Non-stationarity versus long-range dependence in computer network traffic measurements

ALWIN STEGEMAN
University of Groningen
Department of Mathematics
P.O. Box 800
9700 AV Groningen
The Netherlands
A.W.Stegeman@math.rug.nl

January 14, 2002

Abstract

Measurement studies of the workload in computer networks suggest that network traffic exhibits *long-range dependence* (LRD). However, it is well known that slowly decaying autocorrelations can also be the result of non-stationarities, e.g. shifts in the mean or a slowly decaying trend. We show that workload measurements in the Bellcore Ethernet LAN (see Leland et al. [23]) and some of those used in the wide area traffic study by Paxson and Floyd [27], can be modeled by an ordinary ARIMA(p,1,q) process. Since this is not a stationary process, it seems more suitable to assume that the LRD-like phenomena are caused by non-stationarities.

The author's research is supported by a Dutch Science Foundation (NWO) Grant.
AMS 2000 *subject classifications.* Primary **62M10**; secondary 60K30, 62M15.
Key words and phrases. Long-range dependence, long memory, stationarity, teletraffic, network traffic, workload, ARIMA, ARMA.

1 Introduction

In the last decade teletraffic has become a "hot topic"in the statistical community. It is common-hold belief that high-speed network traffic cannot be modeled by traditional traffic models with exponential inter-arrival times and exponentially decaying autocorrelations. Leading to this opinion is a number of empirical studies of traffic measurements, which conclude that the workload in the network exhibits long-range dependence (LRD), i.e. the sample autocorrelations show slowly decaying behavior. The most influential among these studies is due to Leland et al. [23], who provide a statistical analysis of the workload (e.g. number of bytes sent or packet arrivals per time unit) in the Ethernet LAN (Local Area Network) at Bellcore from 1989 until 1992. These measurements are also known as the *Bellcore data.* On the basis of several empirical plots (R/S, variance-time and periodogram) they conclude that the observed workload is consistent with the LRD hypothesis. Another well-known measurement study is done by Paxson and Floyd [27], who analyze WAN (Wide Area Network) traffic from the Lawrence Berkeley Laboratory (1994) and the Digital Equipment Corporation (1993-1995). They find that packet arrivals cannot be modeled by a Poisson process and that WAN workload exhibits LRD. Other empirical studies of network traffic are Fowler and Leland [15] and Beran et al. [4].

In theory, a time series model exhibiting LRD is required to be stationary. In practice, however, stationarity of an observed time series is a matter of belief. Ever since Hurst [19] discovered features resembling LRD (also called the Hurst phenomenon) in the yearly minimal water levels of the Nile River, there has been much dispute about the stationarity assumption. Boes and Salas [6] show that a realization from a sequence of independent random variables with shifts in the mean, can lead one to falsely conclude the presence of LRD on the basis of an R/S plot. See also Klemeš [21]. Another non-stationarity that can cause LRD-like phenomena is a slowly decaying trend. This is studied in Bhattacharya et al. [5] (R/S plot), Künsch [22] (periodogram) and Teverovsky and Taqqu [37] (variance-time plot). In teletraffic research it seems, however, that since Willinger et al. [38] provided a "physical explanation" for the LRD by means of the ON/OFF model with heavy-tailed transmission times (see also Heath et al. [18] and Crovella and Bestavros [11]), the option of non-stationarity has faded into the background. The only critical study we could find is Duffield et al. [12] which concerns the Bellcore data.

In this paper we study the same workload measurements as in Leland et al. [23] and Paxson and Floyd [27]. We find that the LRD-like phenomena in the sample autocorrelations and the periodogram disappear after differencing the series. In fact, for some of the measurements, ordinary ARMA(p,q) is an appropriate model for the differenced data. This means that the original workload can be modeled by an ARIMA(p,1,q) process. Since this is not a stationary model, it seems likely that the LRD-like features are caused by non-stationarities, presumably shifts in the mean.

The paper is organized as follows. In Section 2 we give several definitions of LRD and discuss briefly some exploratory methods to detect LRD in a supposedly stationary time series. Next, the question of non-stationarity versus LRD is addressed and the Bellcore data is considered as an example. In Section 3 we define ARIMA models and discuss two goodness-of-fit tests used to evaluate the fit of an observed time series to an ARMA process. It is also shown that a simulated ARIMA(1,1,1) series, when interpreted as being stationary, can lead to LRD-like phenomena. Finally, we successfully fit the Bellcore data to a non-stationary ARIMA model and thus show that serious doubts concerning the stationarity of the series are justified. In Section 4 we present an analysis of other workload measurements. Some are part of the Bellcore measurements, others were used by Paxson and Floyd [27]. Also, the famous Nile River data from Hurst [19] and a more recent trace from an ATM network (see Gogl [17]) are investigated. Section 5 provides a summary of our findings and raises the question as to whether it is reasonable to assume that there is LRD in the workload measurements.

