$ by

$$\xi_n^{(\ell)}(\lambda) := \sqrt{\frac{f(t_1)}{\pi}} \cos(nt_1) \mathbf{1}_{\{0\}}(\lambda) + \sum_{k=0}^{\ell-1} \sqrt{\frac{f(t_{k+1})}{\pi}} \cos(nt_{k+1}) \mathbf{1}_{(t_k, t_{k+1}]}(\lambda). \quad (2.18)$$

Let $\theta_n^{(\ell)}(\lambda)$ be defined as $\xi_n^{(\ell)}(\lambda)$, but with the cosine terms replaced by sine terms. The first integral in (2.17) is approximated by $\int_0^\pi \xi_n^{(\ell)}(\lambda) dB_1(\lambda)$, which can be computed using Equation (1.4). Since a similar approximation can be made for the second integral, we arrive at the following approximation $\hat{X}_n^{(\ell)}$ of X_n :

$$\hat{X}_n^{(\ell)} := \sum_{k=0}^{\ell-1} \sqrt{\frac{f(t_{k+1})}{\ell}} \left[\cos(nt_{k+1}) U_k^{(0)} - \sin(nt_{k+1}) U_k^{(1)} \right], \quad (2.19)$$

where $U_k^{(i)}$ are again i.i.d. standard normal random variables for $k = 0, \dots, \ell - 1$. The two vectors $U^{(0)}$ and $U^{(1)}$ should also be mutually independent, since B_1 and B_2 are independent as well.

It should be noted that for $H > 1/2$, we approximate the spectral density by functions that have no pole at 0. As already pointed out, this pole is equivalent to long-range dependence. Hence, we approximate a long-range dependent process by a short-range dependent process. Still, we may obtain a sample with a covariance structure that approximates the structure determined by (1.7) and (1.8) very well. This issue is explored further in Section 4.2.

The FFT can be used to calculate (2.19) efficiently. To this end, we define the sequence $(a_k)_{k=0, \dots, 2\ell-1}$ by

$$a_k := \begin{cases} 0 & k = 0; \\ \frac{1}{2} (U_{k-1}^{(0)} + iU_{k-1}^{(1)}) \sqrt{f(t_k)/\ell} & k = 1, \dots, \ell - 1; \\ U_{k-1}^{(0)} \sqrt{f(t_k)/\ell} & k = \ell; \\ \frac{1}{2} (U_{2\ell-k-1}^{(0)} - iU_{2\ell-k-1}^{(1)}) \sqrt{f(t_{2\ell-k})/\ell} & k = \ell + 1, \dots, 2\ell - 1. \end{cases} \quad (2.20)$$

Using (2.11), one can check that the Fourier transform of (a_k) is real and equals (2.19).

Since $\xi_n^{(\ell)}$ approximates ξ_n better for larger ℓ , an interesting question is whether the approximate sample converges to an exact sample as $\ell \rightarrow \infty$. However, different types of convergence of random variables and processes exist; it is at first sight not clear in what sense the approximate sample converges. In the remainder of this subsection, we deal with this convergence issue.

Convergence

We start by deriving the covariance structure of $\hat{X}^{(\ell)}$. From (2.19) it follows that the covariance between $\hat{X}_m^{(\ell)}$ and $\hat{X}_n^{(\ell)}$ for $n, m = 0, \dots, N - 1$ is given by

$$\text{Cov}(\hat{X}_m^{(\ell)}, \hat{X}_n^{(\ell)}) = \sum_{k=0}^{\ell-1} \frac{f(t_{k+1})}{\ell} \cos((m-n)t_{k+1}), \quad (2.21)$$

which depends only on $m - n$. Hence, the spectral simulation method produces stationary approximate samples. Moreover, the covariances approach their exact values as $\ell \rightarrow \infty$:

$$\sum_{k=0}^{\ell-1} \frac{f(t_{k+1})}{\ell} \cos(nt_{k+1}) \rightarrow 2 \int_0^\pi \frac{f(\lambda)}{2\pi} \cos(n\lambda) d\lambda = \frac{1}{2\pi} \int_{-\pi}^\pi f(\lambda) \exp(-in\lambda) d\lambda = \gamma(n), \quad (2.22)$$

where the last equality is (1.10). A more detailed numerical analysis of the covariance structure is deferred to Chapter 4.

It is readily checked with (1.10) that $\text{Cov}(\hat{X}_m^{(\ell)}, \hat{X}_n^{(\ell)})$ converges to $\gamma(|m - n|)$ as $\ell \rightarrow \infty$. From this fact, it is not difficult to deduce that the finite-dimensional distributions of $\hat{X}^{(\ell)}$ converge in distribution to the corresponding finite-dimensional distributions of X as $\ell \rightarrow \infty$. However, we will not prove this, since we can prove an even stronger convergence result: every sample point $\hat{X}_n^{(\ell)}$ converges in \mathcal{L}^2 -norm to the corresponding exact sample point X_n as $\ell \rightarrow \infty$.

The proof of this fact is based on the definition of the stochastic integral appearing in (2.17). Because f is integrable, the function ξ_n is certainly square integrable for $0 \leq n \leq N - 1$. Recall that the discussion on stochastic integration showed that if the sequence of simple functions $(\xi_n^{(\ell)})_\ell$ satisfies, for fixed $0 \leq n \leq N - 1$,

$$\lim_{\ell \rightarrow \infty} \int_0^\pi [\xi_n(\lambda) - \xi_n^{(\ell)}(\lambda)]^2 d\lambda = 0, \quad (2.23)$$

then $\int_0^\pi \xi_n^{(\ell)}(\lambda) dB_1(\lambda)$ converges in \mathcal{L}^2 -norm to $\int_0^\pi \xi_n(\lambda) dB_1(\lambda)$ as $\ell \rightarrow \infty$. It is indeed true that (2.23) holds; a similar result holds for the second integrand of (2.17). A proof of these facts can be found in the appendix of this chapter. By the independence of B_1 and B_2 , we deduce that every sample point $\hat{X}_n^{(\ell)}$ converges to X_n in \mathcal{L}^2 -norm, i.e., in mean square sense.

Since the ‘error’ $\hat{X}_n^{(\ell)} - X_n$ is a centered Gaussian variable for every n (with a variance that decays to zero in ℓ), we have

$$\mathbb{E}|\hat{X}_n^{(\ell)} - X_n|^p = \frac{\Gamma((p+1)/2)}{\sqrt{\pi}} [2\mathbb{E}|\hat{X}_n^{(\ell)} - X_n|^2]^{p/2}. \quad (2.24)$$

Thus, $\hat{X}_n^{(\ell)}$ converges to X_n in \mathcal{L}^p -norm for every $p \geq 1$! Because of the normality, this is equivalent to convergence in probability. This convergence of the individual sample points is readily extended to a joint convergence result: the finite-dimensional distributions of $\hat{X}^{(\ell)}$ converge in probability to the corresponding finite-dimensional distributions of X as $\ell \rightarrow \infty$.

An interesting question is at what rate the convergence takes place. By the way the stochastic integral is constructed, this rate is related to the rate of convergence of (2.23) and its θ_n -counterpart. Let $\epsilon > 0$ be arbitrary, and suppose that $H > 1/2$. As $\sqrt{f(\lambda)/(2\pi)} \cos(n\lambda)$ behaves like $C|\lambda|^{1-2H}$ near $\lambda = 0$, one has on $[0, \pi/\ell]$, for large ℓ ,

$$(C - \epsilon)\lambda^{1/2-H} \leq \sqrt{\frac{f(\lambda)}{2\pi}} \cos(n\lambda) \leq (C + \epsilon)\lambda^{1/2-H}.$$

Set $C' := [(C + \epsilon)/(C - \epsilon)]^{1/(1/2-H)} < 1$, then

$$\begin{aligned} \sigma_{\hat{X}_n^{(\ell)} - X_n}^2 &= \|\xi_n - \xi_n^{(\ell)}\|^2 + \|\theta_n - \theta_n^{(\ell)}\|^2 \\ &\geq \int_0^{\pi/\ell} \left(\sqrt{\frac{f(\lambda)}{2\pi}} \cos(n\lambda) - \sqrt{\frac{f(\pi/\ell)}{2\pi}} \cos(n\pi/\ell) \right)^2 d\lambda \\ &\geq \int_0^{C'\pi/\ell} \left((C - \epsilon)\lambda^{1/2-H} - (C - \epsilon)(C'\pi/\ell)^{1/2-H} \right)^2 d\lambda \\ &\sim \mathcal{K}\ell^{2H-2}, \end{aligned}$$

where \mathcal{K} denotes some constant. Thus, the rate of convergence is quite slow.

Since the length of the output sequence of the FFT algorithm applied to (2.20) must be at least the sample size N , the smallest possible choice for ℓ is $N/2$, although better results are obtained for larger ℓ . For $\ell = N/2$, the spectral simulation approach is closely related to the Paxson method, as will now be made clear.

The Paxson method

Paxson [47] proposes a rather intuitive method for simulating fractional Gaussian noise. By studying the output statistically, he tests if the resulting samples have indeed the desired properties. Unfortunately, the

paper lacks a thorough justification of the proposed procedure, and it remains unclear why the obtained sample should be (close to) Gaussian.

However, with the formulas above, it is possible to make the arguments precise. In the Paxson method, the approximate fGn sample is the Fourier transform of

$$b_k := \begin{cases} 0 & k = 0; \\ \sqrt{\frac{R_k f(t_k)}{N}} \exp(i\Phi_k) & k = 1, \dots, N/2; \\ b_{N-k}^* & k = N/2 + 1, \dots, N - 1, \end{cases}$$

where R_k are independent exponentially distributed random variables with mean 1 for $k \geq 1$, and the asterisk denotes the complex conjugate. Besides $\Phi_{N/2}$, which is set to zero, the Φ_k are independent uniformly distributed random variables on $[0, 2\pi]$ for $k \geq 1$, also independent of the R_k . In this case, t_k equals $2\pi k/N$. Note that the obtained sample is real by construction.

Because R_k is exponentially distributed with mean 1, $\sqrt{2R_k} \exp(i\Phi_k)$ has the same distribution as $U_k^{(0)} + iU_k^{(1)}$, where $U_k^{(0)}$ and $U_k^{(1)}$ are independent standard normal random variables in the usual notation. This fact is also used in the well-known Box-Muller algorithm to simulate Gaussian random variables (e.g., [4]).

Let us now compare this with the spectral simulation method for $\ell = N/2$. In that case, the sample is the Fourier transform of (see (2.20))

$$b'_k := \begin{cases} 0 & k = 0; \\ \sqrt{\frac{f(t_k)}{2N}} (U_{k-1}^{(0)} + iU_{k-1}^{(1)}) & k = 1, \dots, N/2 - 1; \\ \sqrt{\frac{2f(t_k)}{N}} U_{k-1}^{(0)} & k = N/2; \\ \sqrt{\frac{f(t_k)}{2N}} (U_{N-k-1}^{(0)} - iU_{N-k-1}^{(1)}) & k = N/2 + 1, \dots, N - 1, \end{cases} \quad (2.25)$$

where the $U_k^{(i)}$ have their usual meaning. At this point it becomes clear what the relation is between the spectral method and the Paxson method; by comparing b_k to b'_k , we see that the value of $b_{N/2}$ is the only difference (of course, the indexing of the random variables $U^{(i)}$ differs, but this has no impact). However, this effect will vanish for ‘large’ N ; it is readily checked that it does not affect the convergence result of the previous subsection. Thus, the accuracy of the Paxson method increases with the sample size N , and is even exact as $N \rightarrow \infty$ in the sense that each sample point converges to an ‘exact’ sample point in probability.

We can now argue that Paxson’s suggestion to set $\Phi_{N/2} = 0$ is not useful, since this destroys the normality of $b_{N/2}$, which is reflected in *every* sample point, cf. (2.11). This is the reason that Paxson finds in some cases that the resulting sample points have a ‘nearly’ normal distribution. Therefore, we set

$$b_{N/2} = \sqrt{\frac{f(t_{N/2})}{2N}} U_{N/2}^{(0)},$$

although (as before) the imaginary part has no influence on the sample. In the remaining of this paper, we call this *improved* Paxson method simply the Paxson method.

The method is faster than the Davies and Harte method, although still of order $N \log(N)$. This is because the Paxson method requires just one Fourier transform of a sequence of size N , instead of two transforms of size $2N$. Hence, the Paxson method is approximately four times faster than the exact Davies and Harte method, provided N is large. It may thus offer a good alternative to exact simulation when large sample sizes are required. A more detailed runtime comparison is made in Chapter 4.

The fBm samples produced with Paxson fGn samples have the special property that the end point is always 0. This property is easily seen by using the definition of the FFT transform (2.11):

$$\sum_{n=0}^{N-1} \sum_{k=0}^{N-1} b_k \exp\left(2\pi i \frac{nk}{N}\right) = \sum_{k=0}^{N-1} b_k \sum_{n=0}^{N-1} \exp\left(2\pi i \frac{nk}{N}\right) = Nb_0 = 0.$$

This highly undesirable property of Paxson fBm samples can be regarded as a fractional Brownian bridge-effect.

Having recognized that the Paxson method equals the spectral simulation method with $\ell = N/2$, a natural question is what can be said about spectral simulation with $\ell = N$. The next subsection shows that there is a connection with the Davies and Harte method.

The approximate circulant method; connection with the Davies and Harte method

We now make the observation that the spectral simulation approach is related to the circulant diagonalization algorithm of Davies and Harte. In that algorithm, the λ_k are calculated with (2.11), which can be rewritten

as

$$\begin{aligned}\lambda_k &= \sum_{j=0}^{N-1} \gamma(j) \exp\left(2\pi i \frac{jk}{2N}\right) + \sum_{j=1}^{N-1} \gamma(j) \exp\left(2\pi i \frac{(2N-j)k}{2N}\right) \\ &= \sum_{j=-N+1}^{N-1} \gamma(j) \exp\left(2\pi i \frac{jk}{2N}\right).\end{aligned}\quad (2.26)$$

Letting $t_k = \pi k/N$, this can be approximated by the infinite sum $\sum_{j=-\infty}^{\infty} \gamma(j) \exp(ijt_k)$ when N is ‘large’. Note that this is just the spectral density (1.9). In that case, we approximate λ_k with $f(t_k)$ for $k = 0, 1, \dots, N$ and with $f(t_{2N-k})$ for $k = N+1, \dots, 2N-1$. To avoid problems with the pole, $f(0)$ has to be approximated by a finite value. Since we can compute λ_0 directly from (2.26) using (1.7), the choice $f(0) = \lambda_0 = N^{2H} - (N-1)^{2H}$ seems justified.

Instead of using the exact λ_k , we investigate what happens when the approximations $f(t_k)$ are used in the Davies and Harte algorithm to generate a sample. The FFT is then applied to (see (2.13))

$$c_k := \begin{cases} \sqrt{\frac{f(t_k)}{2N}} U_k^{(0)} & k = 0; \\ \sqrt{\frac{f(t_k)}{4N}} (U_k^{(0)} + iU_k^{(1)}) & k = 1, \dots, N-1; \\ \sqrt{\frac{f(t_k)}{2N}} U_k^{(0)} & k = N; \\ \sqrt{\frac{f(t_{2N-k})}{4N}} (U_{2N-k}^{(0)} - iU_{2N-k}^{(1)}) & k = N+1, \dots, 2N-1. \end{cases} \quad (2.27)$$

The first N coefficients of the Fourier transform constitute an approximate fGn sample. Because the only difference with the Davies and Harte method (that was based on a circulant matrix) is the use of *approximate* λ_k , we will refer to this method as *approximate circulant method*. The input coefficients (c_k) closely resemble (2.20) with $\ell = N$; the differences are the coefficients for $k = 0$ and $k = N$. Again, the effects of this difference vanish as the sample size N grows, which indicates that the method is asymptotically exact in the sense described earlier.

Moreover, we can now measure how the spectral simulation method with $\ell = N$ performs in the spectral domain for finite N : it can be calculated what ‘spectral’ error is made by using the spectral density (1.9) rather than the λ_k . A further investigation of this issue is found in Chapter 4.

By approximating λ_k with the value of the spectral density at t_k , the number of applications of the FFT on a sequence of size $2N$ is halved. Instead, the spectral density is evaluated in $2N$ points; this is done in order N time. Therefore, the spectral simulation method with $\ell = 2N$ is theoretically twice as fast as the exact Davies and Harte method for large N . However, the method is in practice only a bit faster than the exact method for reasonable sample sizes.

This observation indicates that it makes no sense to increase ℓ further to $2N$ in the spectral simulation method. Still, sometimes (e.g. in time series analysis) the autocovariances of some relevant process are unknown, although the spectral density has some known closed form. Then, it is an interesting option to simulate such a process with the approximate spectral simulation method with $\ell = 2N$. Note that the proof in the appendix indicates that the convergence result holds for any stationary Gaussian process with spectral density $f(\cdot)$.

Improvements

We have seen that the function $\xi_n^{(\ell)}$ defined in (2.18) converges to ξ_n in the sense of (2.23), which led to all theoretical results in the current section. A natural question is if this simple function is the only possible choice for which this type of convergence holds. A glance at the proof indicates that the answer is negative. The function $\xi_n^{(\ell)}$ was constructed by evaluating the spectral density in the rightmost point of each interval of the grid. Looking at (2.21), another interesting possibility is to replace $f(t_{k+1})/\ell$ by

$$\int_{t_k}^{t_{k+1}} \frac{f(\lambda)}{\pi} d\lambda. \quad (2.28)$$

Unfortunately, it depends on the spectral density $f(\cdot)$ whether this integral can be computed (efficiently). In the fGn case, this integral cannot be computed directly, but should be approximated (e.g., by integrating f numerically). For $k = 0$, the integral can be approximated accurately by replacing $f(\lambda)$ by its Taylor expansion $\sin(\pi H)\Gamma(2H+1)\lambda^{1-2H}$, cf. (1.11).

Instead of using computationally intensive numerical integration techniques to calculate (2.28), it is also possible to interpolate $f(\cdot)$ linearly between the grid points (this is impossible for $k = 0$ when $H > 1/2$

because of the pole at 0, but the Taylor based approximation can then be used). This variant is similar to the approach described above; the spectral density is evaluated at the middle point of each interval on the grid, instead of the rightmost point.

2.2.5 Wavelet-based simulation

An approach that is related to spectral simulation is wavelet-based simulation. Like spectral simulation, the idea is to simulate a process of which the transformed version can be regarded as an approximate sample of fractional Brownian motion. However, this is not the only resemblance between wavelet and spectral simulation; in both cases, frequency information is stored in coefficients. A small example makes the differences clear. Suppose we have a sample with only one frequency (e.g., a scaled sine or cosine). Its Fourier transform is then a spike at that particular frequency. Construct a new double-sized sample by letting the first half equal the original sample, and letting the second half be constant, e.g., zero. Still, the Fourier transform of this new sample has a major spike at the original input frequency, but other frequencies are also considered present in a small amount to account for the second half of the sample. Thus, although the coefficients in the spectral case (i.e., the Fourier transform of a time-domain sample) determine the time-domain sample uniquely, the untransformed coefficients do not provide information on the location of these frequencies in time.

Moreover, it is desirable in applications to have more time resolution (i.e., time information) for high frequencies than for low frequencies. Wavelet coefficients have this property. Thus, instead of simulating the Fourier transform, it is also possible to simulate the wavelet coefficients of a fractional Brownian motion sample. The time-domain sample is then found by transforming the wavelet coefficients.

To understand the basics of wavelets, we will have to go through some theory and technicalities. For more details, we refer to Traas *et al.* [56, in Dutch] or the more technical wavelet lectures by Daubechies [18].

Denote the space of all real-valued square-integrable functions (in Lebesgue sense) on \mathbb{R} by $L^2(\mathbb{R})$. Elements of these spaces are best thought of as classes of functions. Two functions in this space are of the same class if they are almost everywhere equal, i.e., equal up to a set of points with zero Lebesgue measure. As an example, 0 denotes all functions that are almost everywhere zero. An inner-product in this space is defined by

$$\langle f, g \rangle = \int_{\mathbb{R}} f(t)g(t)dt,$$

for $f, g \in L^2(\mathbb{R})$, and a (pseudo-)norm by $\|f\|^2 = \langle f, f \rangle$, to which we refer as the L^2 -norm on \mathbb{R} . Note that this norm is different from the previously encountered L^2 -norm, because that norm was a norm on our probability space (Ω, \mathcal{F}, P) , whereas the L^2 -norm is a norm on the function space $L^2(\mathbb{R})$.

A collection of functions $\{f_k : k \in K\}$ is called orthonormal if $\langle f_i, f_j \rangle = \mathbf{1}(i = j)$ for all $i, j \in K$. It is called a basis for some subset A of $L^2(\mathbb{R})$ if every element $f \in A$ can uniquely be written as

$$f = \sum_{k \in K} \alpha_k f_k$$

for some α_k . When K is infinite but countable, e.g., $K = \mathbb{Z}$, $\sum_{k \in \mathbb{Z}} \alpha_k f_k$ is the unique (in almost everywhere sense) function for which $\lim_{N \rightarrow \infty} \|f - \sum_{k=-N}^N \alpha_k f_k\| = 0$. In this section, orthonormal bases will play an important role, but the results hold (with some minor modifications) as well when a weaker notion than orthonormal basis is used. These ‘weaker’ bases are called Riesz bases. To keep things simple, we refer for the details on the use of Riesz bases to [56].

Instead of a spectral analysis as in the spectral simulation case, a multiresolution analysis (MRA) is performed in the wavelet case. A multiresolution analysis consists of a collection of nested subspaces $\{V_j : j \in \mathbb{Z}\}$ and a function ϕ . Each space V_j is a linear subspace of $L^2(\mathbb{R})$. The subspaces satisfy the following set of properties:

1. $V_j \subset V_{j-1}$ for $j \in \mathbb{Z}$;
2. $\bigcap_{j \in \mathbb{Z}} V_j = \{0\}$;
3. Any function $f \in L^2(\mathbb{R})$ can be approximated by functions in V_j to arbitrary precision in L^2 -norm, i.e., given $f \in L^2(\mathbb{R})$ there exists a sequence $\{g_\ell\} \subset \bigcup_{j \in \mathbb{Z}} V_j$ such that $\lim_{\ell \rightarrow \infty} \|g_\ell - f\| = 0$;
4. $x \in V_j$ if and only if $x(2^j \cdot) \in V_0$;
5. $\phi \in V_0$ and $\{\phi(\cdot - k) : k \in \mathbb{Z}\}$ is an orthonormal basis for V_0 .

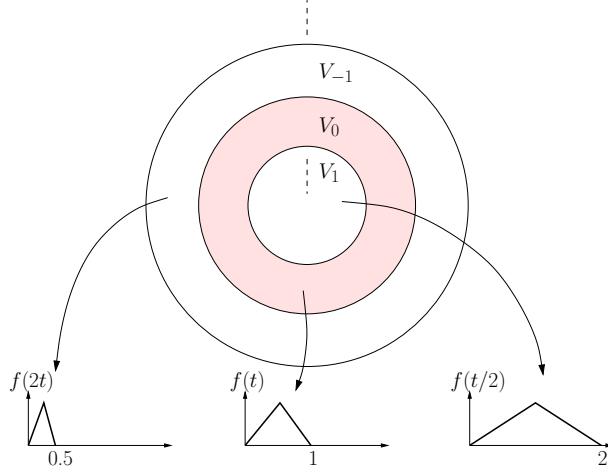


Figure 2.1: Illustration of a multiresolution analysis.

The function ϕ in the fifth property is called the *scaling function*.

Performing a multiresolution analysis of a signal x means successively projecting it onto each of the subspaces V_j . The projection of x onto subspace V_j is called an *approximation* at level j . It can be checked that $\{\phi_{j,k}(t) = 2^{-j/2}\phi(2^{-j}t - k) : k \in \mathbb{Z}\}$ is an orthogonal basis for V_j . Since $V_j \subset V_{j-1}$, the approximation at level j is coarser than the approximation at level $j-1$. The key idea of the MRA consists in examining the loss of information, which is called the *detail*, when going from one approximation to a coarser one. Details can be obtained by subtracting subsequent approximations.

The detail may also directly be computed, but for that we need to define the sets $W_j \subset L^2(\mathbb{R})$ as the set of all elements in V_{j-1} that are orthogonal to V_j . From theory on inner-product spaces follows that the detail of level j is found by the projection of x on W_j . Moreover, the MRA theory shows that there exists a function ψ , called the *mother wavelet*, to be derived from ϕ , such that $\{\psi_{j,k}(t) = 2^{-j/2}\psi(2^{-j}t - k) : k \in \mathbb{Z}\}$ is an orthonormal basis for W_j .

To illustrate the idea of an MRA, the sets V_1 , V_0 and V_{-1} are plotted in a Venn-diagram in Figure 2.1. Possible elements of these spaces are also depicted. An element x of the shaded set can be written as $x = y - z$, where $y \in V_0$, $z \in V_1$, and $\langle z, x \rangle = 0$.

The mother wavelet generated by an MRA should satisfy $\int \psi(t)dt = 0$ because of some regularity issues. Some wavelets even have more: $\int t^k \psi(t)dt = 0$ for $k = 1, \dots, \mathcal{N}$ (and $\int t^k \psi(t)dt \neq 0$ for $k = \mathcal{N} + 1, \dots$). The number \mathcal{N} is called the number of vanishing moments of the wavelet.

A number of ‘standard’ MRA’s have been developed in the literature. The MRA’s that generate the so-called Daubechies family of wavelets are famous examples; they are especially useful since the wavelets have compact support in the time domain. Another example is the Meyer family of wavelets; these are especially useful due to their smoothness.

In an MRA, the information in x is rewritten as a collection of details at different levels and a low-resolution approximation. To illustrate this, let us write $x \in L^2(\mathbb{R})$ as the sum of a coarse approximation at level \mathfrak{J} and the details at all levels smaller or equal to \mathfrak{J} . The approximation is written as a linear combination of the $\phi_{\mathfrak{J},k}$, the detail at level j is written as a linear combination of the $\psi_{j,k}$:

$$x(t) = \sum_{k \in \mathbb{Z}} a_x(\mathfrak{J}, k) \phi_{\mathfrak{J},k}(t) + \sum_{j=-\infty}^{\mathfrak{J}} \sum_{k \in \mathbb{Z}} d_x(j, k) \psi_{j,k}(t). \quad (2.29)$$

The coefficients $a_x(\mathfrak{J}, k)$ and $d_x(j, k)$ are defined by inner products of x with the scaling function resp. mother wavelet:

$$\begin{aligned} a_x(\mathfrak{J}, k) &= \langle x, \phi_{\mathfrak{J},k} \rangle \\ d_x(j, k) &= \langle x, \psi_{j,k} \rangle, \quad \text{for } j \leq \mathfrak{J}. \end{aligned}$$

There exists an algorithm to compute these wavelet coefficients recursively in an efficient way, but we are mainly interested in *construction* of x instead of decomposition.