2 Long-range dependence

2.1 Definitions

We define long-range dependence of a stationary stochastic process as follows.

Definition 2.1 *Let* $(X_t,\, t = 1, 2, \ldots)$ *be a real-valued, weakly stationary stochastic process with autocorrelation function* $\rho(k)$*, i.e.*

$$\rho(k) = \frac{\operatorname{Cov}(X_t,\, X_{t+k})}{\operatorname{Var}(X_t)}\,, \qquad k = 0, 1, 2, \ldots$$

Then X *exhibits* long-range dependence (LRD) *if*

$$\sum_{k=0}^{\infty} |\rho(k)| = \infty\,. \tag{2.1}$$

□

Instead of LRD the term *long memory* is frequently used. In view of (2.1), a process has LRD if the autocorrelations are not absolutely summable. If the autocorrelations are absolutely summable the process is said to have *short-range dependence* or *short memory.* This is true for some important classes of Markov processes and ARMA processes, which are frequently applied in time series analysis. As we see from the definition, for a process to exhibit LRD it is required to be weakly (or second-order) stationary. For a discussion on weak and strict stationarity see Brockwell and Davis [7], Section 1.3. Notice that $\operatorname{Var}(X_t) < \infty$ is required. We will assume in the sequel that the stochastic processes and time series we consider, have finite variance.

In the literature (see e.g. Beran [3] or Cox [10]), LRD is sometimes defined by describing the rate at which $\rho(k)$ decreases to zero as $k \to \infty$. Following this approach, a stationary process has LRD if

$$\rho(k) \sim c_\rho\, k^{-\beta}\,, \qquad k \to \infty\,, \tag{2.2}$$

where c_ρ is a positive constant and $\beta \in (0,1)$. Clearly, (2.2) implies (2.1). From an empirical perspective (2.2) is more suitable than (2.1) when checking for LRD in a supposedly stationary time series. Another reason why (2.2) is used to define LRD is that two very popular processes, fractional ARIMA and fractional Gaussian noise, both satisfy (2.2) (see, respectively, Beran [3], Section 2.5, and Samorodnitsky and Taqqu [30], Section 7.2.3).

An analogous way to define LRD is through the spectral density. The *spectral density* of a weakly stationary process $(X_t,\, t = 1, 2, \ldots)$ with $\sum_k \rho(k)^2 < \infty$ is defined as (see Brockwell and Davis [7], Section 4.3)

$$f(\lambda) = \frac{\sigma^2}{2\pi} \sum_{k=-\infty}^{\infty} \rho(k)\, e^{-ik\lambda}\,, \qquad \lambda \in [-\pi, \pi]\,,$$

where $\sigma^2 = \operatorname{Var}(X_t)$. Since the autocorrelations are not summable under LRD, the spectral density can have a singularity at $\lambda = 0$. One can show that (2.2) is equivalent to

$$f(\lambda) \sim c_f\, \lambda^{\beta-1}\,, \qquad \lambda \to 0\,, \tag{2.3}$$

where c_f is a positive constant depending on β and c_ρ. For a proof see Stegeman [32], Chapter 4.

2.2 Exploratory methods to detect LRD

As mentioned above, in practice one usually tries to verify (2.2) or (2.3) rather than (2.1). There are several exploratory methods to detect the presence of LRD in a time series and more sophisticated statistical methods to estimate the parameter β in (2.2) and (2.3). A detailed description can be found in Beran [3]. See also Abry and Veitch [1], who describe a wavelet method to detect LRD. For a performance analysis of various estimators we refer to Taqqu and Teverovsky [35, 36]. Here, we describe four methods which are frequently used as an exploratory tool to detect LRD in the data.

Suppose we have observed a time series $(X_t,\, t = 1, \ldots, N)$ and we believe it is stationary. One way to detect LRD would be to compute the sample autocorrelations $\hat{\rho}(k)$ and see if they satisfy (2.2) for large k. This is done by making a plot of $\log(\hat{\rho}(k))$ against $\log(k)$, also called the log-log *correlogram*. If LRD in the sense of (2.2) is present, the points in this plot should be randomly scattered around a straight line with slope $-\beta$ for sufficiently large values of k. A disadvantage of this method is that the estimate $\hat{\rho}(k)$ is unreliable for large k with respect to N (see Brockwell and Davis [7], Section 7.2).