Application to fractional Brownian motion

In the fractional Brownian motion case, $d_{B_H}(j, k)$ is computed by

$$d_{B_H}(j, k) = \langle B_H, \psi_{j,k} \rangle = \int_{\mathbb{R}} B_H(t) \psi_{j,k}(t) dt. \quad (2.30)$$

Omitting details, the integral is defined as a limit (in L^2 -norm) of stochastic integrals of approximating simple processes. See Section 1.2.1 for a related procedure.

The covariance structure of $d_{B_H}(j, k)$ is studied in Flandrin [29]. The variance of $d_{B_H}(j, k)$ is $\sigma^2 2^{j(2H+1)}$ for some constant $\sigma^2 > 0$. When $H > 1/2$, the covariances of the wavelet coefficients are quite small, but their decay depends heavily on the mother wavelet ψ (mainly through the number of its vanishing moments). Note that the coefficients are (almost) independent if the covariances are very close to zero. As an example, it is shown that the simplest orthonormal wavelet basis, the Haar system ($\mathcal{N} = 1$), for which

$$\psi(t) = \begin{cases} +1 & \text{for } 0 \leq t < 1/2 \\ -1 & \text{for } 1/2 \leq t < 1 \\ 0 & \text{otherwise} \end{cases},$$

the assumption of independence of the wavelet coefficients is severely violated. Nevertheless, when a wavelet with more than one vanishing moment is used, the wavelet coefficients are ‘approximately’ independent.

When we want to use the above considerations to simulate fractional Brownian motion, it is natural to analyze $B_H(t)$ in an MRA framework. Note that a finite fractional Brownian motion sample is with probability 1 an element of $L^2(\mathbb{R})$. An approximate generation method may then be based on (2.29), as suggested in Wornell [60]. The author defines the process

$$\tilde{B}_H(t) = \lim_{J \rightarrow \infty} \sum_{j=-\infty}^J \sum_{k \in \mathbb{Z}} \tilde{d}_{B_H}(j, k) 2^{-j/2} \psi(2^{-j}t - k), \quad (2.31)$$

where $\tilde{d}_{B_H}(j, k)$ are independent Gaussian random variables with variance $\sigma^2 2^{j(2H+1)}$ (the limit in (2.31) is again to be understood as the limit in L^2 -norm). The author shows that this process has spectral properties like fractional Brownian motion. Therefore, \tilde{B}_H is a fractional Brownian motion-like process which can be used as an approximation for real fractional Brownian motion. In fact, the quality of the approximation improves as \mathcal{N} increases. However, the speed of the method is often affected when a wavelet with more vanishing moments is chosen. The Daubechies wavelets are widely used; the Daubechies \mathcal{N} wavelet has \mathcal{N} vanishing moments, and can be characterized by only $2\mathcal{N}$ numbers, a so-called filter. It can be shown that a wavelet corresponding to a filter of size $2\mathcal{N}$ has at most \mathcal{N} vanishing moments, and moreover that the Daubechies \mathcal{N} wavelet is the only wavelet with this property. In practice, the Daubechies 10 wavelet is considered to be a good choice.

It is argued by Abry and Sellan [2] that this method has two major drawbacks. The first is that the coarsest level approximation is just set to zero, because the infinite sum in (2.31) should be truncated. Their argument is that the coarse approximation includes the low-frequency behavior of the sample, whereas the details describe the high-frequency behavior. Forgetting about the coarse approximation means forgetting about the long-range dependence and self-similarity, which is clearly not desirable. Moreover, they argue that the use of independent wavelet coefficients with orthonormal wavelets results in an approximate correlation structure.

However, they describe another wavelet-based method that is proposed by Sellan and Meyer [53]. To understand the method, one needs some theory of generalized processes (or tempered distributions). Using these notions, one can define a ‘derivative’ of Brownian motion (although its paths are almost surely non-differentiable!). This derivative, sometimes called white noise, can be written as follows:

$$W(t) = \sum_k \mu_k \phi(t - k) + \sum_{j \leq 0} \sum_k \lambda_{j,k} \psi_{j,k}(t), \quad (2.32)$$

where μ_k and $\lambda_{j,k}$ are i.i.d. and standard normally distributed. Based on this representation, one obtains a fractional Brownian motion in two steps. First, one *fractionally* integrates (2.32) to obtain a continuous-time analog of fractional Gaussian noise (which again only exists in a ‘generalized process’ sense). The resulting process is then integrated to get the fractional Brownian motion (in the generalized process sense); a final argument then shows that this is an ordinary stochastic process with continuous sample paths. For details on fractional integration, we refer to [49].

The following representation is obtained from (2.32):

$$B_H(t) = \sum_k b_H(k) \phi_{0,k}^{(s)}(t) + \sum_{j \leq 0} \sum_k \lambda_{j,k} 4^{-s} 2^{js} \psi_{j,k}^{(s)}(t), \quad (2.33)$$

where $s = H+1/2$, $\lambda_{j,k}$ are again i.i.d. Gaussian random variables, $b_H(k)$ is a fractional ARIMA(0, s , 0) process (cf. Section 1.2.3), and $\phi^{(s)}$ and $\psi^{(s)}$ are a suitably defined fractional scaling function and wavelet function, to be derived from ϕ and ψ . As before, we set $\phi_{j,k}^{(s)}(t) = 2^{-j/2} \phi^{(s)}(2^{-j}t - k)$ and $\psi_{j,k}^{(s)}(t) = 2^{-j/2} \psi^{(s)}(2^{-j}t - k)$. Note that since the first term (the coarse approximation) accounts for the long-term behavior, this approximation is expected to have better long-range dependent properties than the Wornell wavelet method.

For the details how to construct these fractional scaling and wavelet function, the reader is referred to [2]. The authors make use of reconstruction algorithms for Riesz bases. Their idea is to calculate the generating sequences of the fractional scaling and wavelet function. These generating sequences fully characterize the corresponding function. To be precise, the generating sequence for the fractional scaling function resp. fractional wavelet function are the sequences that satisfy

$$\begin{aligned} \phi^{(s)}(t/2)/\sqrt{2} &= \sum_k u_k^{(s)} \phi^{(s)}(t - k) \\ \psi^{(s)}(t/2)/\sqrt{2} &= \sum_k v_k^{(s)} \psi^{(s)}(t - k). \end{aligned}$$

The generating sequences of the fractional scaling function and wavelet function $u^{(s)}$ and $v^{(s)}$ are computed using the generating sequences u and v of ϕ and ψ . The sequences u and v are tabulated, e.g., in [18]. The ARIMA(0, s , 0) process is constructed by manipulating Gaussian white noise using the same techniques that were used to compute $u^{(s)}$ from u and $v^{(s)}$ from v . However, this leads to an approximate ARIMA(0, s , 0) sample.

When the coefficients in (2.33) are generated, the so-called fast pyramidal Mallat filter bank algorithm can be applied to compute (2.33) in an efficient way. For this, the infinite sum has to be truncated at a finite number $-J$. The authors claim to obtain already relevant approximations for $J = 5$ or $J = 6$ using the Daubechies 10 wavelet.

As usual, denote the required sample size by N . The finest resolution that is present in a sample of size N is determined by $2^{J_f} = N$, but a truncation parameter of $J = 5$ or $J = 6$ may be interesting to limit the computation time. An approximate sample based on (2.33) is found by starting with a fractional ARIMA sample b_H of size $N2^{-J}$. The sum $\sum_k b_H(k) \phi_{0,k}^{(s)}(t)$ is then computed for $t = 2^{-J}, \dots, N2^{-J}$. Since the number of summation terms in $\sum_k \lambda_{j,k} 4^{-s} 2^{js} \psi_{j,k}^{(s)}(t)$ is $N2^{-J-j}$, this results in an order N^2 algorithm when (2.33) is directly calculated. An additional complication is that no explicit expression for $\phi^{(s)}(t)$ exists (even for ϕ and ψ in the Daubechies 10 case it does not exist). However, only the generating sequences need to be known for the reconstruction, since these sequences serve as input in the pyramidal filter bank algorithm. This algorithm depends heavily on the convolution operation, which can efficiently be implemented using the Fast Fourier Transform (FFT). The complexity is then reduced to $N \log N$. Whereas the generating sequence of the Daubechies 10 wavelet has length 20, the generating sequence of $\phi^{(s)}$ and $\psi^{(s)}$ is in general infinite and should be truncated. Obviously, this has an impact on the speed as well.

There is one more thing to take care of. An important operation in the algorithm is the convolution of sequences. When convolving two sequences, the begin and end points of the resulting sequence are polluted by the lack of data; the so-called border effects. These are dealt with by increasing the required sample size with a safety margin and truncating the output sequence (on both sides!). This makes it hard to make theoretical statements on the required amount of time. We investigate the time complexity empirically in Chapter 4.

Just before completion of this report, we found some related wavelet decompositions in the literature. In Meyer *et al.* [43], it is noted that (2.33) is incorrect and that a so-called infrared correction must be added. In Chapter 4, we analyze samples simulated using (2.33); the bad performance observed there might be caused by the lack of the infrared correction, although this does not seem to be a plausible explanation. We also refer to the recent papers by Ayache and Taqqu [5] and [34] for further details on the wavelet method.

2.2.6 Other methods

Numerous other methods have been proposed as an approximate synthesis method for fractional Brownian motion. In this section, we will describe a few and discuss their advantages and drawbacks. Of course, the list is not claimed to be exhaustive.

Chaotic maps

The chaotic maps approach (Erramilli *et al.* [27]) is a method to simulate the source level behavior. Given an initial random value x_0 and a deterministic function f , a deterministic sequence is constructed by the recursion $x_{n+1} = f(x_n)$. The active and inactive periods are determined by the observed sequence, e.g., the source is active at time n when $0 < x_n < 1/2$. The chaos (irregular or seemingly stochastic behavior) arises from a property known as Sensitive dependence on Initial Conditions. This means that a small deviation from the initial condition results in a completely different sequence. In fact, the absolute difference in x_n increases exponentially in n .

It is shown that simple nonlinear chaotic maps already capture some of the behavior of real teletraffic. However, the maps fail to have a direct interpretation. Therefore, several chaotic maps are tried, hoping that satisfactory results for traffic simulation are obtained. Moreover, the output has to be aggregated, which causes the same problems as with the Aggregating packet processes approach (see Section 2.2.2).

The Decreusefond and Lavaud method

The Decreusefond and Lavaud method [20] is based on the following representation of fractional Brownian motion:

$$B_H(t) = \int_0^t K_H(t, s) dB(s)$$

for some function $K_H(t, s)$, where B denotes ordinary Brownian motion as in Chapter 1.

Assume that we need a fractional Brownian motion sample on the interval $[0, 1]$, this interval is then split in N smaller intervals of equal length. Set $t_n = n/N$ for $n = 0, 1, \dots, N$.

We estimate point n of the trace by

$$B_H(t_n) = \sum_{i=0}^n \frac{1}{t_{i+1} - t_i} \int_{t_i}^{t_{i+1}} K_H(t_n, s) ds (B(t_{i+1}) - B(t_i)). \quad (2.34)$$

The integral $\int_{t_i}^{t_{i+1}} K_H(t_n, s) ds$ may not be approximated by $K_H(t_n, t_i)$ or $K_H(t_n, t_{i+1})$, because $K_H(t_n, t)$ is not continuous with respect to t on $[0, t_n]$. The approximation would then not converge to an exact trace when N is large. In [20], the authors describe a rather technical way to overcome this.

Although this method is not really evaluated and tested, the main advantage is clearly that the integral in (2.34) should only be computed once when more than one trace is needed. However, the number of integrals to be calculated is of order N^2 when a sample of length N is needed. All these values should be stored, which is a drawback of the method. Exact methods that are much faster than this method are also known (see Section 2.1.3), so the method is not very useful in practice.

The Dzhaparidze and van Zanten methods

Dzhaparidze and van Zanten have proven a number of interesting series representations, as applications of spectral theory for fractional Brownian motion.

The first [26] involves the positive real zeros $x_1 < x_2 < \dots$ of the Bessel function J_{-H} of the first kind of order $-H$, and the positive zeros $y_1 < y_2 < \dots$ of J_{1-H} . One then has, in the sense of equality in distribution,

$$B_H(t) = \sum_{n=1}^{\infty} \frac{\sin x_n t}{x_n} X_n + \sum_{n=1}^{\infty} \frac{1 - \cos y_n t}{y_n} Y_n,$$

where X_1, X_2, \dots and Y_1, Y_2, \dots are independent centered Gaussian random variables. Their variances are given by

$$\text{Var}X_n = 2c_H^2 x_n^{-2H} J_{1-H}^{-2}(x_n), \quad \text{Var}Y_n = 2c_H^2 y_n^{-2H} J_{-H}^{-2}(y_n),$$

where $c_H^2 = \pi^{-1} \Gamma(1 + 2H) \sin \pi H$. Both series converge absolutely and uniformly in $t \in [0, 1]$ with probability one [26].

The reason why this method is particularly interesting for simulation purposes is that efficient algorithms exist to compute the zeros of Bessel functions. Moreover, these zeros only have to be computed once, regardless the number of traces that need to be simulated. Of course, the series have to be truncated at some level. The question raises what level is appropriate; we do not discuss this issue in this report, but we note that the following estimate holds [25]:

$$\limsup_{N \rightarrow \infty} \frac{N^H}{\sqrt{\log N}} \mathbb{E} \sup_{t \in [0, 1]} \left| \sum_{n>N} \frac{\sin x_n t}{x_n} X_n + \sum_{n>N} \frac{1 - \cos y_n t}{y_n} Y_n \right| < \infty.$$

It turns out that $N^{-H}\sqrt{\log N}$ is the ‘best’ possible rate for a series expansion of fractional Brownian motion [38].

Interestingly, since the x_i and y_i increase, truncation only concerns the *high* frequencies. Therefore, the method is expected to work especially well if the Hurst parameter H is close to one, cf. Figure 1.2. Notice that the high frequencies (details) are also truncated in the wavelet method.

The second series representation is similar, but only involves the zeros of a single Bessel function, J_{1-H} . For $n \in \mathbb{N}$, define σ_n by

$$\sigma_n^{-1} = \frac{1-H}{H} \frac{\Gamma^2(1-H)\Gamma(3/2-H)}{\Gamma(H+1/2)\Gamma(3-2H)} \left(\frac{\omega_n}{2}\right)^{2H} J_{-H}^2(\omega_n) T^{2-2H},$$

where $\omega_1 < \omega_2 < \dots$ are the positive zeros of J_{1-H} . Let $X, Y_1, Y_2, \dots, Z_1, Z_2, \dots$ be i.i.d. standard normal variables. Then with probability one the series

$$\frac{t}{\sqrt{2-2H}} X + \sum_{n=1}^{\infty} \frac{\sin 2\omega_n t/T}{\omega_n/T} \sqrt{\sigma_n} Y_n + \sum_{n=1}^{\infty} \frac{\cos 2\omega_n t/T - 1}{\omega_n/T} \sqrt{\sigma_n} Z_n$$

converges uniformly in $t \in [0, T]$ and defines a fractional Brownian motion with Hurst parameter H . As the first method, one can prove rate-optimality of this expansion.

Appendix: Proof of (2.23)

In this subsection, we prove that for $n = 0, 1, 2, \dots$:

$$\lim_{\ell \rightarrow \infty} \int_0^\pi \left[\sqrt{\frac{f(\lambda)}{\pi}} \cos(n\lambda) - \xi_n^{(\ell)}(\lambda) \right]^2 d\lambda = 0 \quad \text{and} \quad (2.35)$$

$$\lim_{\ell \rightarrow \infty} \int_0^\pi \left[\sqrt{\frac{f(\lambda)}{\pi}} \sin(n\lambda) - \theta_n^{(\ell)}(\lambda) \right]^2 d\lambda = 0, \quad (2.36)$$

where f denotes the spectral density of fractional Gaussian noise for some $0 < H < 1$ and (with $t_k = k\pi/\ell$)

$$\begin{aligned} \xi_n^{(\ell)}(\lambda) &= \sqrt{\frac{f(t_1)}{\pi}} \cos(nt_1) \mathbf{1}_{\{0\}}(\lambda) + \sum_{k=0}^{\ell-1} \sqrt{\frac{f(t_{k+1})}{\pi}} \cos(nt_{k+1}) \mathbf{1}_{(t_k, t_{k+1}]}(\lambda) \quad \text{and} \\ \theta_n^{(\ell)}(\lambda) &= \sqrt{\frac{f(t_1)}{\pi}} \sin(nt_1) \mathbf{1}_{\{0\}}(\lambda) + \sum_{k=0}^{\ell-1} \sqrt{\frac{f(t_{k+1})}{\pi}} \sin(nt_{k+1}) \mathbf{1}_{(t_k, t_{k+1}]}(\lambda). \end{aligned}$$

We keep n fixed throughout the proof, and start by showing that $\theta_n^{(\ell)}$ converges to $\theta_n := \sqrt{\frac{f(\lambda)}{\pi}} \sin(n\lambda)$ in L^2 -norm as $\ell \rightarrow \infty$. Note that f has a pole at 0, but that the sine-term compensates this pole. Thus, $\theta_n(\lambda)$ is continuous on $[0, \pi]$ for every Hurst parameter $0 < H < 1$. It follows that $[\theta_n(\lambda)]^2$ is Riemann integrable on $[0, \pi]$, which implies that $\lim_{\ell \rightarrow \infty} \int_0^\pi [\theta_n^{(\ell)}(\lambda)]^2 d\lambda = \int_0^\pi [\theta_n(\lambda)]^2 d\lambda$. Since $\theta_n^{(\ell)}(\lambda) \rightarrow \theta_n(\lambda)$ for every $0 \leq \lambda \leq \pi$ as $\ell \rightarrow \infty$, we have the desired convergence of $\theta_n^{(\ell)}$ to θ_n in L^2 -norm, see for instance Theorem 4.5.4 of Chung [13].

More care is needed to prove (2.35). When $0 < H \leq 1/2$, the same reasoning as above applies. However, $\xi_n := \sqrt{\frac{f(\lambda)}{\pi}} \cos(n\lambda)$ has a pole at 0 for $1/2 < H < 1$. Still, f is (Riemann) integrable because f is a spectral density; in fact, $\int_0^\pi f(\lambda) d\lambda = \pi$. For any $\epsilon > 0$, this makes it possible to find a $\delta > 0$ (independent of ℓ) such that

$$\begin{aligned} \int_0^\delta [\xi_n^{(\ell)}(\lambda) - \xi_n(\lambda)]^2 d\lambda &\leq \int_0^\delta [\xi_n^{(\ell)}(\lambda) + \xi_n(\lambda)]^2 d\lambda \leq 4 \int_0^\delta [\xi_n(\lambda)]^2 d\lambda \\ &\leq 4 \int_0^\delta f(\lambda) d\lambda < \epsilon/2, \end{aligned}$$

where the first two inequalities use the fact that $\xi_n(\lambda) \geq \xi_n^{(\ell)}(\lambda) \geq 0$ for small λ . As before, we have $\xi_n^{(\ell)}(\lambda) \rightarrow \xi_n(\lambda)$ for every $\delta \leq \lambda \leq \pi$, and the Riemann integrability of $[\xi_n(\lambda)]^2$ on $[\delta, \pi]$, which implies

that $\lim_{\ell \rightarrow \infty} \int_{\delta}^{\pi} [\xi_n^{(\ell)}(\lambda)]^2 d\lambda = \int_{\delta}^{\pi} [\xi_n(\lambda)]^2 d\lambda$. Hence, it is possible to find an ℓ with the property that $\int_{\delta}^{\pi} [\xi_n^{(\ell)}(\lambda) - \xi_n(\lambda)]^2 d\lambda < \epsilon/2$ and the claim is proven.

Note that the arguments of the proof apply for a general spectral density, not necessarily the spectral density of fGn.

CHAPTER 3

Estimation and testing

We have seen that there exist some fast approximate methods that may offer an alternative for exact methods. We hope that the samples of the approximate methods share several properties with exact samples. Among the most important properties of an approximate trace are its long-range dependence and self-similarity; since we are especially interested in the case $H > 1/2$, we often tacitly impose this restriction throughout this chapter. One of the criteria in the comparison of approximate methods, which is the topic of Chapter 4, is how well the sample incorporates this long-range dependence. In the present chapter, we discuss some popular methods to estimate the long-range dependence parameter H , of which we benefit in the next chapter. In addition, two tests are described that test whether a sample can be distinguished from fractional Gaussian noise with some specified Hurst parameter.

The biggest problem that arises when trying to estimate a Hurst parameter exceeding $1/2$ is that the tail behavior of the autocovariance function cannot be estimated from a finite sample. However, some characteristics of long-range dependence and self-similarity also show up in a finite sample. These characteristics and the related estimation methods are discussed in this chapter.

3.1 Estimation

We start with an overview of some popular estimation methods for the long-range dependence parameter H . It should be said that this section has highly benefitted from the publicly available plotting and estimation programs in the AT&T S-language on the homepage of M.S. Taqqu [54]. Most of the methods below are also described by Taqqu *et al.* [55].

All plots are generated with the same fractional Gaussian noise sample (or its cumulative sum fractional Brownian motion if required), that is simulated using the (exact) Davies and Harte method. The Hurst parameter is $H = 0.8$ and the process is simulated on $\{0, 1/N, \dots, (N-1)/N\}$, whereas the size of the sample is $N = 2^{14}$. In Chapter 1, we denoted a fractional Brownian motion sample on $\{0, \dots, N-1\}$ by Y_0, \dots, Y_{N-1} in Chapter 1, but we now use the same notation for the sample on $\{0, 1/N, \dots, (N-1)/N\}$. The same is done for the corresponding fractional Gaussian noise sample. Note that X_k thus has variance N^{-2H} .

Although specific properties of fractional Gaussian noise are used in the derivation of some estimators, the estimators in this section are also widely used on *general* long-range dependent processes. Some of the estimators in this chapter are heuristically motivated assuming fractional Gaussian noise input, but are in practice applied to any sample of a (stationary) process.

3.1.1 Aggregated variance method

The aggregated variance method is based on the self-similarity property of the sample. Recall the definition of the aggregated processes $X^{(m)}$:

$$X_k^{(m)} = \frac{1}{m} (X_{km} + \dots + X_{(k+1)m-1}),$$

for $k = 0, 1, \dots$. Because of the self-similarity (at least asymptotically), $X^{(m)}$ has the same (finite-dimensional) distributions as $m^{H-1}X$ for big m . In particular, $\text{Var}(X_k^{(m)}) = m^{2H-2}\text{Var}(X_k)$. The variance of $X_k^{(m)}$ is equal for every k and a plausible estimator is

$$\widehat{\text{Var}}(X_k^{(m)}) = \frac{1}{M} \sum_{i=0}^{M-1} \left(X_i^{(m)} - \overline{X^{(m)}} \right)^2, \quad (3.1)$$

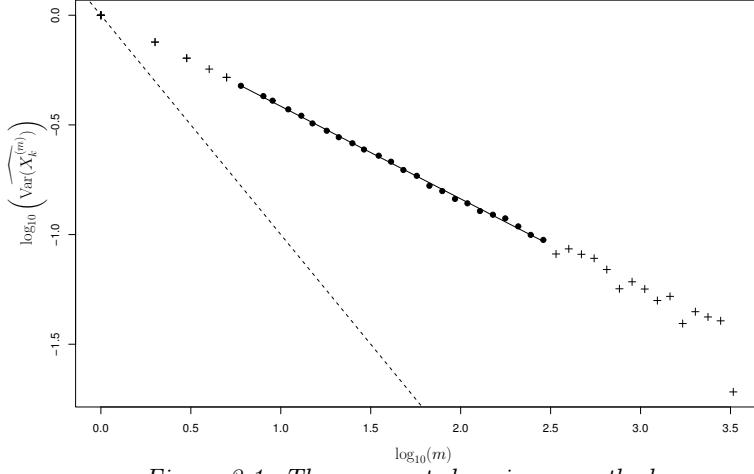


Figure 3.1: The aggregated variance method.

where $\overline{X^{(m)}}$ denotes the sample average of $X^{(m)}$:

$$\overline{X^{(m)}} = \frac{1}{M} \sum_{i=0}^{M-1} X_i^{(m)}.$$

In both formulas, M is the integer part of N/m . The estimator of H is obtained by plotting $\widehat{\text{Var}(X_k^{(m)})}$ versus m on a log-log scale. When the estimates of the variances were equal to their real values, all the points would lie on a straight line with slope $2H - 2$. In practice, the slope is estimated by fitting a straight line through the points. The estimate for H is found from the estimate of the slope.

However, this is very naive, since estimator (3.1) is biased in the presence of non-zero correlations, particularly long-range dependence. It is pointed out in Beran [9] that this bias disappears when M is large (i.e., N is large and m is small), but that this happens very slowly for samples with long-range dependence. In fact, it holds that

$$\mathbb{E}\left(\widehat{\text{Var}(X_k^{(m)})}\right) \sim \text{Var}(X_k^{(m)}) [1 - CM^{2H-2}], \quad (3.2)$$

for some constant C and $M \rightarrow \infty$. In the presence of positive correlations (i.e., $H > 1/2$), C is positive and $\widehat{\text{Var}(X_k^{(m)})}$ tends to underestimate $\text{Var}(X_k^{(m)})$ for small M . This leads also to an underestimate of H (assuming that the estimator of the slope is not too much biased). Moreover, we see that this bias increases in H , which is also found in the empirical study by Taqqu *et al.* [55].

In addition to the already mentioned problems for small M (i.e., ‘large’ m), the estimates of the variances are based on only few observations and therefore unreliable. To avoid these observations from spoiling the estimate for H , those estimates are simply not used in the fitting procedure. The same is done for the observations for ‘small’ m , because those may be based on the short-term correlation structure (although these are high-quality estimates if no short-range dependence is present, as in the fractional Gaussian noise case).

From the theory on least-squares fitting, it is known that the estimator of the slope (and thus for H) is unbiased under some conditions on the so-called disturbance term. Among others, the independence and normality of $\log\left(\widehat{\text{Var}(X_k^{(m)})}\right)$ for a range of values of m should be studied in more detail. If m is large (M small), the normality is questionable, see (3.1). Independence will not occur in general, since the *whole* series is used to compute $\widehat{\text{Var}(X_k^{(m)})}$ for every m . Because all line-fitting estimators in this chapter share these problems, they are all biased.

In Figure 3.1, the log-log plot is given for our sample. The observations that are used in the fitting of the line are dotted. From this plot, the estimated Hurst parameter is 0.7889, quite close to the real value of 0.8. As expected, the estimate is indeed lower than its real value. A weakness of the method is that the fitting region may arbitrarily be chosen (in our plot from 0.7 to 2.5 on a \log_{10} scale), which results in a non-robust estimate. For our sample, the estimated variances for small m lie also on the fitted line, because our sample was an exact fractional Gaussian noise sample without special short-term behavior. The dashed line corresponds to a Hurst parameter of 0.5.

3.1.2 Absolute moments method

The absolute moments method is a generalization of the aggregated variance method. It uses the same principle that $X^{(m)}$ has the same (finite-dimensional) distribution as $m^{H-1}X$ for big m (in the case of an exact sample even for all $m \geq 1$). We first intuitively study the quantity

$$AM_m = \frac{1}{M} \sum_{i=0}^{M-1} \left| X_i^{(m)} - \overline{X^{(m)}} \right|^n$$

for some $n > 0$. In view of (3.2), it might well be possible that the following holds:

$$\begin{aligned} \mathbb{E}AM_m &\sim \mathbb{E} \left| X_i^{(m)} - \mathbb{E}X_i^{(m)} \right|^n \left(1 - C_n M^{n(H-1)} \right) \\ &= m^{n(H-1)} E |X_i - \mathbb{E}X_i|^n \left(1 - C_n M^{n(H-1)} \right), \end{aligned} \quad (3.3)$$

for big m as $M \rightarrow \infty$, where C_n is some constant. Although it is, to our knowledge, not known if (3.3) holds, $\mathbb{E}AM_m$ is proportional to $m^{n(H-1)}$ (see [55]) and an estimate of H can thus be computed from the estimated slope in a regression on a log-log scale. A simulation study in Chapter 4 suggests that this estimator has negative bias, which is confirmed in the intuitive setting of (3.3).

Although every value for $n > 0$ leads to an estimate (n is not necessarily integer-valued), the method is typically used with $n = 1$, the absolute moments case. The method reduces to the aggregated variance method if $n = 2$, which implies that it has the same drawbacks as the aggregate variance method, and produces a similar plot as in Figure 3.1. Applied to our sample, an estimate of 0.7827 is obtained (with the same values for the parameters of the estimation procedure). In what follows, $n = 1$ will be used.

3.1.3 Discrete variations

This method is recently proposed by Coeurjolly [14] and makes use of filtering techniques. Denote by a a filter of length $\ell + 1$ and of order $p \geq 1$, i.e., a sequence that satisfies

$$\sum_{q=0}^{\ell} a_q q^r = 0, \quad \text{for } r = 0, \dots, p-1, \quad (3.4)$$

whereas this does not hold for $r = p$. The process V^a is defined by filtering the sample $Y = \{Y_i = \sum_{k=0}^i X_k : i = 0, \dots, N-1\}$:

$$V_k^a = \sum_{q=0}^{\ell} a_q Y_{k-q}, \quad \text{for } k = \ell, \dots, N-1.$$

An example of a filter of order $p = 1$ is $a = (-1, 1)$, in which case V_k^a reduces to X_k .

The discrete variations method is based on the k -th absolute moment of discrete variations, defined by

$$S(k, a) = \frac{1}{N-\ell} \sum_{i=\ell}^{N-1} |V_i^a|^k, \quad (3.5)$$

for some $k > 0$. Assuming that the sample is Gaussian, the standard formula for the k th moment of a Gaussian variable (see (2.24)) yields

$$\mathbb{E}S(k, a) = \frac{\Gamma((k+1)/2)}{\sqrt{\pi}} N^{-kH} \{2\text{Var}(N^H V_1^a)\}^{k/2}. \quad (3.6)$$

Note that unlike the two preceding methods, the mean $\frac{1}{N-\ell} \sum_{i=\ell}^{N-1} V_i^a$ is not subtracted in (3.5) in the discrete variations method.

The author proposes to estimate H by solving Equation (3.6) to H , where $\mathbb{E}S(k, a)$ is replaced by its estimate. This is possible since $\text{Var}(N^H V_1^a)$ does not depend on H (recall that Y_0, \dots, Y_{N-1} is a sample on $\{0, 1/N, \dots, (N-1)/N\}$) and can explicitly be computed. In fact, we obtain a very good estimate for our sample; the estimate is very close to the input Hurst parameter. However, the Hurst parameter is estimated using only the estimates of ℓ short term covariances. This is not desirable, since long-range dependence mainly concerns the high-lag correlations. Therefore, a sample with a short-range component (like a fractional ARIMA sample) would produce highly biased estimates, which is clearly undesirable for an estimator of the long-range dependence parameter.

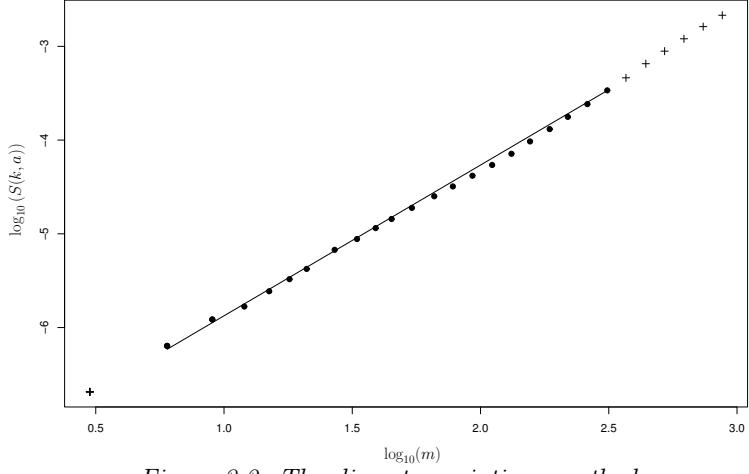


Figure 3.2: The discrete variations method.

A second estimation method that is based on $S(k, a)$ (also proposed in [14]) can be regarded as a generalization of the absolute moments method. Let us define the sequence of filters $(a^m)_{1 \leq m \leq M}$ for some M by

$$a_i^m = \begin{cases} a_j & \text{for } i = jm \\ 0 & \text{otherwise} \end{cases}.$$

Using (3.4) and (3.6) for $r = 0$, it is readily checked that $\mathbb{E}S(k, a^m) = m^{Hk}\mathbb{E}S(k, a)$. An estimator of H is thus deduced from a linear regression of $\{\log S(k, a^m) : 1 \leq m \leq M\}$ on $\{k \log(m) : 1 \leq m \leq M\}$. This method reduces to the absolute moments method for $a = (-1, 1)$ (up to the subtraction of the mean, which is close to zero).

More importantly, the author proves that this estimator converges at a rate of \sqrt{N} to a zero-mean Gaussian distribution with variance $\sigma^2(k, a)$ for some $\sigma^2(k, a) > 0$. He also shows that this variance is minimal for $k = 2$. Although it is not clear that we can neglect the influence of the subtraction of the mean, this suggests that the estimator of the method of moments is inferior to the estimator of the method of aggregated variance, possibly only for large sample sizes.

Note that it is assumed that the sample is Gaussian, or at least that (3.6) holds for a given k . The performance of this method has not been investigated for non-Gaussian samples, but this is no problem for our fractional Brownian motion setting.

To eliminate the impact of short term behavior on the estimator, it is also possible to use only the filters a^m with $M_1 \leq m \leq M_2$ for some M_1, M_2 in the regression. Applied to our sample, this produces a plot as given in Figure 3.2. The bounds M_1 and M_2 are chosen as in Section 3.1.1, whereas the used filter is the Daubechies 4 filter (see Daubechies [18]) for which $p = 2$. The resulting estimate for the Hurst parameter is 0.7913 ($k = 2$).

3.1.4 The Higuchi method

As suggested by its name, this method was proposed by Higuchi [31]. The method is quite similar to the method of absolute moments with $n = 1$. Instead of using non-intersecting blocks, a sliding window is used, which makes the method computationally much more intensive than the above methods.

The quantity on which the Higuchi method is based is

$$L(m) = \frac{N-1}{m^3} \sum_{i=1}^m \frac{1}{M_i} \sum_{k=1}^{M_i} \left| \sum_{j=i+(k-1)m+1}^{i+km} X_j \right|,$$

where M_i signifies the integer part of $(N-i)/m$. The estimate of H is found by plotting $L(m)$ in a log-log plot versus m and adding 2 to the slope of the fitted straight line.

Again, small and large values of m should not be used in the fitting. For our sample, the method gives an estimate of 0.7865, see Figure 3.3.

3.1.5 Periodogram method

The periodogram method estimates the Hurst parameter by fitting a straight line in the spectral domain. This is based on the observation that the spectral density (1.11) behaves like $c_f |\lambda|^{1-2H}$ for $|\lambda| \rightarrow 0$.

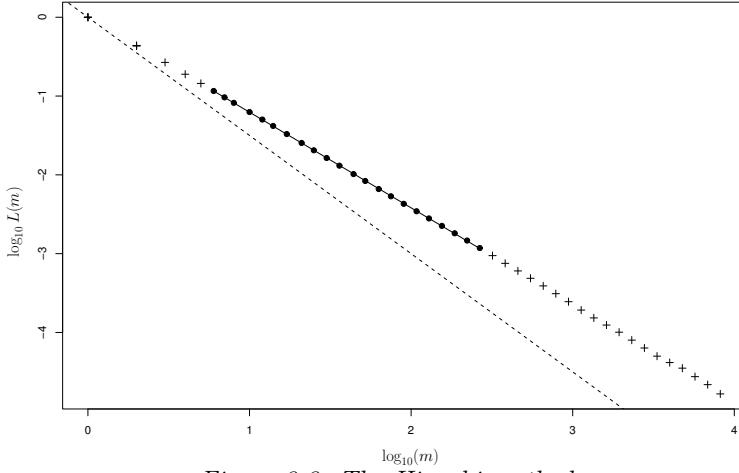


Figure 3.3: The Higuchi method.

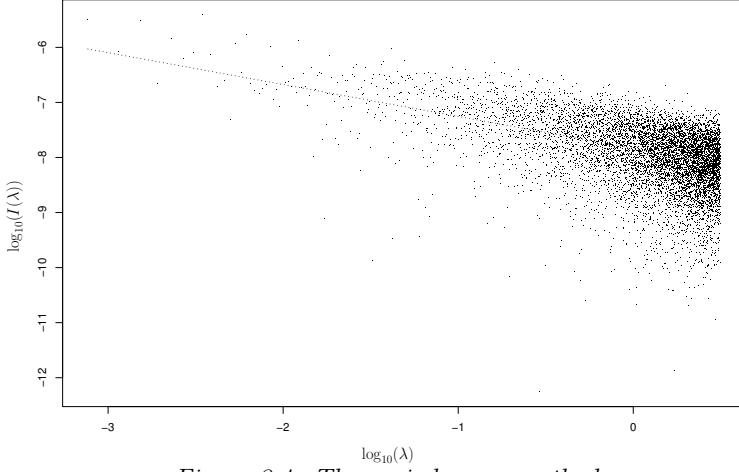


Figure 3.4: The periodogram method.

Equation (1.9) suggests to estimate the spectral density by the *periodogram* defined by

$$I(\lambda) = \sum_{j=-(N-1)}^{N-1} \hat{\gamma}(j) \exp(ij\lambda),$$

with the sample autocovariance $\hat{\gamma}(j)$ function given by

$$\hat{\gamma}(j) = \frac{1}{N} \sum_{k=0}^{N-|j|-1} (X_k - \bar{X})(X_{k+|j|} - \bar{X}),$$

with an obvious definition for the sample average \bar{X} . One can show (see, e.g., Priestley [51]) that this definition of the periodogram equivalent is to

$$I(\lambda) = \frac{1}{N} \left| \sum_{k=0}^{N-1} (X_k - \bar{X}) \exp(ik\lambda) \right|^2. \quad (3.7)$$

It is clear that the periodogram is symmetric around zero, just as the spectral density. The periodogram can be shown to be an asymptotically unbiased estimator of the spectral density f , i.e., [9]

$$\lim_{N \rightarrow \infty} \mathbb{E}[I(\lambda)] = f(\lambda).$$

We compute $I(\lambda_k)$ for $k = 1, \dots, N$, where $\lambda_k = \pi k / N$. In Figure 3.4, the values of the periodogram of our sample at frequencies λ_k are plotted on a log-log scale. As can be seen, the values of the periodogram for contiguous frequencies lie far apart. This can be explained by the fact that for a finite number of frequencies $\lambda_1, \dots, \lambda_N \in (0, \pi)$, the corresponding periodogram ordinates $I(\lambda_1), \dots, I(\lambda_N)$ are ‘approximately’ independent exponential random variables with means $f(\lambda_1), \dots, f(\lambda_N)$ (see Beran [9]). Therefore, the periodogram

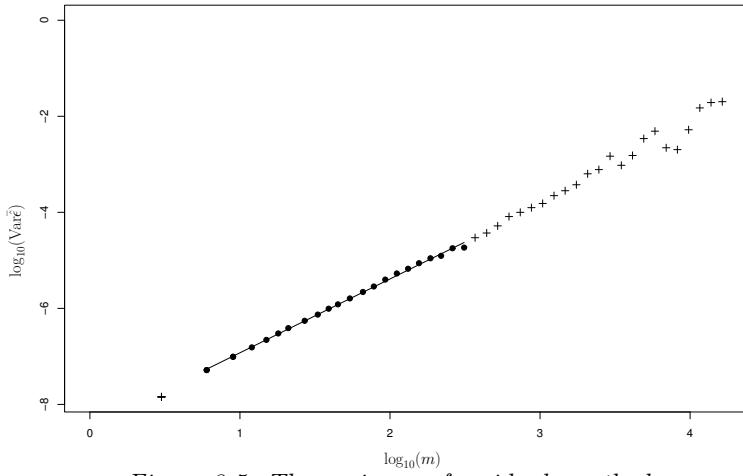


Figure 3.5: The variance of residuals method.

method is not suitable to obtain a high-quality estimation of the Hurst parameter. However, this fact can also be used to perform a regression analysis with an error term based on the exponential distribution, see again [9].

An estimate of H is obtained by fitting a straight line through the data, which has theoretically a slope of $1 - 2H$, in a log-log plot. Since we know that the λ^{1-2H} -behavior is only valid for low frequencies, we use the lowest say 10% of the frequencies. Of course, the choice of this number is subject to ambiguity. The line that is found in this way is also plotted in Figure 3.4, and this leads to an estimate of the Hurst parameter of 0.7897 for our sample.

Several improvements have been proposed for the periodogram method. For example, the modified periodogram approach divides the frequency axis in logarithmically equally spaced boxes, of which the values are averaged to obtain one single value in one box. However, the results are not much better. For details the reader is referred to Taqqu *et al.* [55].

3.1.6 Variance of the regression residuals

This method has been proposed by Peng *et al.* [48, cited according to [55]]. First, the series is broken up into blocks of size m . Within each block k , the partial sums are regressed on a line $\hat{\alpha}^{(k)} + \hat{\beta}^{(k)}i$. The residuals of this regression are given by

$$\hat{\epsilon}_i^{(k)} = \sum_{j=km}^{km+i-1} X_j - \hat{\alpha}^{(k)} - \hat{\beta}^{(k)}i.$$

The sample variance of the residuals is then computed for each block. The average of this sample variance over all blocks is proportional to m^{2H} . For a proof, see [55].

In Figure 3.5, the variance of the residuals are plotted versus m on a log-log scale. This leads to an estimate of 0.7686. As with all line-fitting techniques, we should not take the observations for small and large m very seriously. For Figure 3.5, the same range of values for m is used as in the previous methods to compute the estimate.

The variance of the regression residuals method is extensively studied by Cannon *et al.* [12]. In their terminology, the method is one of the scaled windowed variance methods. Another variant ‘detrends’ within a block by subtracting the line between the first and last point.

3.1.7 R/S analysis

In 1951, Hurst needed to determine what minimum height of the proposed Aswan dam would provide sufficient storage capacity to govern the flow of the Nile River downstream. He observed that measurements of flows showed series of low flow years and high flow years. Inspired by the biblical story on years of drought and prosperity, he called this the Joseph effect, which has become synonymous with long-range dependence. He developed a tool to which is now known as rescaled range (R/S) analysis. The R/S method is widely used in the analysis of long-range dependent processes.

To compute the R/S statistic, the whole series is divided in K non-intersecting blocks that all contain M elements (M is then the greatest integer that is smaller than N/K). The rescaled adjusted range $R(t_i, r)/S(t_i, r)$ is then computed for a number of ranges r , where $t_i = M(i-1)$ are the starting points of the blocks for

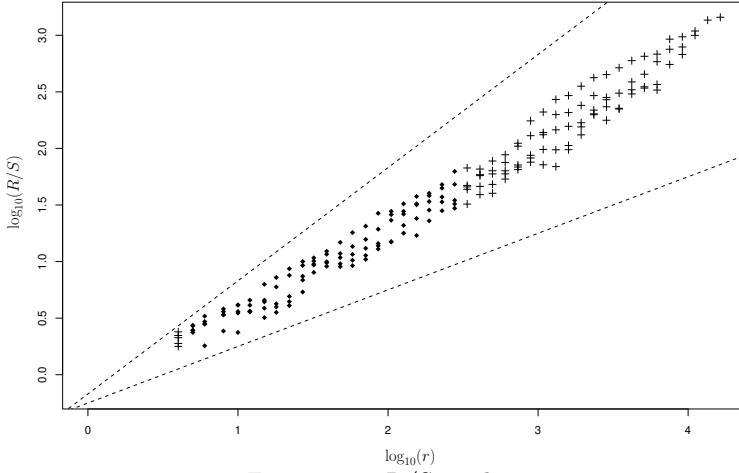


Figure 3.6: R/S analysis.

$i = 1, \dots, K$. For a given range r , $R(t_i, r)$ can only be computed when

$$t_i + r \leq N. \quad (3.8)$$

$R(t_i, r)$ is calculated as follows:

$$R(t_i, r) = \max\{W(t_i, 1), \dots, W(t_i, r)\} - \min\{W(t_i, 1), \dots, W(t_i, r)\},$$

where

$$W(t_i, k) = \sum_{j=0}^{k-1} X_{t_i+j} - k \left(\frac{1}{r} \sum_{j=0}^{r-1} X_{t_i+j} \right), \quad k = 1, \dots, r.$$

Note that $R(t_i, r) \geq 0$, since $W(t_i, r) = 0$ by construction. Define the sample variance $S^2(t_i, r)$ of $X_{t_i}, \dots, X_{t_i+r-1}$ as

$$S^2(t_i, r) = \frac{1}{r} \sum_{j=0}^{r-1} X_{t_i+j}^2 - \left(\frac{1}{r} \sum_{j=0}^{r-1} X_{t_i+j} \right)^2.$$

For each value of r , we obtain a number of R/S samples. When r is small, K R/S samples can be computed, since requirement (3.8) is then satisfied for all $i = 1, \dots, K$. Owing to the same requirement, the number of samples decreases for larger values of r . Note that the resulting samples are then not independent anymore, because they partially use the same sample points. The number of R/S samples is as small as 1 when r approaches N .

For fractional Gaussian noise or fractional ARIMA, the R/S statistic is proportional to r^H as $r \rightarrow \infty$. At first sight, these formulas may look very complicated, and it is not clear why this proportionality should hold. However, the needed averages are all close to zero in our fractional Gaussian noise-case, which makes it possible to get some feeling for the method. Setting the averages to zero, $\{W(t_i, k) : k = 1, 2, \dots, r\}$ is a fractional Brownian motion sample of size r , and the range $R(t_i, r)$ is just the difference between the maximum and the minimum of that sample. The range is thus a measure of the dispersion, like the variance. Since it holds that the variance of the last sample point is proportional to r^{2H} , we tend to believe that this proportionality also holds for the range. Dividing by the square root of the variance reduces the proportionality to r^H . This reasoning is only very intuitive, since one can also argue that the range is proportional to the standard deviation instead of the variance.

The obtained samples are depicted in a log-log plot in Figure 3.6. This plot is also referred to as a Pox-plot. The values of r used for this plot are logarithmically spaced, i.e., $r_k = r_0^k$ for some $r_0 > 1$. This plot is produced with $K = 5$ and $r_0 = 10$. We are yet familiar with the idea of using only the middle part of the plot in a least-squares fitting procedure, which is again indicated by the dotted points. This yields an estimate for the Hurst parameter in our sample of 0.7163; quite small compared to the input parameter of 0.8.

Some variants of R/S analysis are described by Bassingthwaite and Raymond [6]. The authors also study the bias, accuracy and limitations of the method when applied to fractional Brownian motion samples. Unfortunately, their conclusions cannot be interpreted since they use an approximate simulation method.

3.1.8 Whittle's method

In all the methods so far some parameters had to be set arbitrarily, because the methods were based on some graphical estimation procedure. Whittle's method is based on a general approach to obtain an estimate of one or more parameters: maximum likelihood. Since we know that our sample is Gaussian with zero mean and covariance matrix Γ_H (which depends on H), the estimator is obtained by maximizing the likelihood of $X = (X_1, \dots, X_N)'$

$$L(X) = (2\pi)^{-N/2} |\Gamma_H|^{-1/2} \exp\left(-\frac{1}{2} X' \Gamma_H^{-1} X\right)$$

with respect to H . $|\Gamma_H|$ denotes the determinant of Γ_H . When this maximizing is done numerically, the inverse of Γ_H has to be computed for every H , which is very time-consuming.

However, the likelihood may be approximated, which leads to Whittle's approximate likelihood (see, e.g., Beran [9]). This likelihood converges to the exact likelihood as $N \rightarrow \infty$. The estimate for the Hurst parameter is obtained by minimizing

$$Q(H) = \int_{-\pi}^{\pi} \frac{I(\lambda)}{f_H(\lambda)} d\lambda,$$

where I denotes in the usual notation the periodogram and f_H the spectral density (the dependence on H is made explicit). Note that the evaluation of this spectral density in the fractional Gaussian noise case is computationally intensive and this density may be approximated by the Paxson approximation (see Section 1.2.2).

The minimization of Q is further sped up by approximating the integral by a Riemann sum, and evaluating the periodogram and spectral density at the Fourier frequencies $\lambda_k = 2\pi k/m$ for some m , e.g., $m = N$. Using the symmetry in I and f , we minimize

$$\tilde{Q}(H) = \sum_{k=1}^{N/2} \frac{I(\lambda_k)}{f_H(\lambda_k)}$$

to obtain the estimate \hat{H} .

Besides the asymptotical correctness of the Whittle estimator, the main advantage is that it is possible to compute an asymptotic confidence interval. It can namely be shown that

$$\sqrt{N}(\hat{H} - H) \rightarrow Z \sqrt{2 \left[\frac{1}{2\pi} \int_{-\pi}^{\pi} \left(\frac{d}{dH} \log f_H(\lambda) \right)^2 d\lambda \right]^{-1}}, \quad N \rightarrow \infty,$$

where Z is a standard normal random variable and the convergence is to be read as convergence in distribution. In practice, the integration and differentiation are done numerically to find the confidence interval.

For our sample, we give as a starting parameter in the minimization procedure $H = 0.5$, whereafter an estimate of 0.8005 for the Hurst parameter is obtained; very close to the real value. The corresponding confidence interval is [0.7903, 0.8108].

When the spectral density is replaced by $C|\lambda|^{1-2H}$, the method is called the local Whittle method. More on the local Whittle method including references can be found in [54].

Note that the Whittle method calculates the estimate of the Hurst parameter *using the spectral density*, i.e., under the assumption that the sample is fractional Gaussian noise. Since we want to analyze *approximate* fractional Gaussian noise samples, the Whittle method is not perfectly suited for this analysis. However, it is possible to test whether the periodogram of a given sample may be generated by fractional Gaussian noise. We will present this so-called goodness-of-fit test in the next section.

3.1.9 Wavelet method

In Section 2.2.5, a wavelet method to simulate fractional Brownian motion or fractional Gaussian noise was presented. This method was based on the generation of the wavelet coefficients and the key idea was to transform these back to obtain the requested fractional Brownian motion sample. Wavelets are also useful for estimation purposes, and the resulting wavelet estimator has some very nice properties. The wavelet estimator is introduced by Abry *et al.* [1].

Wavelet-based estimation is closely related to the periodogram method. The data are transformed and an estimate is obtained by estimation on the transformed data (in the periodogram method, the estimation is done in the frequency-domain). The transformation of the given sample is not as trivial as it sounds; we first estimate the sequence $(a_{B_H}(\mathcal{J}, k))_k$ for some high resolution (i.e., $-\mathcal{J}$ is big) from the fractional Brownian motion sample. The sequences $(d_{B_H}(j, k))_k$ and $((a_{B_H}(j, k))_k$ can be computed from this sequence for $j = \mathcal{J} + 1, \dots, 0$ using standard wavelet recursion techniques.

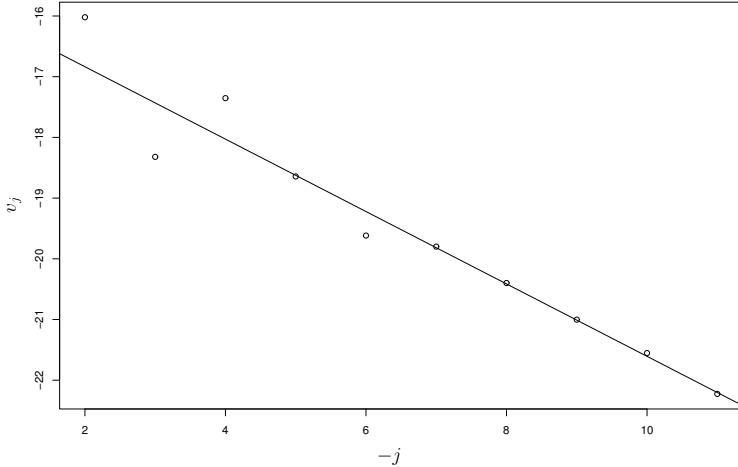


Figure 3.7: Wavelet based estimation.