An alternative method to detect LRD is estimation of the spectrum of X with the aim to verify (2.3). A natural estimator of the spectral density is the *periodogram* (see Brockwell and Davis [7], Section 10.3), which we define by

$$I(\lambda) = \frac{1}{2\pi\, N} \left| \sum_{t=1}^{N} X_t\, e^{-it\lambda} \right|^2 , \qquad \lambda \in [-\pi, \pi] . \tag{2.4}$$

It is common use to evaluate the periodogram at the *Fourier frequencies* $\lambda_j = 2\pi j/N$, $j = -[(N-1)/2], \ldots, [N/2]$. If LRD in the sense of (2.3) is present, a plot of $\log(I(\lambda_j))$ against $\log(\lambda_j)$ would result in an approximately straight line with slope $\beta - 1$ for small frequencies λ_j. By least squares regression a naive estimator of β can be obtained (for a discussion see Geweke and Porter-Hudak [16] and Robinson [29]). One has to take into account, however, that if the autocorrelations are not summable the periodogram fluctuates much more for frequencies tending to zero, than for short memory processes (see Beran [3], Theorem 3.8).

The third method focuses on the variance of the sample mean $\bar{X}_N$. It can be shown (see Stegeman [32], Chapter 4) that if (2.2) or, equivalently, (2.3) holds, then

$$\mathrm{Var}(\bar{X}_N) \sim c_v\, \sigma^2 N^{-\beta} , \qquad N \to \infty , \tag{2.5}$$

where c_v is a positive constant depending on c_ρ and β. For a short memory process, $\mathrm{Var}(\bar{X}_N)$ decreases to zero proportionally to N^{-1}. Given a supposedly stationary time series X we can check if (2.5) holds in the following way. The data are split into blocks of size m (with $m \ll N$). For the jth block we compute the mean value $\bar{X}_j^{(m)}$, for $j = 1, \ldots, [N/m]$. The sample variance S_m^2 of $\{\bar{X}_1^{(m)}, \ldots, \bar{X}_{[N/m]}^{(m)}\}$ is an estimator of $\mathrm{Var}(\bar{X}_m)$. By computing S_m^2 for successive values of m and plotting $\log(S_m^2)$ against $\log(m)$ LRD in the sense of (2.5) can be detected. This is also called a log-log *variance-time* plot. The resulting points should be scattered around a straight line with slope $-\beta$. If the process has short memory the slope should be -1.

The oldest and perhaps best known method to detect LRD is the R/S method. It has its origins in the field of hydrology and was first used by Hurst [19] when he discovered LRD-like features in the yearly minimal water levels of the Nile River. For a stationary process $(X_t,\, t = 1, 2, \ldots)$ with partial sums $Y_n = \sum_{t=1}^{n} X_t$ and sample variance $S_n^2 = n^{-1} \sum_{t=1}^{n} X_t^2 - n^{-2} Y_n^2$, the R/S statistic, or *rescaled adjusted range*, is given by

$$\frac{R(n)}{S(n)} = \frac{1}{S(n)} \left[\max_{0 \le t \le n} \left(Y_t - \frac{t}{n} Y_n \right) - \min_{0 \le t \le n} \left(Y_t - \frac{t}{n} Y_n \right) \right] .$$

It can be shown that if X_t is Gaussian, ergodic and (2.2) holds, then

$$n^{-H} \frac{R(n)}{S(n)} \xrightarrow{d} \xi , \qquad n \to \infty , \tag{2.6}$$

where ξ is a non-degenerate random variable (see Mandelbrot [24] and Taqqu [34]) and $H = 1 - \beta/2$. The parameter H is the so-called *Hurst parameter* and is frequently used as a measure of the strength of LRD in the data (H close to 1 corresponds to a strong presence of LRD). For various short memory processes (2.6) holds with $H = 1/2$ (see Mandelbrot [24] and Feller [14]). For fractional Gaussian noise and fractional ARIMA (with Gaussian innovations), we have

$$E(R(n)/S(n)) \sim c_r \, n^H , \qquad n \to \infty , \tag{2.7}$$

where c_r is a positive constant. For an observed time series $(X_t, t = 1, \ldots, N)$ we can try to verify (2.7) as follows. Partition the series in $[N/m]$ blocks of size m. Then, for each n, compute $R(m_i, n)/S(m_i, n)$, starting at points $m_i = i\,m + 1$, $i = 0, 1, \ldots$, such that $m_i + n \le N$. For values of n smaller than m, we get $[N/m]$ different estimates of $R(n)/S(n)$. For values of n approaching N, we get fewer values, as few as 1 when $n \ge N - m$. Next, plot $\log(R(m_i, n)/S(m_i, n))$ against $\log(n)$ and get, for each n several values on the plot. For large n, the points should lie around a straight line with slope H.