It remains to choose the level \mathcal{J} . When the sample size is N (assumed to be a power of 2), the finest possible level \mathcal{J} satisfies $N = 2^{-\mathcal{J}}$. The sequence $(a_{B_H}(\mathcal{J}, k))_k$ is then estimated by

$$\hat{a}_{B_H}(\mathcal{J}, k) = 2^{\mathcal{J}/2} \sum_i \phi(i) B_H(i + k).$$

More details and an error analysis can be found in [21].

Recall that in Section 2.2.5, it was pointed out that the wavelet coefficients $d_{B_H}(j, k)$ have zero means and variance $\sigma^2 2^{j(2H+1)}$. Therefore, we can perform a wavelet decomposition of the sample, and estimate the variance of the wavelet coefficients by using all available coefficients $d_{B_H}(j, \cdot)$ at resolution j . Since it is clear from the analysis by Flandrin [29] that (although the coefficients $d_{B_H}(j, l)$ and $d_{B_H}(j, k)$ are correlated when $l \neq k$) no long-range dependence is present in the wavelet coefficients for popular wavelets, the variance of the wavelet coefficients is estimated by

$$v_j = 2^j \sum_{k=0}^{2^{-j}-1} (d_x(j, k))^2.$$

Note that this estimator is in general a bad estimator for the variance because there is no independence, but for large 2^{-j} (see the discussion in Section 1.1.2), this estimator estimates the variance quite well, up to a constant. However, the wavelet estimate of the Hurst parameter is not found by computing v_j for the fractional Brownian motion sample and making a least squares fit of $\log(v_j)$ on j (note that the potentially added constant is then irrelevant). Instead of this, the wavelet estimator proposed by Abry and Veitch [3] is found by performing the same procedure on the wavelet coefficients of the corresponding fractional Gaussian noise sample. The slope of the fitted line is then not equal to $2H + 1$ anymore, but equals $2H - 1$.

The cited authors have also improved this method, using the expressions they found for $\mathbb{E}(\epsilon_j)$ and $\text{Var}(\epsilon_j)$, with

$$\epsilon_j = \log_2(v_j) - (2H - 1)j - 2\log_2(\sigma).$$

A new estimator of H is then obtained from a linear regression of $\{\log_2(v_j) - \mathbb{E}(\epsilon_j)\}_{j_1 \leq j \leq j_2}$ on $\{j\}_{j_1 \leq j \leq j_2}$, weighted by $\{1/\text{Var}(\epsilon_j)\}_{j_1 \leq j \leq j_2}$. A confidence interval for the estimate then follows from regression theory.

This method is implemented by Coeurjolly [15] in the S-language, and makes use of the Wavetresh wavelet S-library partly implemented in C. For our sample, H is estimated as 0.7982 with 95% confidence interval [0.7769, 0.8194], when $j_1 = -11$ and $j_2 = -2$. The corresponding plot is given in Figure 3.7. Note that $-j$ instead of j is plotted against v_j .

As mentioned in [3], the wavelet estimator \tilde{H} is asymptotically unbiased when the number of vanishing moments of the wavelet is bigger than $H - 1$, and in practice has very low bias even for short data sets. In addition, the estimator is efficient (has asymptotically the smallest possible variance).

Another big advantage of the method is that potential trends have no influence on the estimator, as long as the number of vanishing moments of the wavelet \mathcal{N} is big enough to get rid of the trend. On the contrary, Whittle's method and the other above mentioned methods do not have this property. For linear trends, a wavelet with $\mathcal{N} = 2$ is powerful enough to exclude the influence of the trend on the estimation. When more complicated trends may be present, \mathcal{N} should be increased. However, this convenient property is of no importance to us, because we will only analyze generated data without trends (strictly speaking, a series with a trend is not stationary anymore, and we only defined long-range dependence for stationary processes).

3.2 Testing

Except for a few estimators, the estimation methods described in the previous section can be used to estimate the Hurst parameter of general long-range dependent samples. In this section on testing techniques, we focus on fractional Gaussian noise. It is possible to test statistically if our sample can be distinguished from fractional Gaussian noise. One could think that this is precisely what we need to evaluate simulation methods of fractional Gaussian noise and fractional Brownian motion, and thus question why all above mentioned estimation methods are needed for our analysis. Before we discuss this, we need to know more about the hypotheses of the tests.

We will study two tests. The null hypothesis is in one case that the spectral density is given by (1.11), in the other case that the covariances of the sample are in accordance with fractional Gaussian noise for some specified Hurst parameter; in both cases, the alternative is that the null hypothesis does not hold. This is the first drawback, because a test where the null and alternative hypothesis are swapped is much more powerful. We would then have statistical evidence against the hypothesis that the sample is not fractional Gaussian noise.

Moreover, it is unimaginable to summarize the sample in one number (the test statistic or *p*-value) and not to lose information on how it was produced. Therefore, it is still useful to have a look at the other methods that estimate the long-range dependence parameter H . By doing this, it is easier to find out what the particular weaknesses of the simulation method are; every estimation method is namely based on some property that a ‘proper’ fractional Gaussian noise (or fractional Brownian motion) sample should have. In other words, we get some feeling for the approximate methods when we see which properties are incorporated in the sample and which are not.

When two approximate methods are performing quite well, we cannot value the two methods by means of the test statistic or its *p*-value of only one sample, because this value is subject to coincidence. This is an additional problem in the use of these tests, which makes it necessary to extract more information from the sample. In fact, we can also exploit the possibility that we can generate as many samples as we like. The conclusions of the tests are then more robust.

3.2.1 A goodness-of-fit test for the spectral density

Beran [8] proposed a goodness-of-fit test statistic based on the spectral properties in the sample. In fact, it is tested whether the sample spectral density (i.e., the periodogram) could be generated by a process with some specified spectral density (which is in our case the spectral density of fractional Gaussian noise with the true H).

However, some problems are involved when this is tested. The observed frequencies in the sample are compared to the expected spectral information contained in f , but nothing is said about the distribution of these frequencies in time. Therefore, it is possible that a non-stationary series has spectral properties that force the test to accept the null hypothesis.

Moreover, the comparison of spectral densities involves only the correlation structure. Recall that the spectral density is defined as $f_H(\lambda) = \sum_{j=-\infty}^{\infty} \gamma(j) \exp(ij\lambda)$, where $\gamma(\cdot)$ is the autocovariance function (note that the dependence on H is again made explicit in the notation). Therefore, it is possible to have a non-centered process that has exactly the same spectral density as fractional Gaussian noise. In addition, some non-Gaussian process may also have the same spectral density. However, this is no problem in our setup, since it is clear from the structure of all approximate generation methods under consideration that a Gaussian zero-mean sample is produced.

The goodness-of-fit statistic is defined as

$$T_N(H) = \frac{A_N(H)}{B_N(H)^2}, \quad (3.9)$$

where

$$A_N(H) = \frac{4\pi}{N} \sum_{k=1}^{N/2} \left\{ \frac{I(\lambda_k)}{f_H(\lambda_k)} \right\}^2 \quad \text{and} \quad B_N(H) = \left\{ \frac{4\pi}{N} \sum_{k=1}^{N/2} \frac{I(\lambda_k)}{f_H(\lambda_k)} \right\}^2.$$

Here, $I(\cdot)$ denotes the periodogram (see (3.7)) of the observed data and $f_H(\cdot)$ the spectral density of fractional Gaussian noise with the real H . Both are evaluated at the Fourier frequencies $\lambda_k = 2\pi k/N$.

It is shown in [8] that, under the null hypothesis, $\sqrt{N}(A_N(H) - \mu_1, B_N(H) - \mu_2)$ converges in distribution to a bivariate Gaussian random variable with mean 0 and covariance matrix Σ as $N \rightarrow \infty$ for some known μ_1, μ_2 and Σ . Using the so-called Delta method, the author deduces that $T_N(H)$ is approximately normally

distributed with mean π^{-1} and variance $2\pi^{-2}N^{-1}$. It is claimed that the approximation of $T_N(H)$ by such a Gaussian random variable is already good for moderately large sample sizes.

From a practical point of view, one of the strengths of this goodness-of-fit test is that the asymptotic distribution of the test statistic is not affected when the Whittle estimated Hurst parameter is substituted in (3.9), see again [8]. However, this is of no importance to us, since we are interested whether an approximate sample can be distinguished from fractional Gaussian noise with the *input* Hurst parameter.

Beran has also implemented this test, and made the code publicly available. For our sample, the test statistic is 0.3213, which corresponds to an asymptotic p -value of 0.3916 (i.e., in this case $2P(T_N(H) > 0.3213) = 0.3916$, since the test is two-sided). This means that the null hypothesis of a spectral density with Hurst parameter $H = 0.8$ is accepted at any reasonable confidence level. Recall that the p -value has to be compared with a given confidence level, e.g., 5%.

However, testing whether this sample is fractional Gaussian noise with Hurst parameter 0.76 leads to a p -value of 0.1158, still accepting the null hypothesis even at a 10% confidence level. This makes us slightly suspicious about the power of this test.

3.2.2 A chi-square test for fractional Gaussian noise

As indicated in the discussion of the goodness-of-fit test for the spectral density, it is desirable to test whether the finite-dimensional distributions are in accordance with fractional Gaussian noise. Although several goodness-of-fit tests exist to test the marginal distribution [9, Sec. 10.1], we restrict ourselves to a chi-square test for the joint distribution. Under the null hypothesis, the sample has zero mean and its auto-covariance function $\gamma(\cdot)$ is given by (1.7). Recall from Chapter 2 that the covariance matrix Γ determined by $\gamma(\cdot)$ has a Cholesky decomposition $\Gamma = LL'$. Denoting again the sample size by N , we can ‘standardize’ a fractional Gaussian noise sample $X = (X_0, \dots, X_{N-1})'$ by computing $Z = L^{-1}X$, such that Z has independent standard normal elements. A chi-square test can be based on this fact.

Letting $Z = (Z_0, \dots, Z_{N-1})$, the test statistic is

$$C_N(H) = \sum_{k=0}^{N-1} Z_k^2, \quad (3.10)$$

which is chi-square distributed with N degrees of freedom under the null hypothesis. Again, the dependence on H is emphasized in the notation (note that Γ and therefore L also depend on H). For a given confidence level α , the hypothesis is rejected when $C_N(H)$ is larger than the upper α -quantile of the χ_N^2 distribution.

Unlike the goodness-of-fit test, it is (to our knowledge) not known if the asymptotic distribution of the test statistic is affected when Z is computed by substituting a consistent estimator \tilde{H} for H : more research has to be done to see if

$$\frac{C_N(\tilde{H}) - N}{\sqrt{2N}}$$

converges in distribution to a standard normal variable as $N \rightarrow \infty$.

The power of the test against alternatives can also be studied in more detail. It is a desirable property of a test that the distribution of the test statistic under alternatives is very different from the distribution under the null hypothesis. The null hypothesis is then rejected with a relatively high probability when the alternative is true. This is called the power of the test. In other words, it is desirable that the distribution of $C_N(H)$ ‘differs’ from χ_N^2 when the elements of Z do not have unit variance and are not independent.

A serious problem of the test is that the matrix L should be computed and be kept in memory to make it possible to invert the matrix. Because the inversion can be done quite fast using backward-substitution, the memory problem is more serious. For instance, it was not possible to obtain a test value for our sample, because it was not possible to allocate the necessary memory space. Without giving the details, the test statistic can also be calculated once the eigenvalues of L are known. As the computation of these eigenvalues is very time-consuming due to the size of L , this is no good option either.

It is possible to solve the memory problem, although the test loses power against specific types of alternatives. Assume that no memory problems are involved when the sample size is K . This K should be chosen as large as possible, and we assume that both K and the sample size $N > K$ are powers of two. We then split our sample in N/K blocks, and compute the test statistic for each block. Under the null hypothesis, the sum of all obtained test statistics over the K blocks has again a χ_N^2 distribution, which is easy to see. By following this procedure, some of the correlations in the sample are not taken into account in the test. We should keep in mind that this has no influence when these correlations are zero, but that the modified test statistic gives the same result when these correlations are non-zero. The modified test yields a p -value of 0.4439 for our sample ($K = 2^9$).

Since we are able to simulate approximate samples, we can also simulate the distribution of the test statistic using a certain approximate simulation method. It is possible to test if this distribution equals the χ^2_N (null)distribution, which provides a means to draw more robust conclusions than a chi-square test on only one simulated sample. The well-known one-sample Kolmogorov-Smirnov test (see, e.g., Gibbons [30]) is the technique that can be used to test if a sample is drawn from a specified distribution.

CHAPTER 4

Evaluation of approximate simulation methods

When proposing a new simulation method for fractional Brownian motion, the inventors of the method relate their method to existing ones. Besides the speed, an important aspect of the method is whether the resulting samples are long-range dependent with the right Hurst parameter. This is checked by estimating the Hurst parameter H , mostly with the Whittle method, and comparing these estimates to the input value of the sample. To the author's knowledge, no articles exists in which attention is paid to the impact of the estimation procedure; the estimates may be biased. Although we will shortly see that simulation results suggest that the Whittle estimator is (asymptotically) unbiased, it should first be checked whether the sample corresponds indeed to fractional Gaussian noise, in particular its spectral properties (see Chapter 3).

When an approximate sample incorporates the right long-range dependent behavior, there is still no reason to assume that the sample is a realization of fractional Gaussian noise or fractional Brownian motion. For instance, we also need that the mean is zero and that the sample is Gaussian. In fact, we know precisely what else is needed: fractional Brownian motion is characterized by the properties on page 6.

We start this chapter with a study on the errors that are made by the approximate simulation methods, using the theory of Chapter 2. After this preliminary study, the structure of this chapter follows closely the properties on page 6. The stationarity of the approximate (incremental) samples is thus checked first, although the detection of stationarity is very difficult. We can regard $\mathbb{E}B_H^2(t) = t^{2H}$ as the self-similarity or long-range dependence property, which enables us to use the estimation techniques presented in Chapter 3. Before it is possible to estimate the Hurst parameter, it should be first tested if the samples cannot be distinguished from fractional Gaussian noise, as already mentioned. The requirement that the paths should be continuous is of no interest to us, since we have discrete-time (sampled) versions of the paths. This is also the case for the normality property, since all approximate simulation methods that we will study generate zero-mean Gaussian sample paths.

The evaluation of approximate simulation methods has been no subject of extensive study. Jennane *et al.* [36] review some methods and test statistically the normality of the process, the stationarity of the increments (essentially only the variance-stationarity) and the self-similarity of the resulting samples. In that respect, the structure of this chapter is closely related to their approach. However, the statistical methods to check the validity of these properties differ. As an example, they use a chi-square test to check the self-similarity property, thus assuming independence that is not present. Moreover, some promising simulation methods are not studied in their article.

Coeurjolly [15] discusses a number of interesting simulation and estimation methods, but does not evaluate these in the detail we will do.

Unless stated otherwise, all plots in this chapter are produced with a Hurst parameter $H = 0.8$.

4.1 Error analysis

The error analysis is first performed on the approximate circulant method, since the requires ideas are already discussed (see Chapter 2). Unfortunately, it is impossible to compute errors of Paxson samples, since no exact reference sample is available. Since the Paxson method and the approximate circulant method are both special cases of the spectral simulation method, we expect that the results concerning convergence are also true for the Paxson method, although it may require larger sample sizes.

After having studied the errors for the approximate circulant method, we address the errors of the RMD_{*l,r*} and the wavelet method.

4.1.1 The approximate circulant method

We first study the approximation error of fractional Gaussian noise in the spectral domain, whereafter we switch to the time domain to gain empirically more insight into the theoretical considerations of Section 2.2.4. The error of fractional Brownian motion samples is also analyzed, and the approximation is also studied in a simple network model.

The error in the spectral domain

In the spectral domain, the approximation error that is made by the approximate circulant method is best analyzed by comparing the eigenvalues of the Davies and Harte method $\lambda_k = \sum_{j=-N+1}^{N-1} \gamma(j) \exp(\pi i j k / N)$ to the spectral density values $f(\pi k / N) = \sum_{j=-\infty}^{\infty} \gamma(j) \exp(\pi i j k / N)$ for $k = 1, \dots, 2N - 1$. Note that no approximation is involved for $k = 0$ by construction. The only question is how to compare these two quantities.

In the following, we denote the approximate circulant sample by $\tilde{X} = \{\tilde{X}_k : k = 0, \dots, N - 1\}$ and the exact Davies and Harte sample by $X = \{X_k : k = 0, \dots, N - 1\}$. Since we are interested in the error in the frequency domain, a natural idea is to study the sequences that serve as input in the FFT algorithm. For the approximate circulant method, this sequence is given by (2.27). To avoid unnecessary notationally complicated expressions, we do not care about the scaling factor for c_0 and c_N ; we set $c_0 = \sqrt{f(0)/(4N)}$ with $f(0) = N^{2H} - (N - 1)^{2H}$ and $c_N = \sqrt{f(\pi)/(4N)}$. Likewise, we set in the FFT-input sequence $(w_k)_k$ of the Davies and Harte method (which follows from (2.12)) $w_0 = \sqrt{\lambda_0/(4N)}$ and $w_N = \sqrt{\lambda_N/(4N)}$. The difference $(d_k)_k$ of the two sequences $(c_k)_k$ and $(w_k)_k$ provide information on the error in the spectral domain. Writing things out and setting $t_k = \pi k / N$, we get

$$d_k = \begin{cases} \left(\sqrt{\frac{\lambda_k}{4N}} - \sqrt{\frac{f(t_k)}{4N}} \right) (U_k^{(0)} + iU_k^{(1)}) & k = 0, \dots, N \\ \left(\sqrt{\frac{\lambda_{2N-k}}{4N}} - \sqrt{\frac{f(t_{2N-k})}{4N}} \right) (U_{2N-k}^{(0)} - iU_{2N-k}^{(1)}) & k = N + 1, \dots, 2N - 1 \end{cases}. \quad (4.1)$$

In the following, we feel free to change the coefficient d_N from time to time for explanatory reasons. Since the expectation of the coefficients is zero, the standard deviation (or variance) of d_k measures intuitively the error in the spectral domain. We will therefore study a scaled version of the standard deviation of d_k ,

$$\sigma_k := \left| \sqrt{\sum_{j=-\infty}^{\infty} \gamma(j) \exp(\pi i j k / N)} - \sqrt{\sum_{j=-N+1}^{N-1} \gamma(j) \exp(\pi i j k / N)} \right| \quad (4.2)$$

and the variance

$$\sigma_k^2 = \left(\sqrt{\sum_{j=-\infty}^{\infty} \gamma(j) \exp(\pi i j k / N)} - \sqrt{\sum_{j=-N+1}^{N-1} \gamma(j) \exp(\pi i j k / N)} \right)^2. \quad (4.3)$$

The variances are plotted as function of k in Figure 4.1 for $N = 2^{10}$. From this figure, it becomes clear that traces generated by the approximate circulant method differ mainly from exact traces by their low and high frequency behavior, but the range of frequencies that are affected by the approximation seems rather small.

Thus, a natural question is whether the spectral performance of the approximate circulant method becomes better as the sample size N increases, because we may think that the range of affected frequencies becomes smaller. To avoid problems of defining when a frequency is affected, we study the mean standard deviation $\frac{1}{2N} \sum_{k=0}^{2N-1} \sigma_k$ and the mean variance $\frac{1}{2N} \sum_{k=0}^{2N-1} \sigma_k^2$ instead. The errors are plotted against the (log) sample size in Figure 4.2. The solid line corresponds to the mean absolute error, and the dashed line to the mean square error. They both show a clear decay.

Although we may find this a satisfactory result, we are mainly interested in the time-domain error, i.e., the error of the output sample. Approximate time-domain traces are used as input in, e.g., queueing systems. Therefore, we leave the frequency domain and change the point of view towards the time domain.

The error of fractional Gaussian noise in the time domain

As already pointed out in Chapter 2, it is possible to compare the time-domain output of the two methods by using the same random numbers to generate the sample. This was done for two generated fractional Brownian

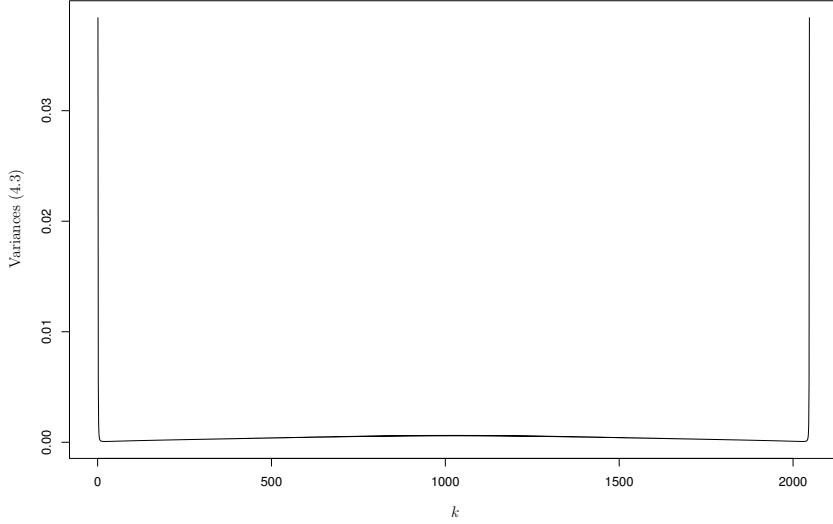


Figure 4.1: The variances σ_k^2 of the differences in the spectral domain measure the error in the spectral domain.

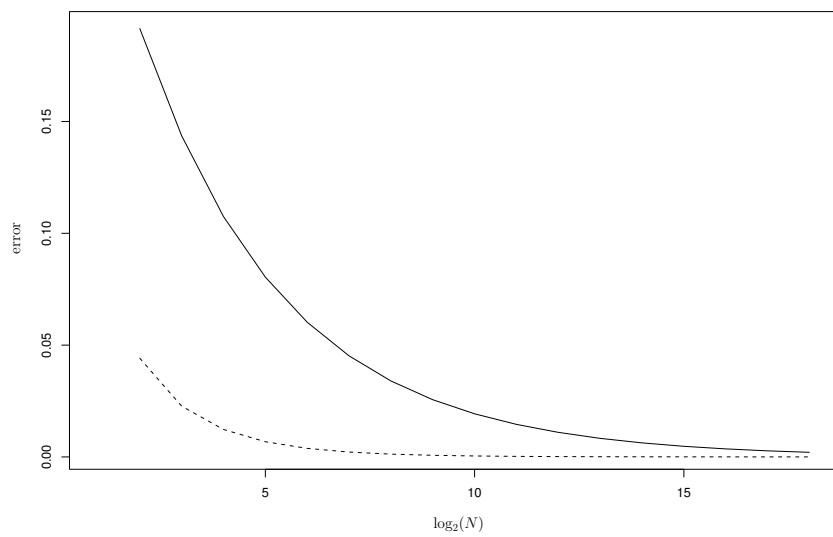


Figure 4.2: The ‘spectral’ performance measures $\frac{1}{2N} \sum_{k=0}^{2N-1} \sigma_k$ (solid line) and $\frac{1}{2N} \sum_{k=0}^{2N-1} \sigma_k^2$ (dashed line) as functions of the (\log_2) sample size.

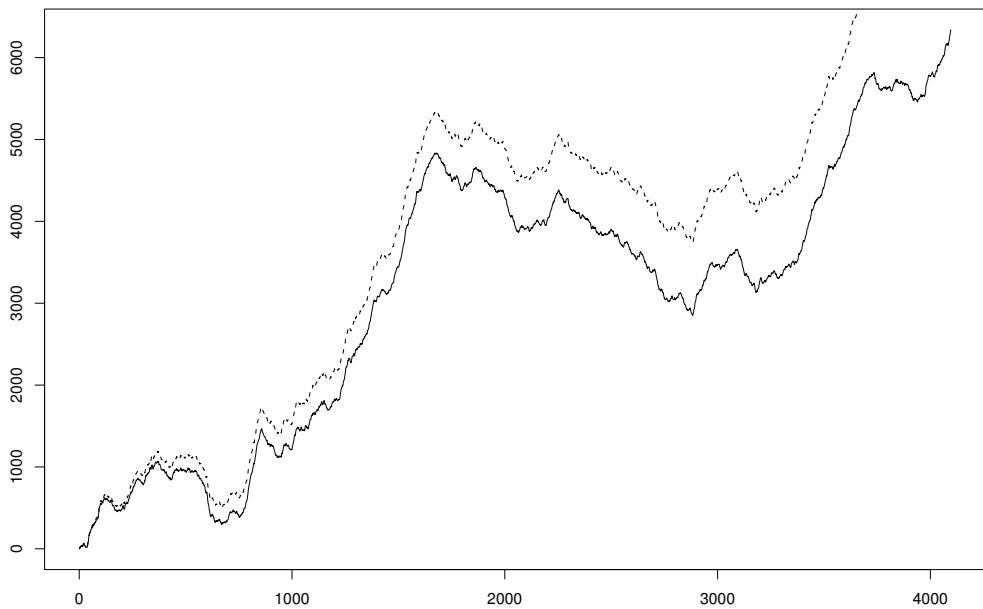


Figure 4.3: Fractional Brownian motion generated by the Davies and Harte method (solid line) and the approximate circulant method (dashed line) using the same random numbers.

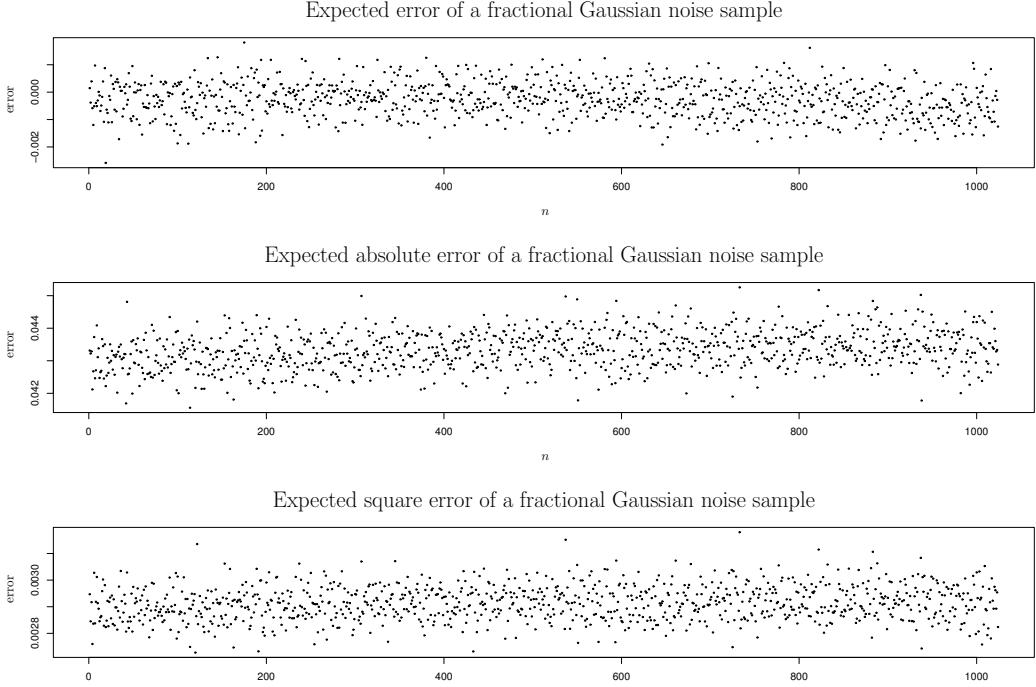


Figure 4.4: Simulation of the expected error, expected absolute error and expected square error made by the approximate circulant method for fractional Gaussian noise.

motion samples in Figure 4.3, with $N = 2^{12}$ and (as usual) $H = 0.8$. The solid line is generated by the Davies and Harte method, the other by the approximate circulant method.

The idea of using common random numbers makes it possible to analyze the (time-domain) errors, since the error of approximate circulant samples can be computed. We perform a small simulation study to analyze the expected error, the expected absolute error and the expected square error. After generating 1000 fractional Gaussian noise samples of length 2^{10} , we compute the three errors for every sample point in each sample. This results in an average for each sample point, which is an estimate for the error in that sample point. The estimates are plotted in Figure 4.4.

To be able to understand these plots, we study the random process $\check{X} - X = \{\check{X}_k - X_k : k = 0, 1, \dots, N-1\}$, which is the real part of the Fourier transform of the sequence $(d_k)_k$ defined by (4.1). It is readily checked that $\check{X}_n - X_n$ is a centered Gaussian variable with variance (not caring about the value of d_N)

$$\sigma_{\check{X}_n - X_n}^2 = \frac{1}{2N} \sum_{k=0}^{2N-1} \sigma_k^2,$$

which does not depend on n anymore. In fact, this variance was already plotted as a function of N in Figure 4.2. The plots in Figure 4.4 are completely in accordance with these considerations; the error seems to be centered and the lower panel consists of (dependent) realizations of a scaled χ_1^2 random variable. Moreover, since $\mathbb{E}|\check{X}_n - X_n|$ is proportional to the square root of $\sigma_{\check{X}_n - X_n}^2$ because of the normality, the expected absolute error also converges to zero as the sample size increases.

In Chapter 2, it was shown that an approximate circulant sample converges to an exact sample in \mathcal{L}^2 -norm. This means that the expected square error converges to zero for every sample point, or equivalently that $\sigma_{\check{X}_n - X_n}^2 \rightarrow 0$ as $N \rightarrow \infty$. Because $\sigma_{\check{X}_n - X_n}^2$ was depicted as the dashed line in Figure 4.2, we see that this theoretical result is indeed numerically confirmed.

The error of fractional Brownian motion

Now that we have both theoretically and empirically seen that the expected absolute and square error converge to zero for fractional Gaussian noise samples, it is an interesting question whether the corresponding fractional Brownian motion samples share these properties. As in Figure 4.4 for fractional Gaussian noise, we plot the mean error, mean absolute error and mean square error for fractional Brownian motion in Figure 4.5.

The shape of the expected error and expected absolute error is quite remarkable: they are perfect straight lines through the origin. This is not quite what we expected, especially in the case of the expected error: the expected error of the fractional Gaussian noise sample is zero, so this must also be the case for the cumulative errors. Let us first study the error of the fractional Brownian motion sample $\check{Y} - Y = \{\check{Y}_n - Y_n =$

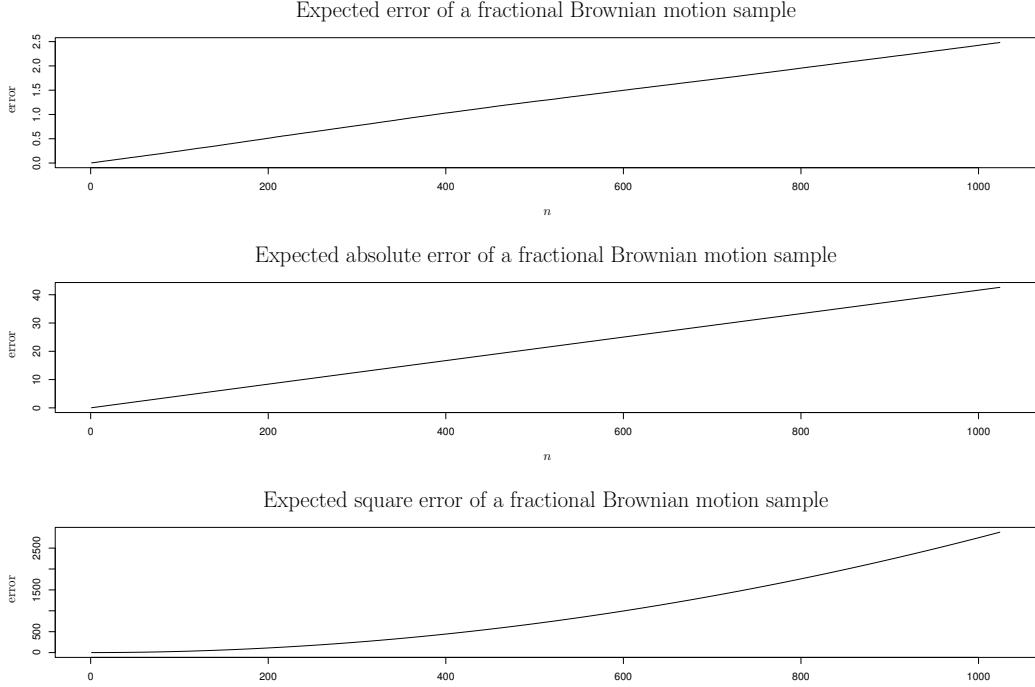


Figure 4.5: Simulation of the expected error, expected absolute error and expected square error made by the approximate circulant method for fractional Brownian motion.

$\sum_{j=0}^n (\check{X}_j - X_j) : n = 0, 1, \dots, N-1\}$. Obviously, this is a centered Gaussian process. It is left to the reader to check that its variance depends on n and is given by

$$\sigma_{Y_n - Y_n}^2 = \frac{1}{2N} \sum_{k=0}^{2N-1} \sigma_k^2 \left[\left(\sum_{j=0}^n \cos^2(\pi j k / N) \right)^2 + \left(\sum_{j=0}^n \sin^2(\pi j k / N) \right)^2 \right].$$

We know from Figure 4.1 that σ_k^2 is almost everywhere zero but for values of k near 0 and $2N-1$. For these values of k , the term between the square brackets is approximately n^2 , since the sine-term vanishes and the cosine-term is very close to 1. The expected square error of the fractional Brownian motion sample $\sigma_{Y_n - Y_n}^2$ is thus proportional to n^2 for large sample sizes, which is precisely what we see in Figure 4.5. The normality accounts for the proportionality to n of the expected absolute error. However, the expected error itself should approximately be zero, whereas it looks like a straight line in Figure 4.5.

The intuitive explanation for the observed phenomenon lies in the fact that the approximate circulant method differs in the low and high frequency behavior from the exact Davies and Harte method. First note that the scale is different in the two upper plots of Figure 4.5, and considerably smaller for the expected error. The exact Davies and Harte sample has a low frequency component that is not incorporated in the corresponding approximate circulant sample. When simulating errors in a fractional Brownian motion sample, a big wave shows typically up. Depending on the realizations of these waves, averaging only 1000 errors (waves) may very well yield a small ‘non-zero wave’, although the ‘expectation’ of that wave is zero. This is precisely what we observe in the upper panel of Figure 4.5.

The intuitive reasoning in terms of waves also explains why the averaged absolute error starts at zero and why it increases. However, it fails to explain why this results in a straight line.

Although the low frequency deviation explains the shape of the graphs, it is still not clear why it looks like there are no high frequencies present in the errors. Recall that low as well as high frequencies are affected by the approximation by the approximate circulant method. In fact, high frequencies are present in all plots of Figure 4.5, but they can not be observed, because this effect is overshadowed by the effect of the low frequencies.

The relative error in a network model

Since the absolute and square error for fractional Brownian motion samples increase obviously in n , the reader might think that this approximate simulation method is of no use to simulate a network traffic process. However, we can not be sure of that at this point. First recall the network traffic process A of Section 1.3. It is a translated version of fractional Brownian motion, where $A(t)$ represented the *cumulative* traffic in the

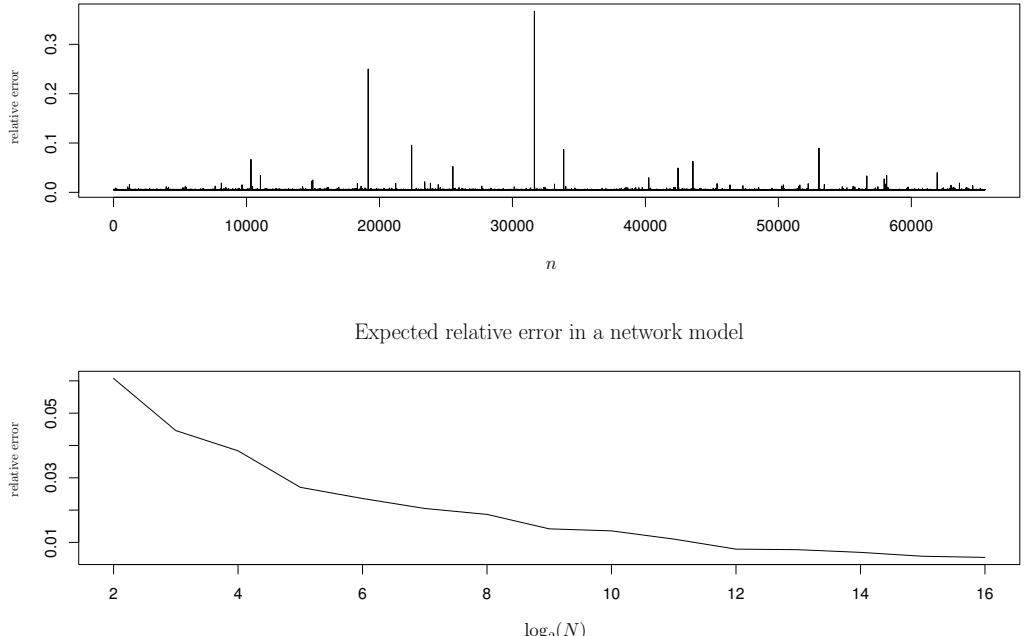


Figure 4.6: The relative error of a typical network trace and the mean relative error as a function of the sample size ($H = 0.8$).

time interval $[0, t]$. The quantity of interest when the network traffic in a time interval $[t, t + T]$ is simulated is thus the absolute (or square) error of the generated value relative to the ‘exact’ value. The theoretical result of Chapter 2 is of no help in this case, so the only remaining possibility is to simulate this relative error.

This relative error is mathematically described by

$$\left| \frac{\sqrt{aM}(X_n - \bar{X}_n)}{M + \sqrt{aM}X_n} \right|$$

for every sample point $n = 0, \dots, N - 1$. It is interesting how this error behaves as the sample size N varies. We will not study this error in detail, but we set $M = 1000$ and $\sqrt{aM} = 250$ and carry out a small simulation experiment to get some feeling for the relative error. We proceed in the same way as before: for a range of sample sizes, we generate 100 traces and compute the relative error in each sample point. Since we hope that the relative error also decreases in the sample size for every sample point, a natural thought is to show graphically that the maximum over the sample points decreases to zero. However, this is not desirable in this case. One realization of X_n of approximately -4 produces a big relative error, which has a huge impact on the maximum. Therefore, we simulate the relative error averaged over all sample points for several values of the sample size N .

The resulting plot is given in the lower panel of Figure 4.6. The upper panel consists of a realization of the relative error for $N = 2^{16}$. As expected, some spikes show up in this plot. We conclude from the lower plot that the approximate circulant method performs better for larger sample sizes, although the result depends clearly on the choice of the parameters M and \sqrt{aM} .

Thinking about approximate samples as input in queueing systems, it is interesting to simulate the errors in the queueing system of interest instead of this simple network model. The performance of the approximate method can then be compared to the Davies and Harte method on some selected measures (e.g., overflow probability). This is especially important when so-called rare events are simulated; rare events are events that occur with a very small probability.

We now perform an error analysis on the $\text{RMD}_{l,r}$ method and compare the results to the observations for the approximate circulant method.

4.1.2 The $\text{RMD}_{l,r}$ method

Unlike the approximate circulant method, no exact reference trace is available for the $\text{RMD}_{l,r}$ method, which makes it impossible to compute and analyze errors. To overcome this, samples produced with relatively small l and r (as typically used in practice) are compared to reference samples with $l = 50$ and $r = 25$, using common random numbers. This trace is then regarded as an exact trace. This approach is also followed in

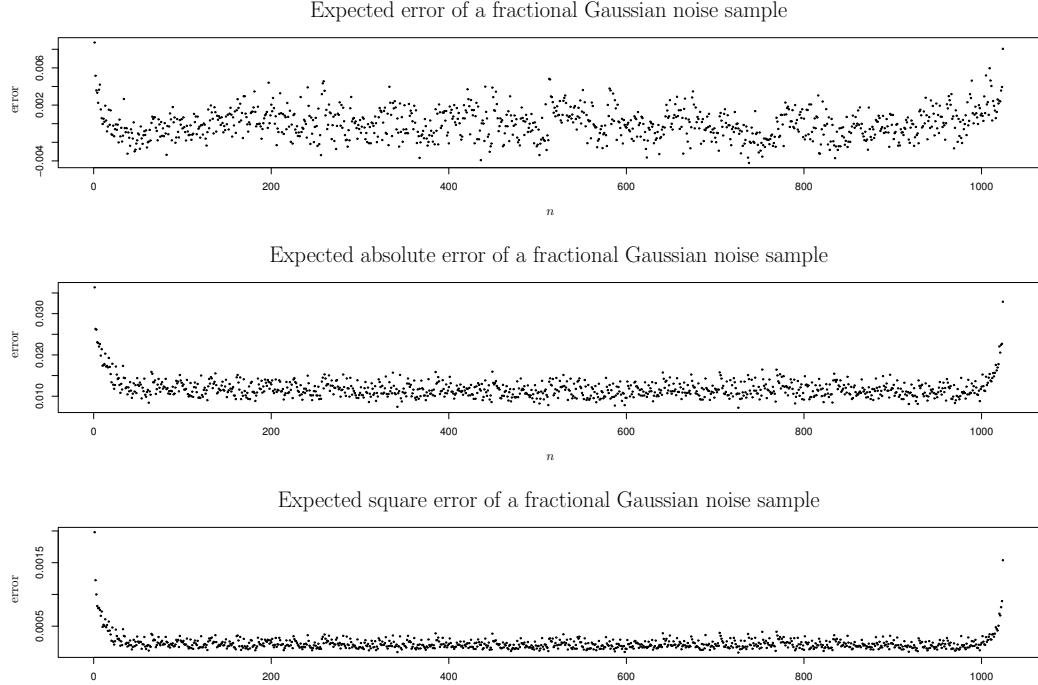


Figure 4.7: Simulation of the expected error and the expected absolute error made by the $\text{RMD}_{1,2}$ method for fractional Gaussian noise.

[46], where the relative error

$$\frac{\sum_{k=0}^{N-1} |X_k - \hat{X}_k|}{\sum_{k=0}^{N-1} |X_k|}$$

is studied, where X denotes the ‘quasi-exact’ $\text{RMD}_{50,25}$ sample and \hat{X} the approximate sample, e.g., $\text{RMD}_{1,2}$.

Instead of this relative error, we focus on the same errors as we did for the approximate circulant method to be able to compare the $\text{RMD}_{l,r}$ errors to the approximate circulant errors. Of course, it makes only sense to study the error in the time domain. Since $l = 1$ and $r = 2$ are claimed to be good choices for the parameters in [46], we will focus the error analysis on $\text{RMD}_{1,2}$ traces. Recall that an $\text{RMD}_{l,r}$ sample is found by simulating the processes $\mathcal{X}_0, \mathcal{X}_1, \dots$, which corresponds to simulating fractional Brownian motion of some coarse equispaced grid and doubling the number of points on the grid (i.e., ‘stochastically interpolating’) in each successive step. This process is stopped when the sample has the required ‘resolution’.

The errors of fractional Gaussian noise and fractional Brownian motion

Plots of the mean error, the mean absolute error and the mean square error of a fractional Gaussian noise sample are given in Figure 4.7. The same quantities are plotted in Figure 4.8 for the corresponding fractional Brownian motion samples. The plots are based on a simulation of only 100 runs, because of the computational effort that is required to generate samples with the $\text{RMD}_{50,25}$ method.

First of all, the error of an $\text{RMD}_{1,2}$ -fractional Gaussian noise sample has zero mean like the approximate circulant method, but the mean absolute and square errors are considerably lower than for a sample generated with the approximate circulant method (cf. Figure 4.4). This is a clear advantage of the $\text{RMD}_{1,2}$ method. It seems, however, that in the first and last part of the sample the expected absolute and square errors are larger, which must be caused by border effects: the number of points that are used in the conditioning decreases when a border is near. Because this is true for every level \mathcal{X}_i , this affects more points than only the first l points near the left border and the last r points near the right border.

More importantly, the errors made in the fractional Gaussian noise sample cancel out each other when cumulative sums are taken, as seen in the two lower panels of Figure 4.8: the absolute error of the fractional Brownian motion sample is non-monotone. This is in contrast to the approximate circulant method, for which there are typically periods in which all errors have the same sign, resulting in a ‘wave’ in the accumulated error, although the individual errors in the fractional Gaussian noise sample are small.

Another interesting property of the plot of the expected absolute error for a fractional Brownian motion sample is that there are roots in $n = 1, n = 1024, n = 2048$ and $n = 4096$. This can be explained by recalling how the sample was constructed. For $l = 1$ and $r = 2$, $X_1 + \dots + X_{2048}$ and $X_{2049} + \dots + X_{4096}$ are simulated in the first iteration, and no approximation is involved. In the second iteration, $X_1 + \dots + X_{1024}$ is simulated

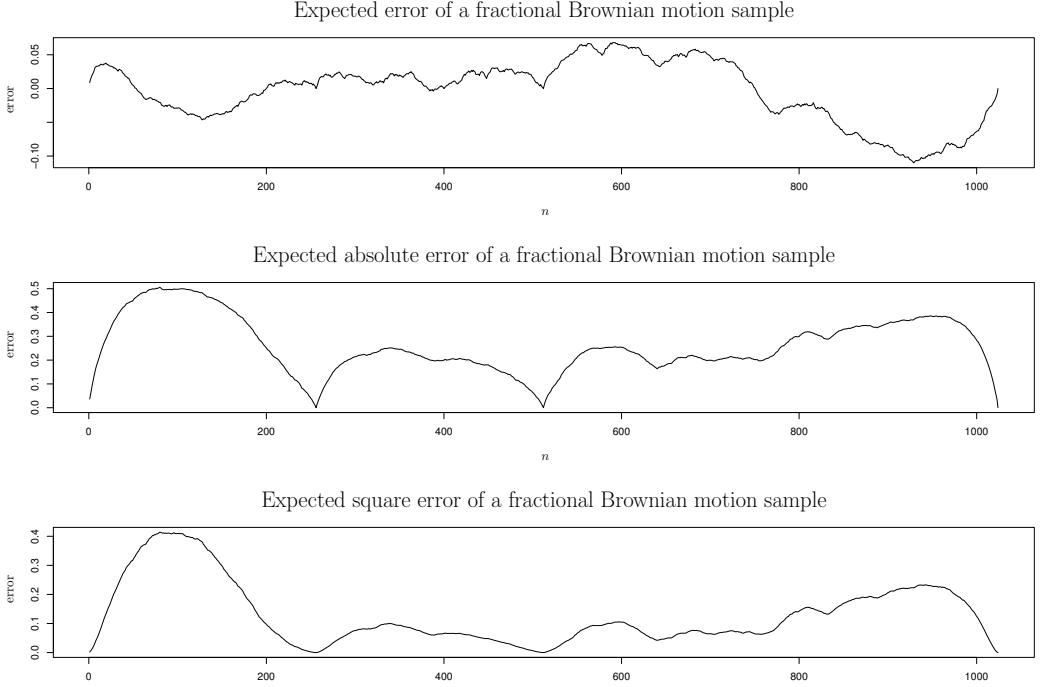


Figure 4.8: Simulation of the expected error and the \sqrt{n} expected absolute error made by the $RMD_{1,2}$ method for fractional Brownian motion.

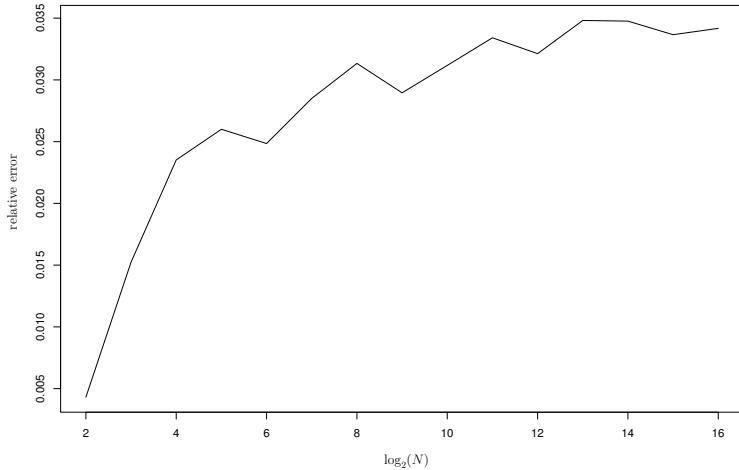


Figure 4.9: The mean relative error as a function of the (\log_2) sample size for $RMD_{1,2}$ traces.

given $X_1 + \dots + X_{2048}$ and $X_{2049} + \dots + X_{4096}$; again exact. Also, $X_{2049} + \dots + X_{3072}$ is simulated given $X_{1025} + \dots + X_{2048}$ and $X_{2049} + \dots + X_{4096}$. This is not exact anymore, because there should be conditioned on $X_1 + \dots + X_{1024}$ as well in an exact algorithm (note that $l = 1!$).

It seems that the first and last ‘mountain’ in the second and third plot of Figure 4.8 are a bit higher than the others. This is in line with the border effects that we observed in Figure 4.7, suggesting that the approximation can be improved by removing some begin and end points of the sample.

From the explanation of the mountain-shape becomes clear that the size of the mountains can be reduced by choosing l and r larger, which is intuitively clear. Choosing $RMD_{\infty,\infty}$ as an (exact) reference instead of $RMD_{50,25}$ results in the same shape for the absolute error, up to small ‘hops’. Furthermore, we expect that a doubling of the sample size has no influence on the shape of the expected absolute error, since the new sample points are added between the existing ones.

The relative error in a network model

This should also be observed in a plot of the relative error in the same fractional Brownian traffic model as in the previous section. Based on only 10 runs, the relative error is depicted as a function of the sample size in Figure 4.9. It shows that the relative error indeed does not improve for larger sample sizes and increases even for small sample sizes. The latter is easily explained, e.g., for $N = 2^2$. Only four points should be generated

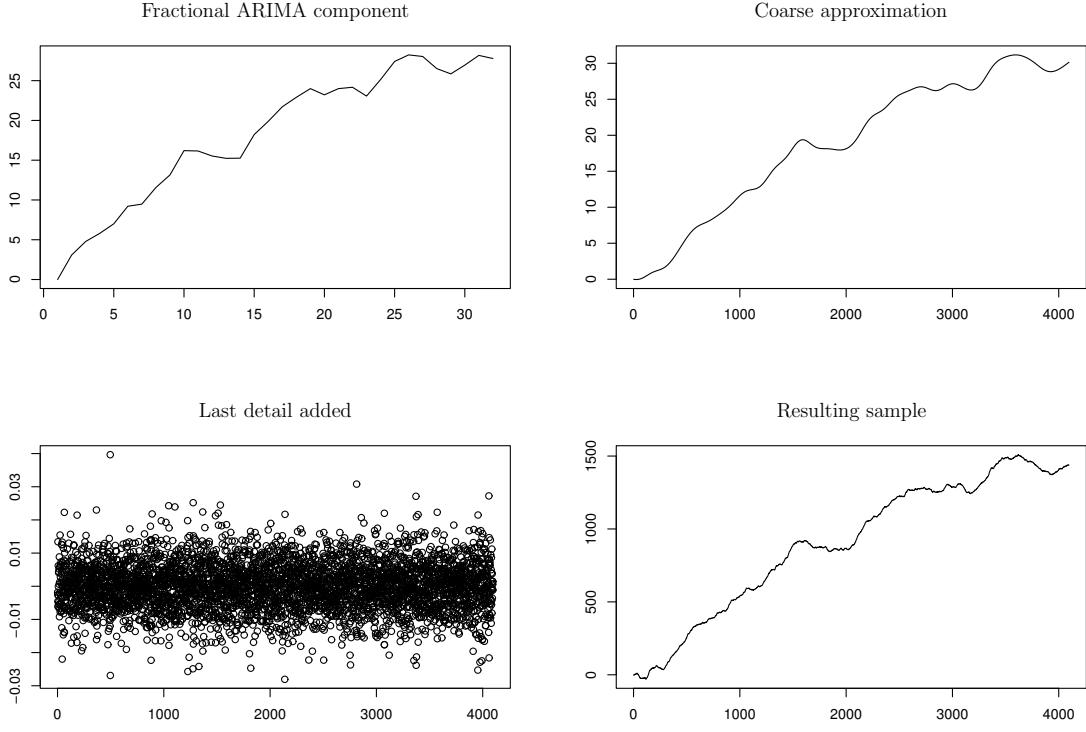


Figure 4.10: The construction of a fractional Brownian motion with the wavelet method.

while conditioning on maximal two points on the left and one point on the right. This yields a very small (relative) error.