2.3 Application to the Bellcore data

We apply the four methods described above to a well-known workload series measured on the Ethernet LAN at the Bellcore company in August 1989. The time series, which we refer to as **BC-pAug**, represents the number of packet arrivals per second. It has length $N = 3142$ and, for now, we assume it to be stationary. In Figure 1 the series, its sample ACF and its log-periodogram are depicted. The slowly decaying ACF and the peak of the periodogram at zero suggest the presence of LRD. In Figure 2 the corresponding log-log correlogram, log-log periodogram, log-log variance-time and log-log R/S plots are drawn. Least squares regression in these plots yields naive estimates of β and $H = 1 - \beta/2$ (see Table 1). In the variance-time and R/S plots also the lines representing the short memory case are depicted (with slopes -1 and $1/2$, respectively). The estimates of β and H are all in the range of LRD and do not heavily depend on the method used to obtain them. Although, it is natural that the estimates obtained using the log-log correlogram or the log-log periodogram are rather sensitive to the number of points included in the regression. Therefore, only the variance-time and R/S methods will be used in the sequel.

method	**estimate β**	**estimate H**
log-log correlogram	0.35	0.83
log-log periodogram	0.33	0.83
log-log variance-time	0.32	0.84
log-log R/S	0.24	0.88

Table 1: Estimates of the strength of LRD in **BC-pAug** using four naive methods.

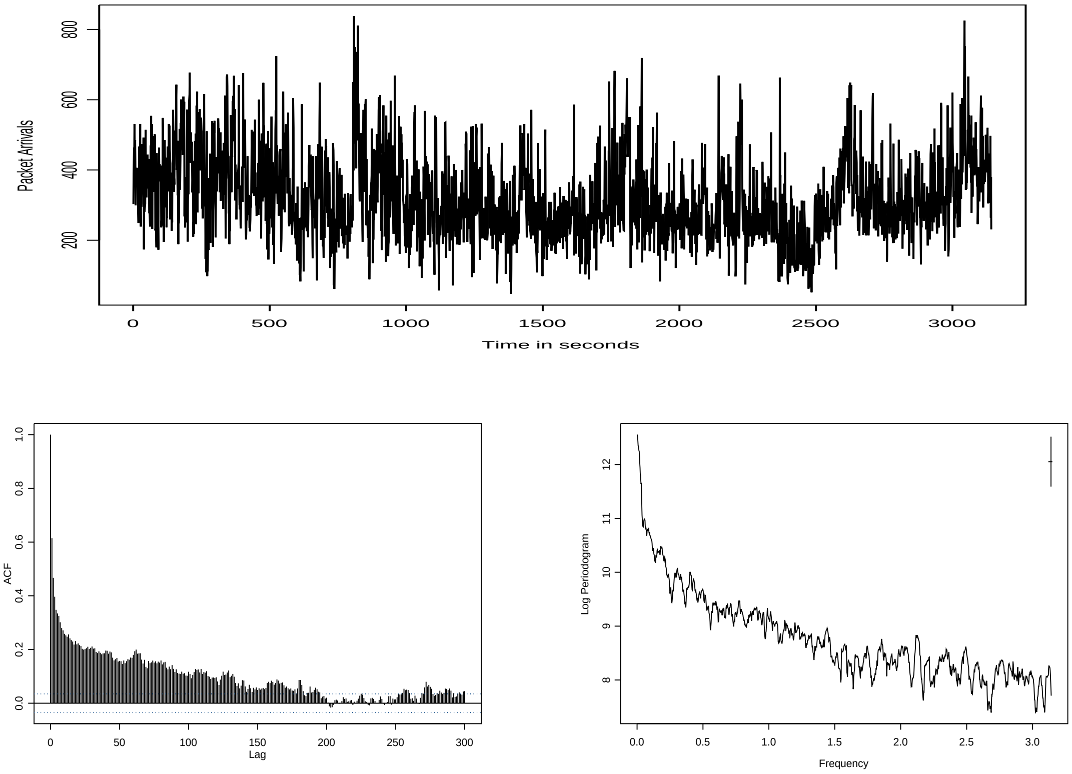

Figure 1: Time series **BC-pAug** (top) with sample autocorrelations (bottom-left) and smoothed log-periodogram (bottom-right).

2.4 LRD or non-stationarity?

The question as to whether an observed time series can be considered as a realization from a stationary stochastic process is more philosophical than mathematical in nature. A stationary stochastic process is usually thought of as a process whose underlying distributional characteristics do not change over time. This concept has been made mathematically precise through the definitions of strict and weak stationarity (see e.g. Brockwell and Davis [7], Section 1.3). There is, however, no general statistical test available for the hypothesis of stationarity of a real-life time series. In practice, the assumption of stationarity is often justified by first fitting the data to a stationary model and then performing goodness-of-fit tests. Also, additional information can be used to decide whether it is reasonable to assume stationarity. However, as mentioned by Klemeš [21], any finite time series can be equally likely regarded as coming from a stationary or non-stationary model. Deviations from stationarity like trends, cycles or level shifts can be explained as "local non-stationarities" in the first case or as "model characteristics" in the second. An analogous reasoning applies to a time series with no apparent non-stationarities. In this sense, the matter of stationarity or non-stationarity becomes one of personal belief.