Note that it becomes also clear at this point that it depends highly on the application which approximate method performs better; the absolute error of the RMD_{1,2} method in Figure 4.8 is clearly smaller than the absolute error of the approximate circulant method, but the relative error in the network model is smaller for the approximate circulant method when reasonable sample sizes are required (compare Figure 4.6 and Figure 4.9)!

4.1.3 The wavelet method

As mentioned in Chapter 2, it has been proposed to use the wavelet method with the truncation parameter $-J$ such that $J = 5$ or $J = 6$. A reference trace is produced with J_f determined by $N = 2^{J_f}$. Although it is in theory possible to simulate the details when going from level J to J_f , this is problematic in practice for computational reasons; as we will shortly see, the method is already very slow for $J = 6$. Moreover, the reference trace is not exact anyway, since the truncation is not the only approximation involved; recall that the fractional ARIMA component was also approximated.

However, we can gain some more insight into the mechanism of the wavelet method by generating some plots that clarify the successive steps in the algorithm. In Figure 4.10, the ‘building blocks’ of a fractional Brownian motion sample with $N = 2^{12}$ and $J = 7$ are depicted. The Daubechies 10 wavelet is used in the algorithm. The program to generate fractional Brownian motion samples with wavelets was implemented in the C-language based on the S-program by Coeurjolly, see [15].

The plots in Figure 4.10 are best understood with the help of the truncated form of Equation (2.33). Since this formula concerns the simulation of a sample of size N on $\{0, 2^{-J}, \dots, (N-1)2^{-J}\}$, the self-similarity property is used to find the wavelet sample $(\tilde{B}_H(n))_{n=0}^{N-1}$. This sample is given by

$$\tilde{B}_H(n) = 2^{HJ} \left(\sum_k b_H(k) \phi_{0,k}^{(s)}(2^{-J}n) + \sum_{j=-J}^0 \sum_k \lambda_{j,k} 4^{-s} 2^{js} \psi_{j,k}^{(s)}(2^{-J}n) \right). \quad (4.4)$$

In the upper left corner, the fractional ARIMA component $b_H(k)$ is plotted (note that the length of this sample does not equal N , but $2^{-J}N$, cf. the discussion in Section 2.2.5). The upper right corner shows the (scaled) output sample when no details are added, i.e., $\sum_k b_H(k) \phi_{0,k}^{(s)}(2^{-J}n)$. The last added detail $\sum_k \lambda_{-J,k} 4^{-s} 2^{-Js} \psi_{-J,k}^{(s)}(2^{-J}n)$ is plotted in the lower left corner of Figure 4.10, whereas the lower right corner consists of the resulting sample $(\tilde{B}_H(n))_{n=0}^{N-1}$.

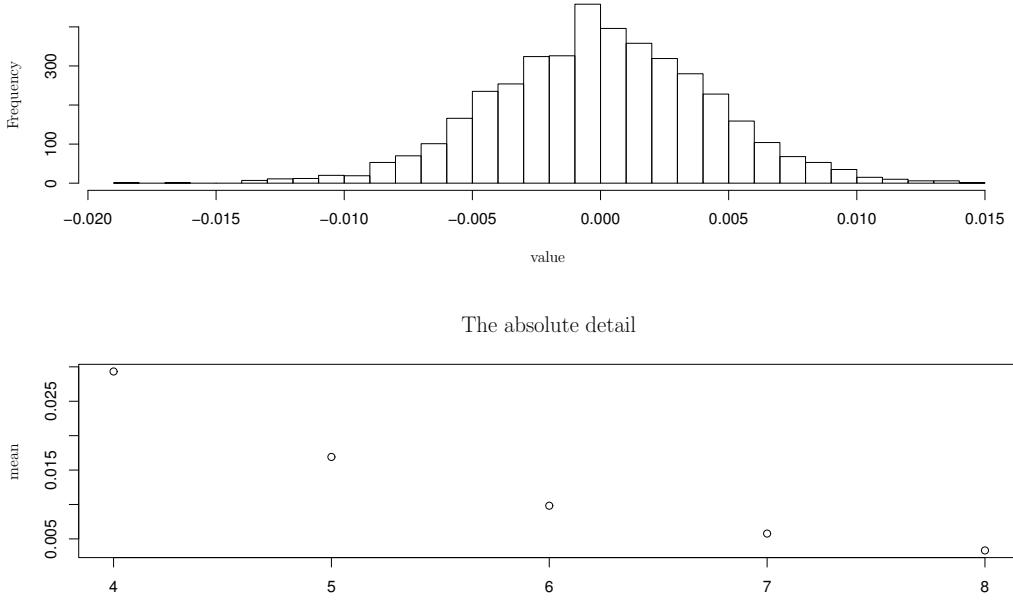
Distribution of the detail for $-j = 8$ 

Figure 4.11: The details studied in more detail.

Let us assume that the approximation in $b_H(k)$ is negligible. It is then natural to study the details at the subsequent approximation levels. Intuitively, the details add higher and higher frequencies to the sample. Although setting the details to zero except the first $J + 1$ details leads to a loss in quality of the sample, it is very well possible that the influence of the omitted details is very small.

There are many ways to measure this influence, but we will study the behavior of the mean of the absolute value of the details. The reason for this is that the detail $\sum_k \lambda_{-J-1,k} 4^{-s} 2^{-(J+1)s} \psi_{-J-1,k}^{(s)}(2^{-J}n)$ can be regarded as the most important component of the error

$$\sum_{j=-J_f}^{-J-1} \sum_k \lambda_{j,k} 4^{-s} 2^{js} \psi_{j,k}^{(s)}(2^{-J}n). \quad (4.5)$$

Recall that we assumed that this is the only error made in the approximation, and that it is computational too intensive to compute the sum (4.5) for reasonable sample sizes on a 200 MHz Pentium machine. We use the detail $\sum_k \lambda_{-J-1,k} 4^{-s} 2^{-(J+1)s} \psi_{-J-1,k}^{(s)}(2^{-J}n)$ as a proxy for error (4.5). Since we already studied the expected absolute error for two other approximate generation methods, we now focus on the mean absolute detail.

It is clear that the details have the same distribution for $n = 0, \dots, N - 1$ and a given level j , although they are not independent. In Figure 4.11, the simulated distribution of the detail $\sum_k \lambda_{j,k} 4^{-s} 2^{js} \psi_{j,k}^{(s)}(2^{-J}n)$ is plotted for $-j = 8$ (upper panel), together with the mean absolute value for different values of j (lower panel). The normality of the details is guaranteed, since the detail is a finite sum of independent centered normal random variables. By visual inspection of the lower panel, it becomes clear that the mean absolute detail (error) decays very fast in $-j$.

The fast decay of the studied ‘error’ indicate that the wavelet method is an interesting simulation method, but it is not clear from the above arguments how accurate the wavelet method is. In addition, we did not analyze the long-term part, the coarse approximation. From the plots in Figure 4.10 becomes clear that a wavelet-based sample depends heavily on this fractional ARIMA component.

4.2 Stationarity and covariance analysis

Now that we have a rough idea of the errors made by some of the approximate methods, we proceed by checking the characterizing properties on page 6. As already pointed out, the remaining part of the accuracy analysis follows these properties. We start by checking if the approximate fractional Brownian motion sample has stationary increments or, equivalently, if the corresponding approximate fractional Gaussian noise sample is stationary. Thereafter, the tests and estimation techniques of Chapter 3 are used in the evaluation.

It is in general extremely difficult to show statistically that a sample is stationary (more precisely: that a sample is a realization of a stationary process). Many tests are only suitable in certain cases. In addition, much literature on tests for stationarity deal with testing if one particular sample can be regarded as stationary, assuming a specific form of the underlying stochastic process. In the presence of long-range dependence, it is even more difficult to distinguish a stationary process with long memory from a non-stationary process; there appear to be local trends and cycles, which are spurious and disappear after some time [9]. In our situation, we can simulate as many samples as we like, which may be exploited in the stationarity check.

Instead of testing the stationarity statistically, we investigate theoretically whether an approximate method produces stationary samples.

Spectral methods

An example of a method for which it is possible to show that stationary samples are produced is the spectral simulation method (see Section 2.2.4).

Although the spectral simulation method, the Paxson, and approximate circulant method are closely related, we cannot conclude from this fact that Paxson and approximate circulant samples are also stationary. Whereas we could afford some sloppiness with respect to some input FFT-coefficients in the error analysis, it is possible that these coefficients distort the stationarity.

Fortunately, this is not the case. Direct computations for the Paxson method using (2.25) show that for Paxson samples holds that

$$\text{Cov}(X_m, X_n) = \sum_{k=1}^{N/2-1} \frac{f(2\pi k/N)}{N/2} \cos(2\pi(m-n)k/N) + \frac{f(\pi)}{2N} (-1)^{m+n}, \quad (4.6)$$

which depends only on $m - n$, since $(-1)^{m+n} = (-1)^{m-n}$. We write $\tilde{\gamma}_P(\cdot)$ for the resulting autocovariance function. Similar computations for the approximate circulant method show that

$$\text{Cov}(X_m, X_n) = \sum_{k=1}^{N-1} \frac{f(\pi k/N)}{N} \cos(\pi(m-n)k/N) + \frac{N^{2H} - (N-1)^{2H}}{2N} + \frac{f(\pi)}{2N} (-1)^{m-n}. \quad (4.7)$$

From this expression it becomes clear that approximate circulant samples are stationary as well. The resulting autocovariance function is denoted by $\tilde{\gamma}_{AC}(\cdot)$.

Some problems arise for the Paxson method at this point. Recall that in a sample of size N , the covariances $\gamma(0), \dots, \gamma(N-1)$ are approximated by $\tilde{\gamma}_P(0), \dots, \tilde{\gamma}_P(N-1)$. From (1.7) follows that the autocovariance function $\gamma(k)$ decreases in k for $H > 1/2$, but this is not the case for the approximation $\tilde{\gamma}_P(k)$, since this function is symmetrical around $k = N/2$.

To illustrate this problem, the autocovariance function $\tilde{\gamma}_P$ is plotted together with the autocovariance function γ for $H = 0.8$ and $N = 2^8$ in the upper panel of Figure 4.12. The lower panel consists of the functions $\tilde{\gamma}_{AC}$ and γ .

Besides the symmetry problem of $\tilde{\gamma}_P$, the differences with the exact autocovariance function are relatively large; even negative covariances are present in Paxson samples. Still, the Paxson method passed many checks (see Paxson [47]), even without the normality improvement of Chapter 2. This shows how much care should be taken in the evaluation of approximate simulation methods! It should be said that $\tilde{\gamma}_P$ approximates γ better as the sample size N increases, which is probably the reason for Paxson's empirical finding that satisfactory samples are obtained for $N = 2^{15}$.

The autocovariance function $\tilde{\gamma}_{AC}$ of the approximate circulant method is more promising, indicated by the plot in the lower panel of Figure 4.12. Except for high-lag autocovariances, the function is almost indistinguishable from γ , which even improves for $N \rightarrow \infty$. In fact, from Equation (4.6) and (4.7) it becomes clear that $\tilde{\gamma}_P$ as well as $\tilde{\gamma}_{AC}$ converge to γ as $N \rightarrow \infty$ (compare (2.22)):

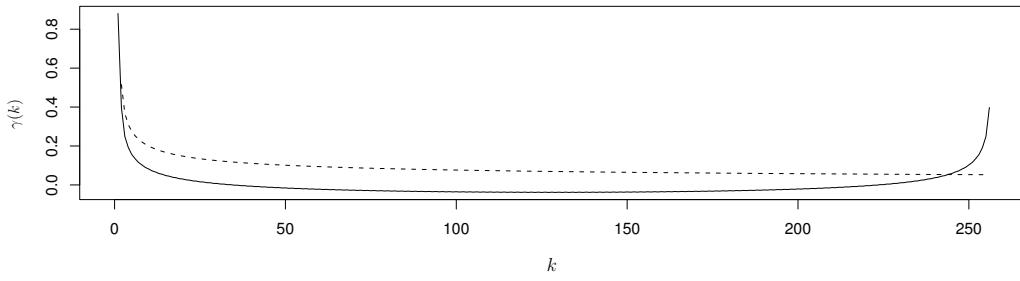
$$\tilde{\gamma}_{P/AC}(k) \rightarrow 2 \int_0^\pi \frac{f(\lambda)}{2\pi} \cos(k\lambda) d\lambda = \gamma(k).$$

However, we are particularly interested in the rate of convergence to zero of $\tilde{\gamma}_{AC}$, which cannot be observed in this plot¹. For a long-range dependent process with Hurst parameter $H = 0.8$, the graph of γ on a log-log scale is a straight line with slope $2H - 2 = -0.4$ for large k .

To check if this is the case for the autocovariance function of the approximate circulant method with $H = 0.8$, the log-log plot is given in Figure 4.13 for $N = 2^{11}$ and $N = 2^{13}$. Since we are only interested in the tail behavior, the plot of the autocovariance function starts with $k = 2^{10}$.

¹I am grateful to Marius Ooms for pointing this out to me.

Autocovariance function of the Paxson method



Autocovariance function of the approximate circulant method

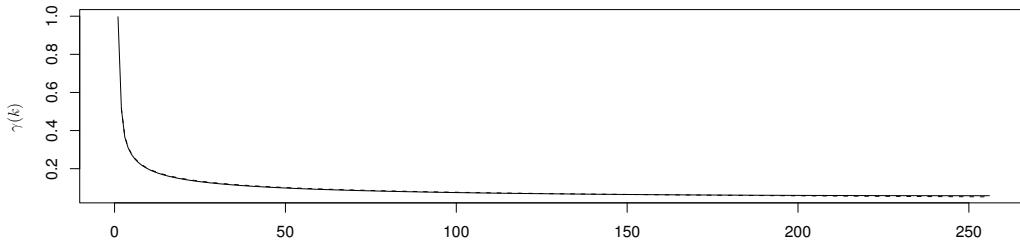
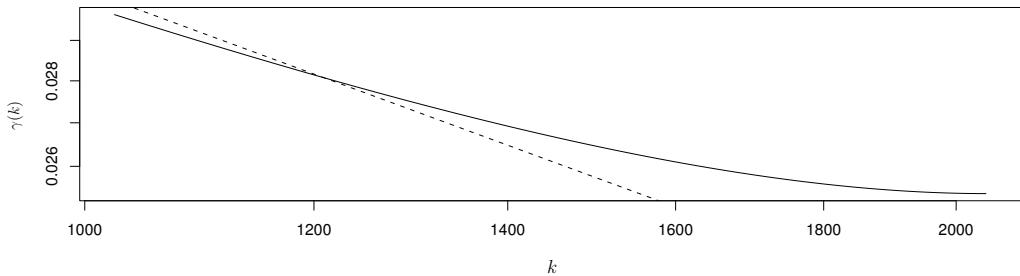


Figure 4.12: The autocovariance function of the Paxson method and the approximate circulant method with the autocovariance function of fractional Gaussian noise (dashed).

Autocovariance function of the approximate circulant method with $N = 2^{11}$



Autocovariance function of the approximate circulant method with $N = 2^{13}$

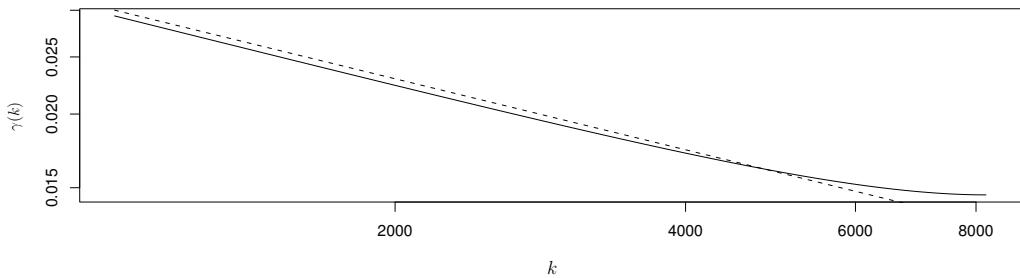


Figure 4.13: The tail of the autocovariance function of the approximate circulant method on a log-log scale. The dashed line is the exact autocovariance function.

We see that the exact autocovariance function is indeed a straight line, but that this is not the case for the autocovariance function of the approximate circulant method. The differences at the left end of the graph are quite small, and can presumably even be made smaller by implementing one of the suggested improvements of Section 2.2.4. It is more interesting that the solid line curves upwards for the approximate circulant method. Because the graph was depicted on a log-log scale, this involves many high-lag autocovariances. From these plots it becomes indeed plausible that the autocovariance function converges pointwise to the exact autocovariance function, although in each case approximately half of the autocovariances is badly approximated. If a sample of size N is needed, an obvious improvement is thus to take the first N elements of an approximate circulant sample of size $2N$. This also makes the method twice as slow.

The $\text{RMD}_{l,r}$ method

It is not always possible to find an explicit expression for the covariance function. For instance, it is not clear how the covariances of the wavelet method can be computed. Norros *et al.* [46] make a plot of the covariance matrices of the $\text{RMD}_{0,1}$ and $\text{RMD}_{1,1}$ method. From this becomes clear that the $\text{RMD}_{l,r}$ method approximates the covariance matrix by a non-stationary ‘terrace architecture’. However, it is not discussed in [46] how to compute the covariance matrix of the $\text{RMD}_{l,r}$ method for general $l \geq 0$ and $r \geq 1$. We will now describe how this covariance matrix can numerically be found. Since the same techniques are used as in the generation method, we will only discuss the main idea in a small example.

Recall that $\mathcal{X}_i = \{X_{ij} = B_H(j2^{-i}) - B_H((j-1)2^{-i}) : j = 1, \dots, 2^i\}$. Consider the probability density $f_{X_{2,1}, X_{2,2}, X_{2,3}, X_{2,4}}$ of the random vector $(X_{2,1}, X_{2,2}, X_{2,3}, X_{2,4})$. This density can be computed by noting that (assume that $l \geq 1$ and $r \geq 2$)

$$\begin{aligned} f_{X_{2,1}, X_{2,2}, X_{2,3}, X_{2,4}}(x_1, x_2, x_3, x_4) &= f_{X_{2,1}, X_{2,1}+X_{2,2}, X_{2,3}, X_{2,3}+X_{2,4}}(x_1, x_1+x_2, x_3, x_3+x_4) \\ &= f_{X_{2,1}, X_{2,3}|X_{2,1}+X_{2,2}=x_1+x_2, X_{2,3}+X_{2,4}=x_3+x_4}(x_1, x_3)f_{X_{2,1}+X_{2,2}, X_{2,3}+X_{2,4}}(x_1+x_2, x_3+x_4) \\ &= f_{X_{2,1}, X_{2,3}|X_{1,1}=x_1+x_2, X_{1,2}=x_3+x_4}(x_1, x_3)f_{X_{1,1}, X_{1,2}}(x_1+x_2, x_3+x_4) \\ &= f_{X_{2,1}|X_{1,1}=x_1+x_2, X_{1,2}=x_3+x_4}(x_1)f_{X_{2,3}|X_{2,1}=x_1, X_{2,2}=x_2, X_{1,2}=x_3+x_4}(x_3) \\ &\quad \cdot f_{X_{1,1}, X_{1,2}}(x_1+x_2, x_3+x_4). \end{aligned}$$

In the notation of Section 2.2.3, it is readily checked that the previous formula implies that

$$\begin{aligned} (x_1, x_2, x_3, x_4)\Gamma_2^{-1}(x_1, x_2, x_3, x_4)' &= \frac{(x_1 - e(2, 0)(x_1 + x_2, x_3 + x_4)')^2}{v(2, 0)} \\ &\quad + \frac{(x_3 - e(2, 1)(x_1, x_2, x_3 + x_4)')^2}{v(2, 1)} \\ &\quad + (x_1 + x_2, x_3 + x_4)\Gamma_1^{-1}(x_1 + x_2, x_3 + x_4)'. \end{aligned}$$

From this, the inverse Γ_2^{-1} of the covariance matrix of $(X_{2,1}, X_{2,2}, X_{2,3}, X_{2,4})$ is determined uniquely and can be computed once the inverse Γ_1^{-1} of the covariance matrix of $(X_{1,1}, X_{1,2})$ is known. This is done by writing every term on the right hand side in the form $(x_1, x_2, x_3, x_4)A(x_1, x_2, x_3, x_4)'$ for some matrix A , thus rewriting the whole right hand side in this form, which gives the matrix Γ_2^{-1} . The same idea can be used to compute Γ_3^{-1} , Γ_4^{-1} , and so on. The truncation parameters l and r show up when for instance $r = 1$ and $f_{X_{2,1}|X_{1,1}=x_1+x_2, X_{1,2}=x_3+x_4}(x_1)$ is approximated by $f_{X_{2,1}|X_{1,1}=x_1+x_2}(x_1)$.

When the covariance matrix of the $\text{RMD}_{l,r}$ method for a sample size $N = 2^g$ should be computed, this process is stopped once Γ_g^{-1} is known. The inverse of Γ_g^{-1} is then the required covariance matrix.

In Figure 4.14, the covariance matrices of the $\text{RMD}_{1,1}$ and $\text{RMD}_{1,2}$ method are plotted. In the terminology of [46], the ‘terrace architecture’ is clearly present in the $\text{RMD}_{1,1}$ matrix, whereas it seems to have disappeared in the $\text{RMD}_{1,2}$ matrix. The $\text{RMD}_{1,2}$ matrix is visually almost indistinguishable from the exact covariance matrix. However, the matrices are not equal. As already pointed out, the approximate covariance matrix does not correspond to a stationary process. Therefore, it is impossible to analyze the covariances of the $\text{RMD}_{l,r}$ method in the same way as we did for the spectral methods: the diagonal and its parallel vectors do not have constant elements anymore. Therefore, we plot the minimum and maximum value of each ‘diagonal’ vector. From this, we can still get an idea about the decay of the elements in the covariance matrix. The resulting plot is given in Figure 4.15 for the $\text{RMD}_{1,2}$ and $\text{RMD}_{3,3}$ methods with sample size $N = 2^8$.

Note that no information is provided by Figure 4.15 on the *distribution* of the ‘diagonal’ vectors; it is well possible that all covariances are exact, except for two covariances: the minimum and the maximum value. In fact, the lines that correspond to the first and third quartile of this distribution lie very close to the exact value except for the rightmost part (these quartile-lines were not depicted to prevent them from spoiling the visual effect of the plot). Like the observations in the error analysis, this suggests that these methods are very

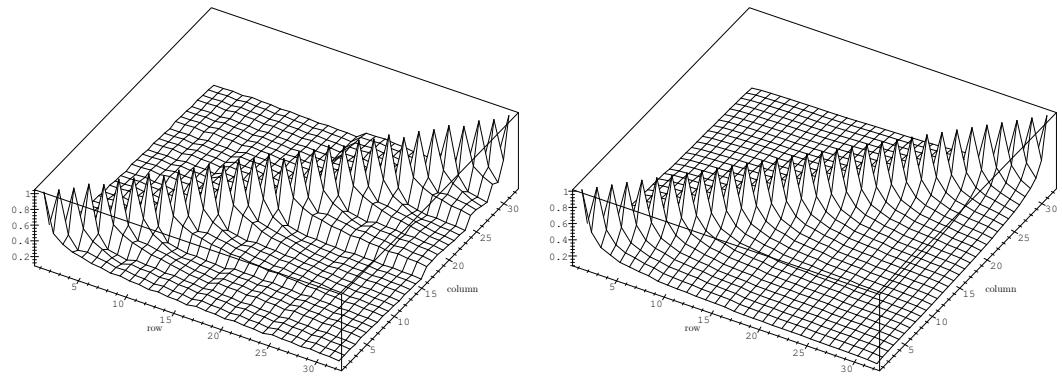


Figure 4.14: The 32×32 covariance matrices of the $RMD_{1,1}$ (left) and the $RMD_{1,2}$ method (right). $H = 0.8$.

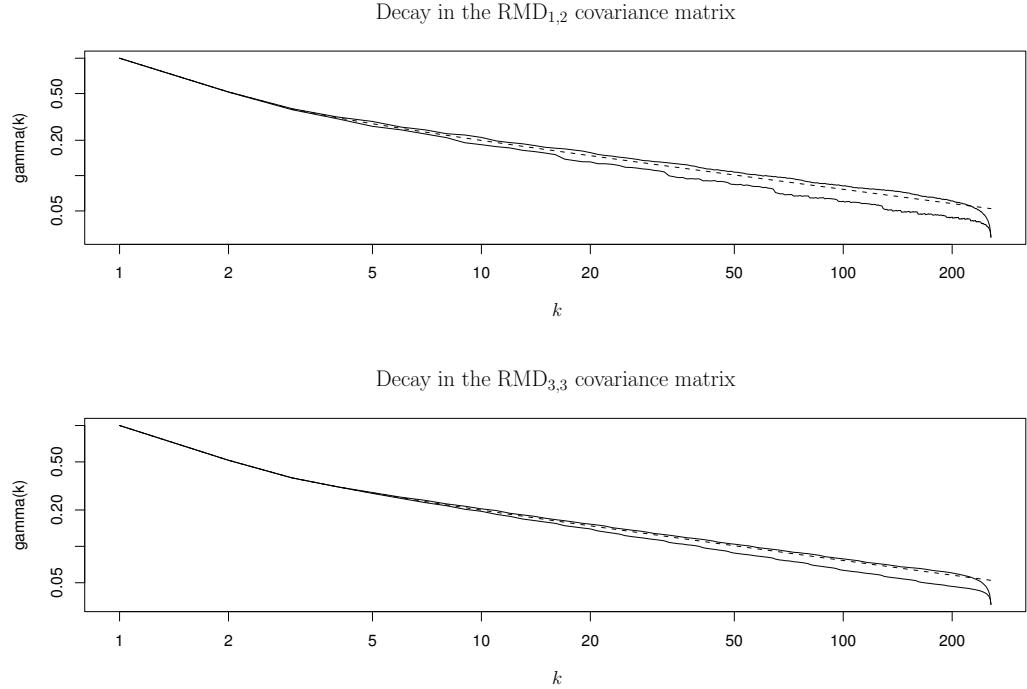


Figure 4.15: The decay in the covariance matrix of the $RMD_{1,2}$ and $RMD_{3,3}$ method on a log-log scale. Again, the dashed line is the exact autocovariance function.

	Davies & Harte	RMD _{1,2}	RMD _{2,1}	RMD _{3,3}	Paxson	approx. circulant	wavelet $J = 6$	wavelet $J = 7$
mean	0.008159	0.008615	0.01958	0.008291	0.1090	0.04486	0.02084	0.02144
max	0.04359	0.05154	0.1961	0.05010	0.3458	0.08606	0.5157	0.3549

Table 4.1: Sample mean and maximum absolute value of the elements in the error matrix.

interesting for simulation purposes. Although the minimum and maximum value of the RMD_{3,3} method seems to lie closer to the exact value, the differences between the RMD_{1,2} and the RMD_{3,3} method are definitely not so large. An important observation is that the rightmost part of the graph curves downwards, which is not desirable. As already suggested in the error analysis, this may be solved by removing some sample points at the beginning and the end of the trace.

The wavelet method; a simulation experiment

To gain more insight into the approximation of the covariance structure of the wavelet method, we perform a simulation experiment to estimate the covariances. This sample covariance matrix is then compared to the theoretical covariance matrix determined by the autocovariance function $\gamma(\cdot)$.

To find a sample covariance matrix, we generate n samples of size N with the wavelet method. Denote the i th sample by $X^{(i)} = \{X_k^{(i)} : k = 0, \dots, N - 1\}$. From these samples, we estimate element (l, k) of the covariance matrix by

$$\frac{1}{n-1} \left(\sum_{i=0}^{n-1} X_l^{(i)} X_k^{(i)} - \frac{1}{n} \left[\sum_{i=0}^{n-1} X_l^{(i)} \sum_{i=0}^{n-1} X_k^{(i)} \right] \right),$$

for $l, k = 0, \dots, N - 1$. After that, the theoretical covariance matrix is subtracted from the estimated matrix to obtain the ‘sample error’ matrix.

The average absolute value and the largest absolute value of the elements in the error matrix provide information on how well the wavelet method performs. To make it possible to compare the wavelet error matrix, the procedure is also followed for the Davies and Harte method and the other approximate methods. The largest absolute error and the mean absolute error are tabulated in Table 4.1. All samples are of size $N = 2^8$ and were generated with input Hurst parameter $H = 0.8$. The number of simulation runs is $n = 10000$. We see that the performance of the RMD_{2,1} method is rather disappointing, whereas the RMD_{1,2} and RMD_{3,3} methods have estimated covariance matrices quite close to the real covariance matrices, as already suggested by the previous theoretical covariance analysis. This is especially true for the RMD_{3,3} method, in which more sample points are used in the conditioning. The resulting (non-stationary) samples have covariances that are even closer to the required covariances than the (stationary) approximate circulant samples! As expected, the approximate circulant method performs better than the Paxson method, which is reflected by smaller values in Table 4.1. In fact, the ‘max’ value is close to the maximum absolute distance between the two covariance functions in the upper panel of Figure 4.12, 0.3456. The two wavelet methods perform quite badly, indicated by the large values in the table. However, we should keep in mind that the covariance matrix was *simulated*, which makes it hard to draw robust conclusions.

4.3 Testing the approximate samples

According to the properties on page 6 that outline the structure of this accuracy analysis, we should test if long-range with the right Hurst parameter is present in the approximate samples. However, we will first test if the approximate simulation methods produce samples that can be regarded as fractional Gaussian noise. Because it is an assumption in the Whittle estimation procedure that the sample is fractional Gaussian noise (see Section 3.1.8), this must be done *before* the Hurst parameter is estimated. To test this assumption statistically, we have the two tests described in Chapter 3 at our disposal.

4.3.1 The goodness-of-fit test for the spectral density

The first is Beran’s goodness-of-fit test for the spectral density, which is especially important since the Whittle estimator is also based on a ‘spectral’ approximation of the likelihood function. Using several approximate methods, 1000 samples of length $N = 2^{13}$ are generated for $H = 0.5$, $H = 0.65$ and $H = 0.8$. For these samples, the test statistic is computed and compared to its theoretical (asymptotic) Gaussian distribution using the Kolmogorov-Smirnov test. Although no problems will emerge for the exact Davies and Harte

H	Davies & Harte	RMD _{1,2}	RMD _{2,1}	RMD _{3,3}	Paxson	approx. circulant	wavelet $J = 6$	wavelet $J = 7$
0.5	0.7837	0.4692	0.4990	0.7174	0.1256	0.2348	0.2683	0.4378
0.65	0.5871	0.6329	0.4219	0.4458	0.7149	0.6693	0.0000	0.0000
0.8	0.8872	0.4717	0.0000	0.4626	0.4079	0.3356	0.0000	0.0000

Table 4.2: Kolmogorov-Smirnov p -values for the goodness-of-fit test with $N = 2^{13}$.

Davies & Harte	RMD _{1,2}	RMD _{2,1}	RMD _{3,3}	Paxson	approx. circulant	wavelet $J = 6$	wavelet $J = 7$
0.08686	0.8771	0.0000	0.8507	0.0000	0.5799	0.0000	0.0000

Table 4.3: p -values of the Kolmogorov-Smirnov test on the chi-square test statistic ($H = 0.8$).

method, the test is also performed on this method. However, our main interest is the performance of the test on approximate samples. The results are summarized in Table 4.2 by the p -values of the Kolmogorov-Smirnov test statistic.

The three RMD _{l,r} methods produce samples that may be regarded as fractional Gaussian noise for $H = 0.5$ and $H = 0.65$, but the situation is different for the RMD_{2,1} method with $H = 0.8$. In fact, the RMD _{l,r} method is exact for $H = 0.5$ for every choice $l \geq 0, r \geq 1$, see [39]. For samples generated with the RMD_{2,1} method with $H = 0.8$, it is rejected at any reasonable confidence level that the spectral properties of the samples are in accordance with fractional Gaussian noise. The RMD_{3,3} method uses more sample points in the conditioning, which must lead to better results than the RMD_{1,2} and RMD_{2,1} methods. However, the p -values in the table show that we cannot distinguish RMD_{1,2} and RMD_{3,3} samples with this goodness-of-fit test (note that it was also reflected in Figure 4.15 that the RMD_{3,3} method is only *slightly* more accurate than the RMD_{1,2} method). This is also the reason why we make use of various estimators of the Hurst parameter: some estimators may be sensitive enough to conclude that the RMD_{3,3} method is better than the RMD_{1,2} method; we need to analyze the approximate samples in more detail to get enough feeling for the methods to draw conclusions.

The two spectral simulation methods, the Paxson and approximate circulant method, produce satisfactory results. On the other hand, the results for the wavelet method are dramatically bad but for $H = 0.5$. Therefore, it is interesting to investigate whether the chi-square test rejects that wavelet samples are fractional Gaussian noise for $H \neq 0.5$.

4.3.2 The chi-square test

In Section 3.2.2, a chi-square test was discussed to test if a sample is fractional Gaussian noise with some specified parameter H .

As earlier, we simulate a large number of samples (in this case even 10000) to make it possible to test with the Kolmogorov-Smirnov test if the distribution of the test statistic (3.10) equals the theoretical distribution under the null hypothesis. To avoid the memory allocation problems discussed in Section 3.2.2, we simulate samples of length $N = 2^8$. We set $H = 0.8$. The p -values of this Kolmogorov-Smirnov test can be found in Table 4.3.

It is remarkable that the p -value of the Davies and Harte method is this low, since we expect a high value for this exact method. This must be coincidence, but it is definitely no coincidence that the p -value for the Paxson method is 0.0000. Apparently, the chi-square test is able to detect that the covariance function of the Paxson method differs from the covariance function of fractional Gaussian noise (see Figure 4.12). The goodness-of-fit test in the spectral domain was not able to detect this. The other p -values are in accordance with the results of the goodness-of-fit test in Table 4.2.

4.4 Long-range dependence in approximate samples

The long-range dependence in new approximate simulation methods for fractional Brownian motion is often studied using a couple of estimators discussed in Chapter 3, e.g., by Norros *et al.* [46] and Paxson [47]. The resulting estimates for the Hurst parameter H are then compared with the input parameter of the algorithm. In some cases, the same approach is followed to find out what the best choice is for some parameter that controls the approximation. An example is the RMD _{l,r} method [46].

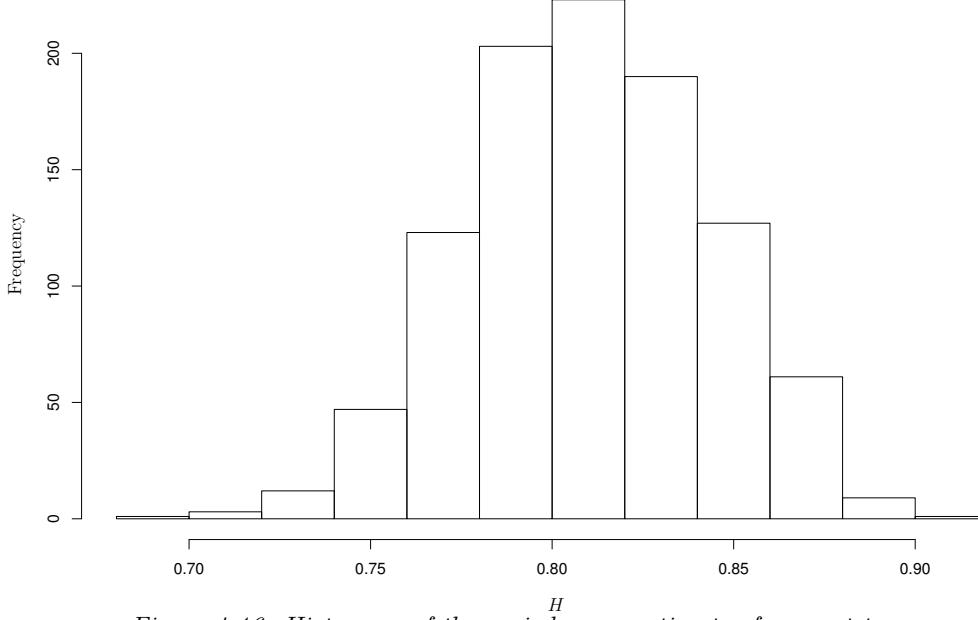


Figure 4.16: Histogram of the periodogram estimates for exact traces.

Without knowing what the properties of the estimator are, this may lead to wrong conclusions, e.g., when the estimator is biased. Assume that the estimator has negative bias; the estimate of the Hurst parameter in exact samples is structural lower than the input parameter H . Let us also assume that we have two approximate methods. One produces samples with exact Hurst parameter H , and one the other produces samples with a Hurst parameter $\tilde{H} > H$. Typically, the *estimated* Hurst parameter is smaller than H for the first approximate method. It is possible that the estimate of the Hurst parameter of the second method equals approximately H . The naive conclusion is then that the second approximate method is better!

It is mostly impossible to study the properties (e.g., bias) of the estimator theoretically. As an example, we consider the periodogram estimator. If the disturbance term in the linear fitting is a centered Gaussian variable with equal variance for every sample point and moreover independent, the resulting least-square estimator is unbiased and normally distributed. However, the distribution of the disturbance term is exponential (see Beran [9]), so we can not be sure that the estimator is unbiased and Gaussian. The histogram of 1000 estimates for the Davies and Harte method is given in Figure 4.16 ($H = 0.8$, $N = 2^{14}$), and suggests that the estimator is indeed Gaussian, although slightly biased.

In some cases, it is possible to derive a theoretical result concerning the limiting distribution of the estimator. As an example, Coeurjolly [15] uses the asymptotical distribution of his discrete variations estimator (see Section 3.1.3) in his comparison of some approximate simulation methods.

As with all estimators in this section, the parameters of the estimation procedures are chosen as in Chapter 3, unless stated otherwise.

To overcome the problem that the distribution of the estimator under the null hypothesis (that the traces have the right Hurst parameter) is not known, we simply simulate this distribution. After simulating this distribution for each approximate simulation method as well, it is tested whether the distributions are equal. In this section, we simulate the distributions of the estimators of Chapter 3 and perform the test of equality in distribution. This makes it possible to draw conclusions using estimators that are highly biased, which solves the earlier mentioned problem.

Several nonparametric tests are available to test the equality of distributions.

Assume we have two populations, X and Y , with cumulative distribution functions denoted by F_X and F_Y , respectively. We have an independent sample of size m drawn from the X population, X_1, \dots, X_m and another independent sample of size n drawn independently from the Y population, Y_1, \dots, Y_n (note that this m and n have nothing to do with the previously used notation). It is assumed that there are no ties, which is a reasonable assumption in our case, since we are dealing with continuous distributions. We would like to test the hypothesis that the two samples are drawn from the same distribution, $F_X(x) = F_Y(x)$ for all x . The most general two-sided alternative hypothesis is simply $F_X(x) \neq F_Y(x)$ for some x , but the available tests under this hypothesis, like the Kolmogorov-Smirnov two-sample test, have little power against small deviations in the distribution.

In our case, we expect the differences in the two samples to be rather small, so we should make some

concessions and restrict the alternative hypothesis a bit more: we will test against differences in location (mean/median) and scale (interquartile range/variance).

For this, the Wilcoxon test and the Freund-Ansari-Bradley will be used. More details on these tests and some other nonparametric tests with this null hypothesis can be found in Gibbons [30], on which the following two subsections are based.

4.4.1 The Wilcoxon test

The Wilcoxon test is based on a so-called linear rank statistic. The rank r of an observation is the rank of the observation in the combined sample (i.e., a sample of size $N = n + m$) after sorting. This means

$$\begin{aligned} r(X_i) &= \sum_{k=1}^m \mathbf{1}(X_i \geq X_k) + \sum_{k=1}^n \mathbf{1}(X_i \geq Y_k) \\ r(Y_i) &= \sum_{k=1}^m \mathbf{1}(Y_i \geq X_k) + \sum_{k=1}^n \mathbf{1}(Y_i \geq Y_k). \end{aligned}$$

It is easier to denote the rank in terms of a vector of indicator random variables as follows. Let

$$Z = (Z_1, Z_2, \dots, Z_N),$$

where $Z_k = 1$ if the k th random variable in the combined ordered sample is an X -observation and $Z_k = 0$ otherwise. A linear rank statistic is a linear combination of the elements of the vector Z .

The Wilcoxon test tests the null hypothesis of equality in distribution of the X and Y sample against the alternative of a shift in location:

$$H_1 : F_X(x) = F_Y(x - s) \text{ for all } x \text{ and some } s \neq 0.$$

The idea is that the ranks of the X -observations will generally be larger or smaller than the ranks of the Y -observations when the samples have a shift in location. When an X -observation has rank k in the combined sample, the rank may be computed from Z by kZ_k . The Wilcoxon test statistic is

$$W_N = \sum_{k=1}^N kZ_k.$$

The probability distribution of W_N may be computed by enumeration or recursion (see [30]), but for larger sample sizes (as in our case) a normal approximation to the distribution is very accurate. The expectation and variance of W_N can easily be computed, but we will not go into details.

The test can shown to be consistent, which means that the probability of rejecting the null hypothesis approaches 1 as the sample size increases if the alternative is true. Although the test statistic is not efficient, it has a relatively low asymptotic variance when the samples are Gaussian.

4.4.2 The Freund-Ansari-Bradley test

Once the equality of the location parameter of the two samples is not rejected, we test whether the samples differ in scale. A nonparametric test that is designed for this is the Freund-Ansari-Bradley test. This test has an alternative hypothesis of a shift in scale:

$$H_1 : F_{X-M}(x) = F_{Y-M}(sx) \text{ for all } x \text{ and some } s > 0, s \neq 1,$$

where M denotes the (common) median of the X - and Y -observations. The idea of the test is that the deviation of the rank of the k th ordered variable about its mean rank $(N+1)/2$ provides information on the variability of the X -observations relative to the Y -observations. When the X -observations are more spread out around the median, the deviation about the mean rank is bigger. The Freund-Ansari-Bradley test statistic is

$$F_N = \sum_{k=1}^N \left(\frac{N+1}{2} - \left| k - \frac{N+1}{2} \right| \right) Z_k.$$

Again, the exact distribution of F_N can be found by enumeration and recursion. A normal approximation is already suitable for relatively moderate sample sizes. Moreover, the test is consistent.

We will use these tests to study whether the approximate simulation methods produce samples with the same long-range dependent behavior as exact samples.

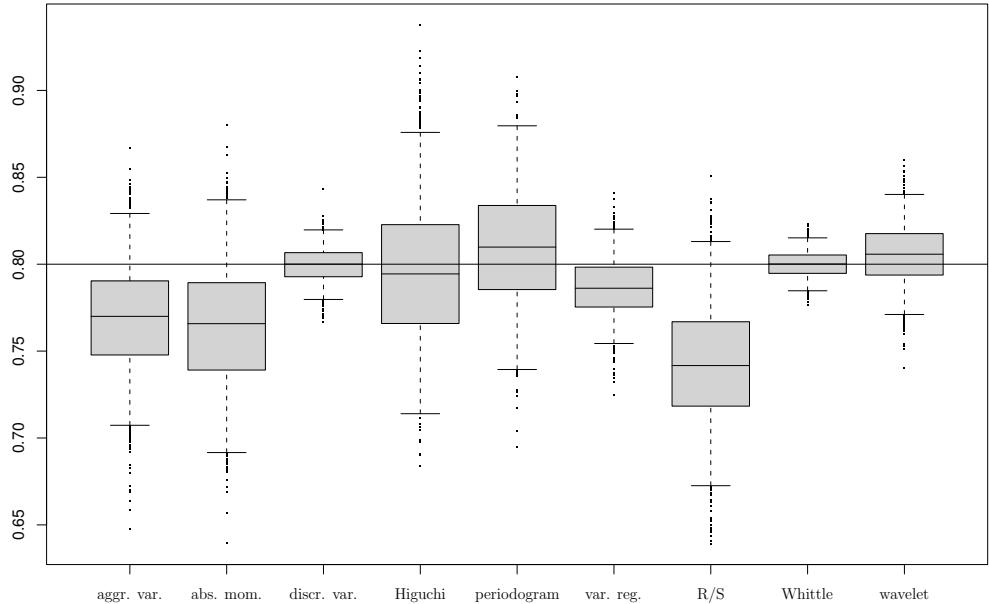


Figure 4.17: Box plot for the exact Davies and Harte method.

4.4.3 Estimation of H for the approximate methods

When estimating the Hurst parameter H , specific properties of the estimators come into play. Following Taqqu *et al.* [55], we produce a so-called box plot for the exact Davies and Harte method to get some feeling for these properties. This box plot, given in Figure 4.17, provides information on the distribution of each estimator. It was produced with 1000 exact fractional Gaussian noise traces of length $N = 2^{13}$. For each trace, the Hurst parameter was estimated with all estimators described in Chapter 3, which led to an empirical distribution for each estimator.

Box plots have the property that half of the estimates lie in the region marked by the gray box, i.e., the box gives the interquartile range. The horizontal line through each box represents the median of the sample, whereas the whiskers end are chosen such that approximately 95% of the estimates lie between them. The estimates that do not lie between the whisker ends are depicted as tiny dots. A big horizontal line is plotted at the real Hurst parameter of the sample, $H = 0.8$.

We thus expect that the estimates of a good estimator lie close to this line. In fact, an unbiased estimator has the property that the median equals the real Hurst parameter when the distribution of the estimator is symmetric. We see that only the discrete variations and the Whittle method produce satisfactory results in this respect, although some other estimators have only a small bias. However, we should keep in mind that the distributions of the estimators are simulated, making it more difficult to draw conclusions concerning the bias of the various estimators.

It is now possible to see what happens when we evaluate an approximate simulation method only by means of the input Hurst parameter. Using R/S estimates, we would clearly say that Davies and Harte samples do not have the right long-range dependent behavior. This is definitely a wrong conclusion, because we know that the Davies and Harte method is exact! Fortunately, the most popular estimator in evaluating approximate methods is Whittle's estimator, which produces satisfactory results.

Long-range dependence in approximate samples

It was already pointed out that we test whether the right Hurst parameter is present in the approximate samples by comparing the empirical distributions of various estimators for H to the simulated distribution under the null hypothesis that the sample is exact. The nonparametric tests at the beginning of this section make it possible to test whether the simulated distributions of the estimators differ in location (the Wilcoxon test) and in scale (the Freund-Ansari-Bradley test).

The p -values of the Wilcoxon test are tabulated in Table 4.4 for several approximate generation methods and estimators. Given the estimator, the p -values were obtained by performing the test on 1000 (independent) estimates of the Hurst parameter for exact samples of length $N = 2^{13}$ (the X -observations in the terminology of the beginning of this section) as well as 1000 estimates for approximate samples of the same length (the Y -observations). The distribution of the test statistic was approximated by a normal distribution, but this has no influence on the p -values, since the sample size (1000) is relatively large. The p -value is interpreted

Estimator	RMD _{1,2}	RMD _{2,1}	RMD _{3,3}	Paxson	approx. circulant	wavelet $J = 6$	wavelet $J = 7$
Aggr var	0.9002	0.0009	0.5106	0.0013	0.4505	0.7750	0.5136
Abs mom	0.6273	0.2117	0.2352	0.8705	0.1576	0.0519	0.1563
Disc var	0.0400	0.0000	0.2679	0.2429	0.2719	0.0000	0.0000
Higuchi	0.6187	0.0175	0.9331	0.0000	0.0001	0.0001	0.0878
Per	0.5679	0.4240	0.2193	0.7806	0.3253	0.0427	0.2983
Var regr	0.7718	0.0000	0.6222	0.6211	0.8842	0.0000	0.0000
R/S	0.6853	0.0000	0.8757	0.3277	0.8722	0.0000	0.0000
Whittle	0.5543	0.0000	0.9722	0.0474	0.0584	0.0000	0.0000
Wavelet	0.1511	0.0000	0.4725	0.1768	0.2127	0.0000	0.0000

Table 4.4: p -values for the Wilcoxon test of location.

Estimator	RMD _{1,2}	RMD _{2,1}	RMD _{3,3}	Paxson	approx. circulant	wavelet $J = 6$	wavelet $J = 7$
$N = 2^{10}$	0.5294	0.0000	0.4438	0.5294	0.0137	0.0000	0.0000
$N = 2^{11}$	0.3000	0.0000	0.5936	0.1883	0.9311	0.0000	0.0000
$N = 2^{12}$	0.8609	0.0000	0.8853	0.3152	0.2988	0.0000	0.0000
$N = 2^{13}$	0.5543	0.0000	0.9722	0.0474	0.0584	0.0000	0.0000
$N = 2^{14}$	0.4552	0.0000	0.6016	0.7838	0.1696	0.0000	0.0000

Table 4.5: Wilcoxon p -values for the Whittle estimator when varying the sample size N .

by its definition; given that the test statistic is w , it is the probability that the test statistic W_N is ‘more extreme’ (i.e., further away from its mean) than w under the condition that the null hypothesis holds; this probability is then multiplied by two since we perform a two-sided test.

It is worth noticing that the RMD_{3,3} gives the best results of all tested RMD _{l,r} methods, which is precisely what we expected. The RMD_{1,2} produces also quite satisfactory results. In fact, the RMD_{1,2} method was already called a ‘surprising exception’ by Norros *et al.* [46], but it is not clear how it can be explained why this is the case. Especially the difference in the performance of the RMD_{1,2} and the RMD_{2,1} method is quite remarkable.

As was already clear from theoretical considerations, the Paxson and approximate circulant method share many properties for large sample sizes. This is confirmed by the results in Table 4.4. The variance structure of the approximate fractional Brownian motion samples seems to be affected by the Paxson approximation, because the aggregated variance estimator yields a very low p -value for this method. The explanation is that the covariance structure of Paxson samples differs significantly from the requested covariance structure, see Section 4.2. The aggregated variance method is apparently influenced by this. It turns out that the Higuchi method is sensitive to the approximation by spectral densities that is used in both methods. The conclusion from the p -values in the table is that it is not likely that the median of the distribution of the Higuchi estimator using the approximate methods equals the median of the distribution of this estimator using an exact method. In addition, we see that the p -values for the Whittle method are rather low; at a 5% confidence level, we reject the Wilcoxon test for the Paxson method and the approximate circulant method is ‘close to rejection’.

Looking only at the p -values of the wavelet method, it seems very questionable that long-range dependence with the right Hurst parameter is present. Note that the zeros do not indicate that no long-range dependence is present, but that they indicate that it does not correspond to the requested Hurst parameter. To illustrate this, the sample mean of the discrete variations estimates is 0.7379 for the wavelet method ($J = 7$) and 0.7996 for the Davies and Harte method (the estimated variances are both close to 0.010). The differences are not always this large; the sample mean of the wavelet estimator (performed on the same samples as the discrete variations method!) is 0.7980 for the wavelet method and 0.8054 for the Davies and Harte method, still rejecting the equality of both sample medians. Since the wavelet and discrete variations estimator do not have enough bias to explain this huge difference (see Figure 4.17), it is definitely a weakness of the wavelet method that the estimate depends so highly on the used estimation method.

The reader may be eager to find out whether a similar table as Table 4.4 for, say, $H = 0.65$ lead to the same conclusions. However, we will not study this here for two reasons. The first is the practical argument that it is simply too time-consuming (in terms of computer time) to perform another simulation study like this. More important is that our main use of the simulation methods is the simulation of data traffic. The

Estimator	RMD _{1,2}	RMD _{2,1}	RMD _{3,3}	Paxson	approx. circulant	wavelet $J = 6$	wavelet $J = 7$
Aggr var	0.3983	—	0.8099	—	0.8715	0.4648	0.7045
Abs mom	0.0215	0.3108	0.5152	0.0035	0.2369	0.0619	0.1584
Disc var	—	—	0.1039	0.2102	0.0889	—	—
Higuchi	0.4185	—	0.2501	—	—	—	0.0002
Per	0.4930	0.1332	0.0694	0.8564	0.6171	—	0.4288
Var regr	0.2512	—	0.8209	0.5560	0.8399	—	—
R/S	0.2605	—	0.0461	0.3956	0.2174	—	—
Whittle	0.5670	—	0.8200	—	0.4473	—	—
Wavelet	0.5745	—	0.3198	0.8044	0.7787	—	—

Table 4.6: *p*-values for the Freund-Ansari-Bradley test of scale.

measurements on Ethernet LAN traffic by Willinger *et al.* [57] showed that the Hurst parameter is about 0.80.

Another interesting question is what happens to the *p*-values when the sample size N varies. To avoid the mentioned computational problems, we only study the influence on the Whittle estimator. The *p*-values of the Wilcoxon test are found in Table 4.5. We see that the methods with a bad performance for $N = 2^{13}$, the RMD_{2,1} and wavelet methods, do not show improvement when the sample size is varied. The results for the other methods look also familiar, although the *p*-values for the two spectral simulation methods show some improvement, except for the approximate circulant method with $N = 2^{10}$. This must be the result of random fluctuations, since the theoretical results in Chapter 2 suggest that the approximate circulant method produces qualitatively better traces than the Paxson method (compare the autocovariance function in Figure 4.12).