From the definition of LRD (see (2.1)), it is clear that making a statement concerning the presence of LRD in a time series involves the assumption of stationarity. Usually, diagnostic plots as in Figure 2 are used as indicators of LRD. It is well-known, however, that non-stationarities

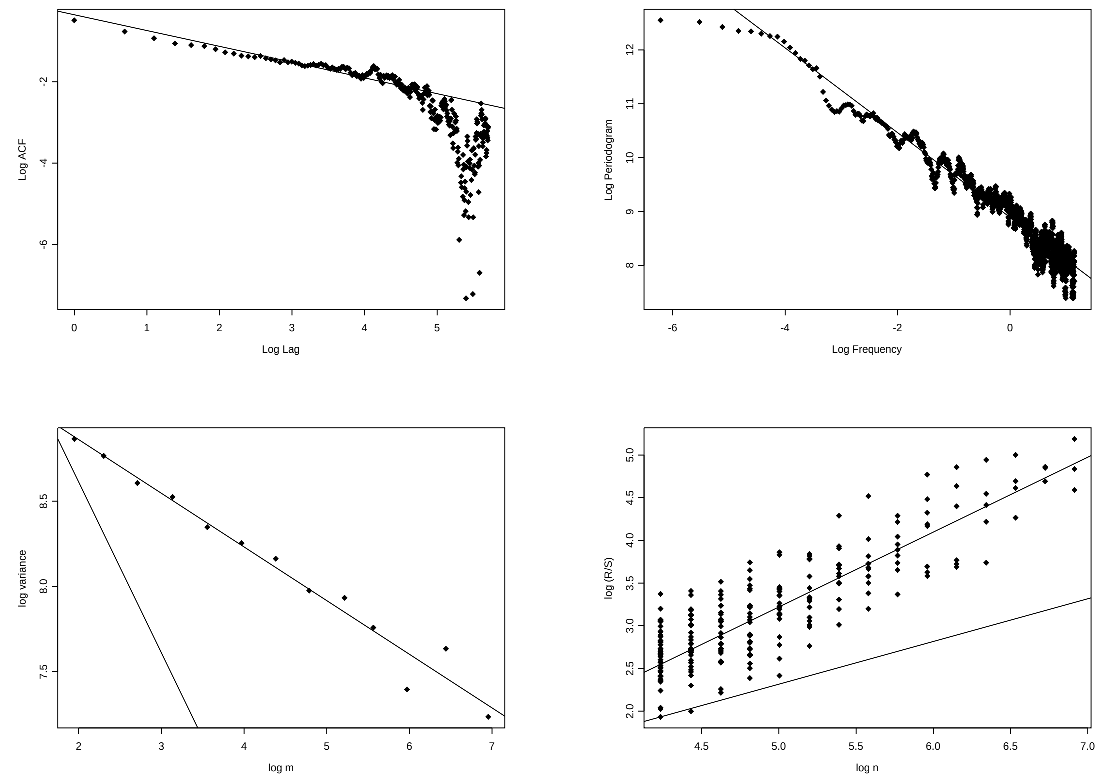

Figure 2: Exploratory plots to detect LRD in **BC-pAug**: log-log correlogram (top-left), log-log periodogram (top-right), log-log variance-time (bottom-left) and log-log R/S (bottom-right). Least squares regression yields heuristic estimates of β and H (see also Table 1). In the variance-time and R/S plot the lines representing the short memory case are depicted (with slopes -1 and $1/2$, respectively).

like shifts in the mean or a slowly decreasing trend can cause LRD-like features in these plots (see Klemeš [21], Boes and Salas [6], Bhattacharya et al. [5] and Teverovsky and Taqqu [37]). A way to distinguish between non-stationarities and LRD is to estimate the Hurst parameter H for various parts of a time series. If the changes in the estimated value of H are "reasonably large", it is not plausible that the series exhibits LRD (see e.g. Duffield et al. [12]). Notice that this procedure is not statistically sound, since confidence bands for the estimated value of H are not available. As an example, we consider four 500-second parts of **BC-pAug**. The parts are chosen such that they show significant differences in the qualitative behavior of **BC-pAug**. In Table 2 we give the estimated values of H based on the variance-time and R/S plots. It is evident that almost all estimates are significantly smaller than the overall values of 0.84 and 0.88, respectively. The estimates obtained from the R/S plot are very close to $H = 1/2$, which is the short memory case. If the variance-time method is used the estimated values increase as the behavior of **BC-pAug** becomes more wild and "less stationary". It is worthwhile mentioning, however, that the wavelet based estimator of Abry and Veitch [1] does find consistent values of $H \approx 0.8$ for different segments of **BC-pAug**. The fact that the estimates in Table 2 differ from those obtained from the whole series may, therefore, also be the result from a lack of robustness of the variance-time and R/S estimation procedures. This would also explain the large differences between the estimated values obtained by the variance-time and R/S methods.