Having tested whether the location of the distributions of each estimator in the exact sample case differs from the approximate sample case, we will now test the variance of these distributions. To be able to perform the Freund-Ansari-Bradley test, we need equality in location of the two distributions.

In Table 4.6, the Freund-Ansari-Bradley *p*-values are given for our samples of the distribution of the estimators. If the equality of locations was rejected by the Wilcoxon test at a 5% confidence level, the Freund-Ansari-Bradley test was not performed. The results show that the variances can be considered equal in most cases. The Freund-Ansari-Bradley test does not change the impression that the RMD_{1,2}, the RMD_{3,3} and the approximate circulant method produce relatively good approximate samples.

Although the equality of the sample variances of the distribution of the estimator is clearly a desirable property, it is more desirable that the Hurst parameter of the samples is close to the input parameter H . In that respect, the results of the Wilcoxon test should be valued more than the results of the Freund-Ansari-Bradley test.

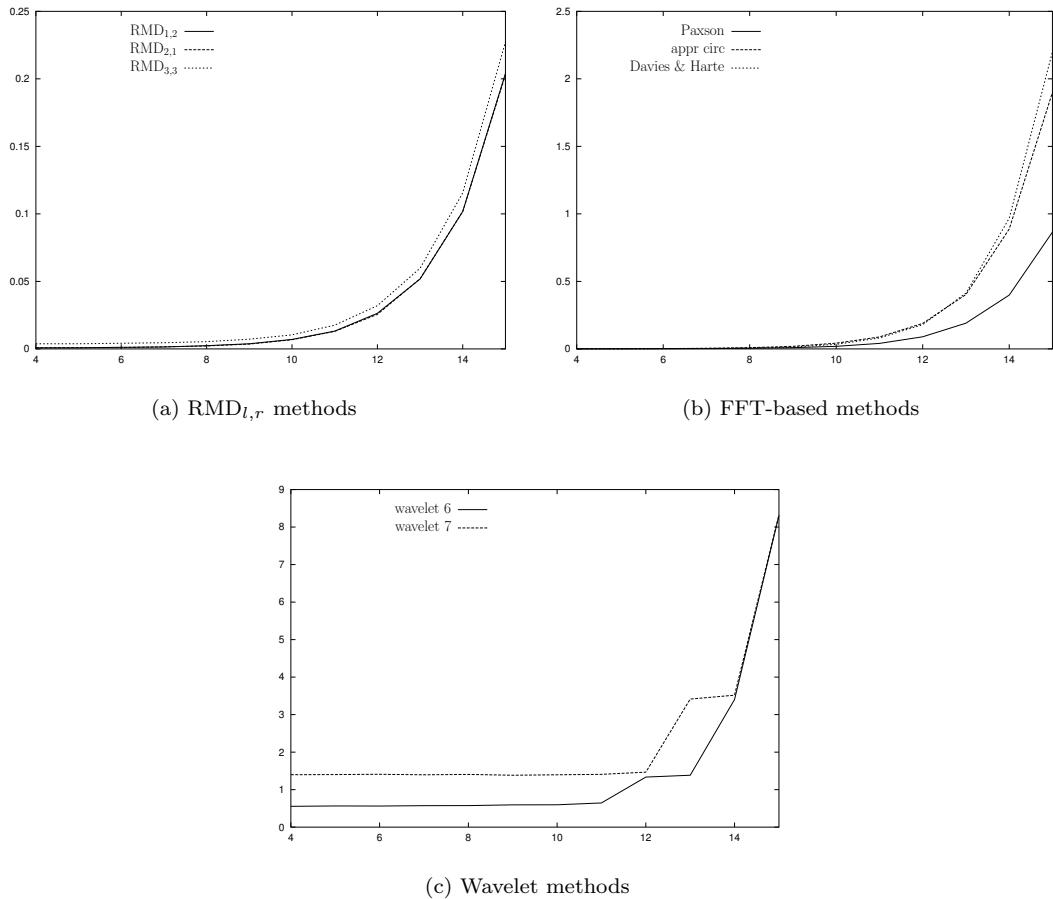


Figure 4.18: The amount of time (in seconds) needed to generate fractional Brownian motion samples for several methods as function of the (\log_2) sample size.

4.5 Time complexity analysis

Now that we have seen that some of the approximate simulation methods may offer an alternative for the exact Davies and Harte method in terms of accuracy, it is interesting to investigate the speed of these methods. To make a fair comparison, the algorithms were all implemented in C and use the same random number package Ranlib, available via Netlib (<http://www.netlib.org>). The matrix algebra package Meschach was used for basic matrix algebra and the fast Fourier transform (FFT). Meschach is available via <http://www.netlib.org/c/meschach>. However, no efforts have been made to tailor and optimize the code.

For a given trace length, we compute the CPU time by generating 50 traces and average the needed time. This time complexity analysis was performed on an Intel Pentium 200 MMX processor. The CPU time as a function of the (\log_2) sample size is plotted in Figure 4.18.

In Figure 4.18(a), the needed time of the three $RMD_{l,r}$ methods under study is depicted. As expected, the $RMD_{1,2}$ and $RMD_{2,1}$ produce samples in the same amount of time. We clearly see that it takes more time for the $RMD_{3,3}$ method to initialize the algorithm, since more matrices used in the conditioning have to be computed. After this is done, the method is approximately as fast as the $RMD_{1,2}$ or $RMD_{2,1}$ method, which makes the gain in accuracy particularly interesting when large sample sizes are required (theoretically, the computations after initialization are a little more time-consuming, because the multiplications with vectors of the type $e(i, k)$ take more time. Moreover, the vectors and matrices should only be computed once when more than one sample is required (e.g., in a simulation study).

Figure 4.18(b) consists of the time that is needed by the two spectral simulation methods and the Davies and Harte method. We clearly see that the Paxson method is the fastest, as expected from the theoretical considerations in Chapter 2. The approximate circulant method requires about twice the amount of computing time of the Paxson method, which was also foreseen. However, it is a bit surprising that the approximate circulant method and the Davies and Harte method are approximately equally fast. This must be caused by the fact that the spectral density is evaluated at N points. This is done in order N seconds, but it only speeds

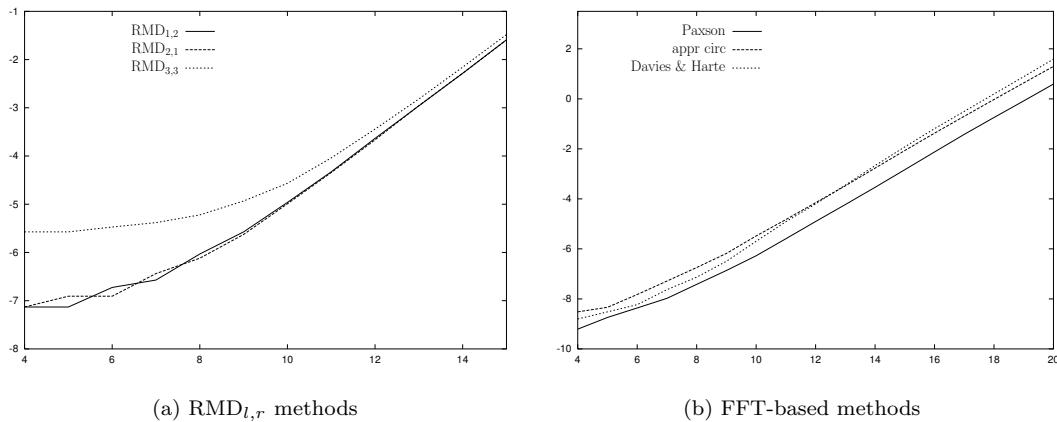


Figure 4.19: A function of the amount of time such that a straight line should appear for large N .

up the algorithm when N is large enough (since an FFT operation of order $N \log(N)$ seconds is saved). We also see this in the plot, because the curve for the Davies and Harte method is steeper for larger sample sizes.

At first sight, the plot of the computing time for the wavelet method in Figure 4.18(c) looks a bit strange, but it is possible to explain its most important features. For small sample sizes, the graph looks almost flat and increases only a bit. In fact, the length of the fractional ARIMA trace that is used as input in the algorithm increases only by a small amount when the number of required sample points is doubled. This has to do with the fact that this input trace has to be convolved a couple of times. When convolving sequences, one has to deal with border effects, i.e., the effect that the resulting sequence is spoiled by the lack of data at the ‘border’ of the sequences. To avoid these border effects from influencing the sample, the length of the input fractional ARIMA trace is increased by a fixed amount; we can think of this as a safety margin. This safety margin is relatively large for small sample sizes, whereas its influence on the speed vanishes asymptotically.

Although this may sound reasonable, it is not plausible that this causes such a flat behavior in the computing time. Something else must play a role, which has to do with the convolution operation as well. This convolution was implemented with the FFT algorithm, but the lengths of the input sequences are not necessarily a power of two. When this is not the case, zeros have to be padded to make the length a power of two. It is very well possible that an increase in the requested sample size has then no influence on the length of the extended (i.e., padded) sequences. This explains that the computation time doubles instantaneously for certain sample sizes. It can not be observed in the plot that this happens instantaneously, since the graph was interpolated; the time complexity was only computed for $N = 2^4, 2^5, \dots, 2^{15}$.

Theoretically, the $\text{RMD}_{l,r}$ method is an order N algorithm. However, this can not be checked from the plot in Figure 4.18(a). It can also not be observed for the FFT-based methods and the wavelet method that their complexity is $N \log(N)$. Still, we can graphically check these assertions for the $\text{RMD}_{l,r}$ and the FFT-based methods. For this, plots that should yield straight lines (at least for large N) are most suitable, see Figure 4.19.

Instead of the time, the log time is plotted against the log sample size for the $\text{RMD}_{l,r}$ method in Figure 4.19(a), yielding indeed a straight line for ‘large’ $\log(N)$. To achieve the same for the FFT-based methods, $\log \log(N)$ is subtracted from the log time. A plot for the wavelet method is not included in Figure 4.19, because only three ‘efficient’ data points are available, since much computing time is spoiled by the padded zeros in the FFT algorithm. These efficient data points are the points at which the graph in Figure 4.18(c) ‘jumps’.

A clear straight line can be observed in Figure 4.19(a) for large N , which is also true for the Paxson and approximate circulant method in Figure 4.19(b). Note that the sample size ranges from $N = 2^4$ to $N = 2^{20}$ to make it possible to see a similar behavior for the Davies and Harte method, although it may require even larger sample sizes to make it more clear. Recall from Chapter 2 that for very large N , the approximate circulant method should be about twice as fast as the Davies and Harte method. Since the Paxson method is twice as fast as the approximate circulant method, the approximate circulant line should lie exactly in the middle between the Davies and Harte line and the Paxson line. This is not the case (yet) for $N = 2^{20}$.

Now that we have understood the shape of the time complexity graphs, we are able to draw conclusions on the accuracy-speed trade-off.

CHAPTER 5

Conclusions

In this report, several methods to simulate fractional Brownian motion were described and discussed. The three (theoretically) most promising ‘families’ of methods were evaluated:

- The $\text{RMD}_{l,r}$ method proposed by Norros *et al.* [46];
- The Paxson method proposed by Paxson [47] and a generalization of the Paxson method, the approximate circulant method;
- A wavelet-based simulation method by Abry and Sellan [2].

A fair comparison between different methods is a difficult task. Almost any method can be favoured above the others by changing the comparison criterion. In this report, the comparison was made by an error analysis and by checking theoretically or statistically a number of properties that fully characterize fractional Brownian motion: besides that the sample should be centered and Gaussian, its increments should be stationary. The property that the variance at time t should be t^{2H} was checked by estimating the Hurst parameter with various estimators.

We arrive at the following conclusions:

- So-called spectral simulation provides the theoretical foundation for the Paxson method, which clarifies many empirical observations and leads to an important improvement of the method. Moreover, a natural extension of the Paxson method, the approximate circulant method, has a direct connection with the exact Davies and Harte method. The obtained insights show that the accuracy of both methods increases as the sample size grows, and that the methods are even asymptotically exact in the sense that the finite-dimensional distributions of the approximate sample converge in probability to the corresponding distributions of the exact sample.
- These insights make it also possible to perform an error analysis for the approximate circulant method, in which the theoretical results are confirmed. The same errors can be analyzed for the $\text{RMD}_{1,2}$ method, which shows better performance.
- The covariance structure of approximate Paxson samples, approximate circulant samples and $\text{RMD}_{l,r}$ samples can numerically be calculated. Since all approximate methods under study produce zero-mean Gaussian samples, these covariances fully characterize the approximate method. While the errors in the covariances are quite big for the Paxson method (although they vanish asymptotically), the autocovariance function of the approximate circulant method is visually almost indistinguishable from the exact autocovariance function. Still, this autocovariance function does *not* show the desired hyperbolical tail behavior.
- Since the $\text{RMD}_{l,r}$ produces in general non-stationary samples, it takes more effort to study the tail behavior of the covariances. We showed how the covariances of an $\text{RMD}_{l,r}$ sample can numerically be computed for general $l \geq 0$ and $r \geq 1$. An analysis of these covariances of the $\text{RMD}_{1,2}$ and the $\text{RMD}_{3,3}$ method showed that these methods also do not have the desired decay in the covariances. This can be improved by removing some begin and end points in the approximate $\text{RMD}_{l,r}$ sample.
- The overall impression from the accuracy analysis is that only the $\text{RMD}_{1,2}$, the $\text{RMD}_{3,3}$, and the approximate circulant method are reasonable alternatives for the exact Davies and Harte method.

- The wavelet method studied in this report is disappointing, in terms of both speed and accuracy, although theoretically the time complexity is approximately $N \log(N)$.
- The $\text{RMD}_{l,r}$ method is the only studied method of which the time complexity grows linearly in the trace length. Since the extra time needed by the $\text{RMD}_{3,3}$ compared to the $\text{RMD}_{1,2}$ method is almost negligible, the $\text{RMD}_{3,3}$ method is the most interesting.
- The Paxson method is faster than the exact Davies and Harte method, while the approximate circulant method is only faster for relatively large sample sizes. As mentioned earlier, their accuracy also improves as the trace length grows. Since they are based on the FFT, these methods have time complexity $N \log(N)$.

From these observations, a final choice is readily made: the $\text{RMD}_{3,3}$ method. Moreover, it is possible to generate $\text{RMD}_{3,3}$ samples on-the-fly, which is an advantage that is not shared by the Davies and Harte method (nor by the two spectral methods). In both methods it is possible to ‘recycle’ some computations when more samples of the same size are needed, but the $\text{RMD}_{3,3}$ method stays an order N method, and the Davies and Harte an order $N \log(N)$ method.

However, in a simulation study, for instance, it may be far more desirable to know that the samples are exact. Therefore, the Davies and Harte method should be used in principle. When long traces or many runs are needed in a simulation study, it may computationally be impossible to generate exact samples. The $\text{RMD}_{3,3}$ method is then a good alternative.

Before celebrating the $\text{RMD}_{l,r}$ method too much, we have also seen in a study of the relative error in a network model that the approximate circulant method performs in this model better than the $\text{RMD}_{1,2}$ method for reasonable sample sizes. This indicates that our conclusions are not as robust as they may seem. If the approximate method is not first evaluated in the framework in which it is used, it is quite risky to just use an approximate method. This is particularly the case when rare events are simulated, e.g., loss probabilities in queues. The use of approximate methods should thus be avoided, especially since a quite fast exact algorithm exists.

Bibliography

- [1] P. ABRY, P. GONÇALVÈS, AND P. FLANDRIN, *Wavelets, spectrum analysis and $1/f$ processes*, in Wavelets and statistics, A. Antoniadis and G. Oppenheim, eds., vol. 103 of Lecture Notes in Statistics, Springer-Verlag, New York, 1995, pp. 15–30.
- [2] P. ABRY AND F. SELLAN, *The wavelet-based synthesis for fractional Brownian motion proposed by F. Sellan and Y. Meyer: remarks and fast implementation*, Applied and computational harmonic analysis, 3 (1996), pp. 377–383.
- [3] P. ABRY AND D. VEITCH, *Wavelet analysis of long range dependent traffic*, IEEE Transactions on Information Theory, 44 (1998), pp. 2–15.
- [4] S. ASMUSSEN, *Stochastic simulation with a view towards stochastic processes*. Notes of a Concentrated Advance Course at MaPhySto, Aarhus, Denmark, February 1999.
- [5] A. AYACHE AND M. S. TAQQU, *Approximating fractional Brownian motion by a random wavelet series: the rate optimality problem*. To appear in J. Fourier Anal. Appl., 2003.
- [6] J.B. BASSINGTHWAIGHTE AND G.M. RAYMOND, *Evaluating rescaled range analysis for time series*, Annals of Biomedical Engineering, 22 (1994), pp. 432–444.
- [7] ———, *Evaluation of the dispersional analysis method for fractal time series*, Annals of Biomedical Engineering, 23 (1995), pp. 491–505.
- [8] J. BERAN, *A goodness-of-fit test for time series with long range dependence*, J. R. Statist. Soc. B, 54 (1992), pp. 749–760.
- [9] ———, *Statistics for long-memory processes*, no. 61 in Monographs on statistics and applied probability, Chapman and Hall, 1994.
- [10] 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), pp. 1566–1579.
- [11] P.J. BROCKWELL AND R.A. DAVIS, *Time series: Theory and methods*, Springer-Verlag, New York, 1987.
- [12] M.J. CANNON, D.B. PERCIVAL, D.C. CACCIA, G.M. RAYMOND, AND J.B. BASSINGTHWAIGHTE, *Evaluating scaled windowed variance methods for estimating the Hurst coefficient of time series*, Physica A, 241 (1997), pp. 606–626.
- [13] K. L. CHUNG, *A Course in Probability Theory*, Academic Press, San Diego, third ed., 2001.
- [14] J.F. COEURJOLLY, *Inférence statistique pour les mouvements browniens fractionnaires et multifractionnaires*, PhD thesis, Université Joseph Fourier, Dec. 2000.
- [15] ———, *Simulation and identification of the fractional Brownian motion: a bibliographical and comparative study*, Journal of Stat. Software, 5 (2000). Available from <http://www.jstatsoft.org>.
- [16] D.R. COX, *Long-range dependence: a review*, in Statistics: an appraisal, H.A. David and H.T. David, eds., Iowa State University Press, 1984, pp. 55–74.
- [17] M. S. CROUSE AND R. G. BARANIUK, *Fast, exact synthesis of Gaussian and nonGaussian long-range dependent processes*. Submitted to IEEE Transactions on Information Theory, 1999.
- [18] I. DAUBECHIES, *Ten lectures on wavelets*, CBMS-NSF Regional Conference Series, SIAM, 1992.

- [19] R.B. DAVIES AND D.S. HARTE, *Tests for Hurst effect*, Biometrika, 74 (1987), pp. 95–102.
- [20] L. DECREUSEFOND AND N. LAVAUD, *Simulation of the fractional Brownian motion and application to the fluid queue*, in Proceedings of the ATNAC '96 conference, 1996.
- [21] B. DELYON AND A. JUDITSKY, *Estimating wavelet coefficients*, in Wavelets and statistics, A. Antoniadis and G. Oppenheim, eds., vol. 103 of Wavelets and statistics, Lecture Notes in Statistics, Springer-Verlag, New York, 1995, pp. 151–168.
- [22] A. B. DIEKER AND M. MANDJES, *On spectral simulation of fractional Brownian motion*, Probab. Engrg. Inform. Sci., 17 (2003), pp. 417–434.
- [23] C.R. DIETRICH AND G.N. NEWSAM, *Fast and exact simulation of stationary Gaussian processes through circulant embedding of the covariance matrix*, SIAM Journal Sci. Comput., 18 (1997), pp. 1088–1107.
- [24] P. DOUKHAN, G. OPPENHEIM, AND M. S. TAQQU, eds., *Theory and applications of long-range dependence*, Birkhäuser Boston Inc., Boston, MA, 2003.
- [25] K. DZHAPARIDZE AND H. VAN ZANTEN, *Optimality of an explicit series expansion of the fractional Brownian sheet*. preprint, 2003.
- [26] ———, *A series expansion of fractional Brownian motion*. preprint, 2003.
- [27] A. ERRAMILLI, R.P. SINGH, AND P. PRUTHI, *Chaotic maps as models of packet traffic*, in Proc. 14th Int. Teletraffic Cong., 6-10 June 1994, vol. 1, North-Holland Elsevier Science B.V., 1994, pp. 329–338.
- [28] P. FLANDRIN, *On the spectrum of fractional Brownian motion*, IEEE Transactions on Information Theory, 35 (1989), pp. 197–199.
- [29] ———, *Wavelet analysis and synthesis of fractional Brownian motion*, IEEE Transactions on Information Theory, 38 (1992), pp. 910–917.
- [30] J.D. GIBBONS, *Nonparametric Statistical Inference*, McGraw-Hill, Inc., 1971.
- [31] T. HIGUCHI, *Approach to an irregular time series on the basis of the fractal theory*, Physica D, 31 (1988), pp. 277–283.
- [32] J.R.M. HOSKING, *Modeling persistence in hydrological time series using fractional differencing*, Water resources research, 20 (1984), pp. 1898–1908.
- [33] C. HUANG, M. DEVETSIOTIS, I. LAMBADARIS, AND A.R. KAYE, *Fast simulation of queues with long-range dependent traffic*, Commun. statist. — Stochastic models, 15 (1999), pp. 429–460.
- [34] H. HULT, *Approximating some Volterra type stochastic integrals with applications to parameter estimation*, Stochastic Process. Appl., 105 (2003), pp. 1–32.
- [35] J. HÜSLER AND V. PITERBARG, *Extremes of a certain class of Gaussian processes*, Stochastic Process. Appl., 83 (1999), pp. 257–271.
- [36] R. JENNANE, R. HARBA, AND G. JACQUET, *Estimation de la qualité des méthodes de synthèse du mouvement brownien fractionnaire*, Traitement du Signal, 13 (1996), pp. 289–302.
- [37] I. KARATZAS AND S.E. SHREVE, *Brownian motion and stochastic calculus*, no. 113 in Graduate Texts in Mathematics, Springer-Verlag, New York, 2nd ed., 1991.
- [38] T. KÜHN AND W. LINDE, *Optimal series representation of fractional Brownian sheets*, Bernoulli, 8 (2002), pp. 669–696.
- [39] W-C. LAU, A. ERRAMILLI, J.L. WANG, AND W. WILLINGER, *Self-similar traffic generation: the random midpoint displacement algorithm and its properties*, in Proceedings of ICC '95, 1995.
- [40] W.E. LELAND, M.S. TAQQU, W. WILLINGER, AND D.V. WILSON, *On the self-similar nature of Ethernet traffic (extended version)*, IEEE/ACM Trans. on Networking, 2 (1994), pp. 1–15.
- [41] G. LINDGREN, *Lecture notes on Stationary and related stochastic processes*, Lund University, 1999. Available from <http://www.maths.lth.se/matstat/staff/georg>.

- [42] B.B. MANDELBROT AND J.W. VAN NESS, *Fractional Brownian motions, fractional noises and applications*, SIAM Review, 10 (1968), pp. 422–437.
- [43] Y. MEYER, F. SELLAN, AND M. S. TAQQU, *Wavelets, generalized white noise and fractional integration: the synthesis of fractional Brownian motion*, The Journal of Fourier Analysis and Applications, 5 (1999), pp. 465–494.
- [44] I. NORROS, *Studies on a model for connectionless traffic, based on fractional Brownian motion*, in Conf. On Applied Probability in Engineering, Computer and Communication Sciences, Paris, June 16–18, 1993.
- [45] ———, *A storage model with self-similar input*, Queueing Systems, 16 (1994), pp. 387–396.
- [46] I. NORROS, P. MANNERSALO, AND J.L. WANG, *Simulation of fractional Brownian motion with conditionalized random midpoint displacement*, Advances in Performance Analysis, 2 (1999), pp. 77–101.
- [47] V. PAXSON, *Fast, approximate synthesis of fractional Gaussian noise for generating self-similar network traffic*, Computer Communication Review, 27 (1997), pp. 5–18.
- [48] C.K. PENG, S.V. BULDYREV, M. SIMONS, H.E. STANLEY, AND A.L. GOLDBERGER, *Mosaic organization of DNA nucleotides*, Physical Review E, 49 (1994), pp. 1685–1689.
- [49] V. PIPIRAS AND M. S. TAQQU, *Fractional calculus and its connections to fractional Brownian motion*, in Doukhan et al. [24].
- [50] V. PITERBARG, *Large deviations of a storage process with fractional Brownian motion as input*, Extremes, 4 (2001), pp. 147–164.
- [51] M.B. PRIESTLEY, *Spectral analysis and time series*, vol. 1, Academic Press, 1981.
- [52] G. SAMORODNITSKY AND M. S. TAQQU, *Stable non-Gaussian random processes*, Chapman & Hall, New York, 1994.
- [53] F. SELLAN, *Synthèse de mouvements browniens fractionnaires à l'aide de la transformation par ondelettes*, Compte Rendus Acad. Sci. Paris Série I, 321 (1995), pp. 351–358.
- [54] M.S. TAQQU, *Homepage*. Available from <http://math.bu.edu/people/murad>.
- [55] M.S. TAQQU, V. TEVEROVSKY, AND W. WILLINGER, *Estimators for long-range dependence: an empirical study*, Fractals, 3 (1995), pp. 785–798.
- [56] C.R. TRAAS, H.G. TER MORSCHE, AND R.M.J. VAN DAMME, *Splines en Wavelets*, Epsilon Uitgaven, Utrecht, 2000.
- [57] 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), pp. 67–85.
- [58] W. WILLINGER, M.S. TAQQU, R. SHERMAN, AND D.V. WILSON, *Self-similarity through high-variability: statistical analysis of Ethernet LAN traffic at the source level*, IEEE/ACM Transactions on Networking, 5 (1997), pp. 71–86.
- [59] A.T.A. WOOD AND G. CHAN, *Simulation of stationary Gaussian processes in $[0, 1]^d$* , Journal of Computational and Graphical Statistics, 3 (1994), pp. 409–432.
- [60] G.W. WORNELL, *A Karhunen-Loève like expansion for $1/f$ processes via wavelets*, IEEE Transactions on Information Theory, 36 (1990), pp. 859–861.