part	log-log variance-time	log-log R/S
51-550	0.64	0.49
851-1350	0.69	0.54
1601-2100	0.79	0.53
2643-3142	0.87	0.51

Table 2: Estimates of the Hurst parameter H for four 500-second parts of **BC-pAug**.

3 ARMA and ARIMA processes

3.1 Definitions

The class of ARMA processes is the class of models most frequently applied for time series that exhibit no apparent deviations from stationarity and have rapidly decreasing autocorrelations. Before we define ARMA processes we introduce the following notation. Let $(Z_t,\, t = 0, \pm 1, \pm 2, \ldots)$ be a sequence of uncorrelated and identically distributed random variables with mean 0 and variance σ_Z^2. The sequence (Z_t) is called *white noise*. For non-negative integers p and q, we define the polynomials ϕ and θ by

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p \quad \text{and} \quad \theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q\,.$$

The *backward shift operator* B is defined by

$$B^j X_t = X_{t-j}\,, \quad j = 0, 1, 2, \ldots$$

The process $(X_t,\, t = 0, \pm 1, \pm 2, \ldots)$ is an ARMA(p,q) process if (X_t) is stationary and if for every t

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}\,. \tag{3.1}$$

The ARMA processes are a class of parametric processes defined by the set of linear difference equations (3.1). Equation (3.1) can be compactly written as $\phi(B)X_t = \theta(B)Z_t$. The name ARMA is an abbreviation of *autoregressive moving average*, the polynomial ϕ being the autoregressive part and θ the moving average part of the process. The parameters p and q are the *order* of the process. Notice that an ARMA process has mean zero. We say that (X_t) is an ARMA(p,q) process with mean μ if $(X_t - \mu)$ satisfies (3.1). If the polynomials ϕ and θ have no common zeros and ϕ has no zeros inside or on the unit circle in the complex plane, the process (X_t) is called *causal.* A causal ARMA process can be written as an infinite moving average $X_t = (\theta(B)/\phi(B))Z_t$, which is the unique stationary solution to (3.1). An ARMA process has short memory in the sense that the autocorrelation function $\rho(k)$ satisfies

$$|\rho(k)| \le b\, a^k\,, \tag{3.2}$$

for some constants $b \in (0, \infty)$ and $a \in (0, 1)$. Clearly, (3.2) guarantees absolutely summable autocorrelations. More properties of ARMA process can be found in Brockwell and Davis [7], Chapter 3.

The ARMA class is very suitable for modeling stationary short memory time series. In practice, however, apparent stationarity of a time series is not always observed. Often, transformations are applied to "make the series look more stationary". A frequently applied method is differencing, i.e. instead of the original series (X_t) one considers

$$(1-B)X_t = X_t - X_{t-1}\,.$$

This will remove a global linear trend. Differencing twice, i.e. $(1-B)^2X_t$, will remove second degree polynomials. In practice the number of times differencing required to get an apparent stationary series is often not larger than one or two. This is because, on a finite interval, many functions can be well approximated by polynomials of a reasonably low degree.

A process that, after differencing finitely many times, reduces to an ARMA process, is called an ARIMA process (autoregressive integrated moving average). An ARIMA process that is ARMA(p,q) after d times differencing is denoted by ARIMA(p,d,q). Here, d is a non-negative integer and is called the *order of differencing*. Hence, if (X_t) is ARIMA(p,d,q) then $(1-B)^dX_t$ is ARMA(p,q). By definition, ARIMA(p,0,q) is ARMA(p,q). Notice that if $d \geq 1$ the ARIMA process is not stationary.

3.2 Two goodness-of-fit tests

Here we describe two methods for testing whether a time series is well modeled by a specified ARMA process. We assume that the orders p and q have been selected and that σ_Z^2 and the parameters of the polynomials ϕ and θ have been estimated by Gaussian maximum likelihood (for details, see Brockwell and Davis [7], Chapters 8 and 9).

The first method considers the sample autocorrelations of the estimated residuals. It can be shown (see Brockwell and Davis [7], Example 7.2.1) that the sample autocorrelations $\hat{\rho}_Z(k)$, $k \geq 1$, of an iid sequence $Z_1, \ldots, Z_N$ with $EZ_t^2 < \infty$ are for large N approximately iid with distribution $N(0, 1/N)$. Assuming that our fitted ARMA model is generated by an iid white noise sequence, the same approximation should be valid for the sample autocorrelations of the estimated residuals. Then the statistic $N\sum_{j=1}^k \hat{\rho}_Z^2(j)$ should be approximately chi-squared with k degrees of freedom. However, since the estimated residuals are a function of the maximum likelihood estimators of the parameters of the ARMA process, they do not constitute an iid sequence. This results in a reduction of the degrees of freedom by $p+q$ (see Brockwell and Davis [7], Section 9.4). Hence, a goodness-of-fit test statistic for the estimated ARMA model is

$$Q_Z(k) = N\sum_{j=1}^{k} \hat{\rho}_Z^2(j)\,, \tag{3.3}$$

which is approximately chi-squared with $k-p-q$ degrees of freedom. The adequacy of the model is therefore rejected at level α if

$$Q_Z(k) > \chi^2_{1-\alpha}(k-p-q)\,.$$

This test is known as the *Portmanteau test*. The number of lags k should not be chosen too large with respect to the length of the time series, since, for large lags, the sample autocorrelations are unreliable estimates.

The second method compares the periodogram $I(\lambda)$ of the time series with the theoretical spectral density $f(\lambda)$ of the estimated ARMA model. The spectral density of an ARMA process is given by

$$f(\lambda) = \frac{\sigma_Z^2}{2\pi}\frac{|\theta(e^{-i\lambda})|^2}{|\phi(e^{-i\lambda})|^2}\,, \qquad \lambda \in [-\pi, \pi]\,.$$

The idea, due to Bartlett [2], is to consider partial sums of the ratios $I(\lambda_j)/f(\lambda_j)$ for the positive Fourier frequencies λ_j. Self-normalization results in the statistics

$$T_h = \frac{\sum_{j=1}^{[Nh/2]} I(\lambda_j)/f(\lambda_j)}{\sum_{j=1}^{[N/2]} I(\lambda_j)/f(\lambda_j)}, \qquad h = 1/[N/2], 2/[N/2], \ldots, 1\,.$$

It can be shown that, for large N, the process $\sqrt{N/2}\,(T_h - h)$ is approximately a standard Brownian bridge on $[0,1]$ (see Priestley [28]). As a goodness-of-fit statistic one can take

$$T^* = \max_{1\le h\le [N/2]} \sqrt{N/2}\,|T_h - h|\,, \tag{3.4}$$

which has the same limit distribution as the Kolmogorov-Smirnov statistic. The 95 and 99 percent quantiles of T^* are 1.36 and 1.63, respectively (see Shorack and Wellner [31], Table 1 on p. 143). For an overview of goodness-of-fit tests in the spectral domain we refer to Priestley [28], Section 6.2.6.

3.3 ARIMA models and LRD

Here we indicate that a simulated ARIMA series, generated by a non-stationary ARIMA model, can lead to LRD-like phenomena in the sample autocorrelation, periodogram, variance-time and R/S plots. We simulated an ARIMA(1,1,1) model with $\phi_1 = 0.4$, $\theta_1 = -0.95$ and $\sigma_Z^2 = 8000$. The parameter values correspond to the estimated ARIMA model for **BC-pAug** in Section 3.4. The series length is $N = 3000$. In Figure 3 the simulated ARIMA series together with its sample autocorrelations, smoothed log-periodogram, log-log variance-time and log-log R/S plots are depicted. The autocorrelations show slow decay, the periodogram has a peak at zero and the variance-time and R/S plots yield a least squares line which differs significantly from the short memory case. The estimates of the Hurst parameter H in the latter two plots are 0.88 and 1.01, respectively. Hence, if the simulated series is interpreted as being stationary, these four exploratory methods are likely to yield the conclusion that LRD is present in the data.

3.4 Fitting BC-pAug to an ARIMA model

In Section 2.4 we discussed the stationarity of **BC-pAug**. Here, we show that ARIMA(3,1,1) is an appropriate model for **BC-pAug**. Combined with the observation that a non-stationary ARIMA process can exhibit LRD-like phenomena (see Section 3.3), this is an argument in favor of the point of view that the LRD features of **BC-pAug** are actually caused by non-stationarities.

In Figure 4 it can be seen that the sample autocorrelations and the smoothed log-periodogram of the (first order) differences of **BC-pAug** show no sign of LRD at all. The peak at zero of the periodogram disappeared and beyond lag 5 almost all autocorrelations do not differ significantly from zero. From the shape of the periodogram we may conclude that, in the differenced series, low frequencies do not play a significant role while high frequencies are all almost equally important. Since low frequencies correspond to long cycle periods this indicates their complete absence in the differenced series. Sometimes, also the autocorrelations of the absolute and squared values of the series are computed. This procedure is used for the analysis of financial time series in order to detect non-linearities in the series, see e.g. Embrechts et al. [13], Section 8.4. If they are outside the 95 percent confidence bounds at large lags, this is seen as an indication that LRD is present in $(|X_t|)$ or (X_t^2) (this could, however, also be caused by non-stationarities; see Mikosch and Stărică [25]). However, from Figure 4 it is clear that these autocorrelations are also small. Notice that the

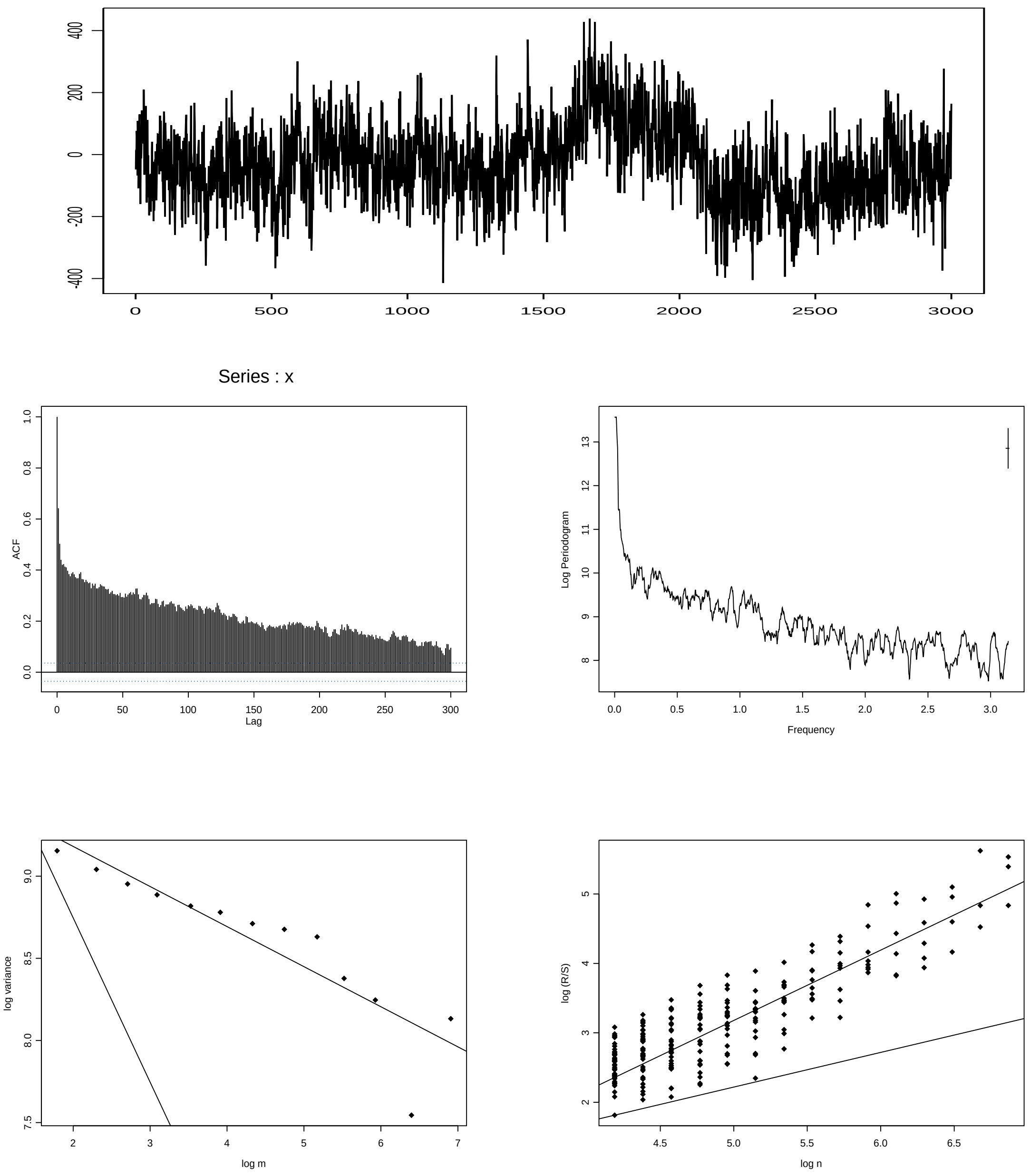

Figure 3: Simulated ARIMA(1,1,1) series (top) with sample autocorrelations (middle-left), smoothed log-periodogram (middle-right), log-log variance-time (bottom-left) and log-log R/S (bottom-right) plots.

variance-time and R/S methods are useless to check whether $H \approx 1/2$, i.e. whether the differenced sequence has short memory, since the telescopic sum property of the differences results in artificially small cumulative sums.

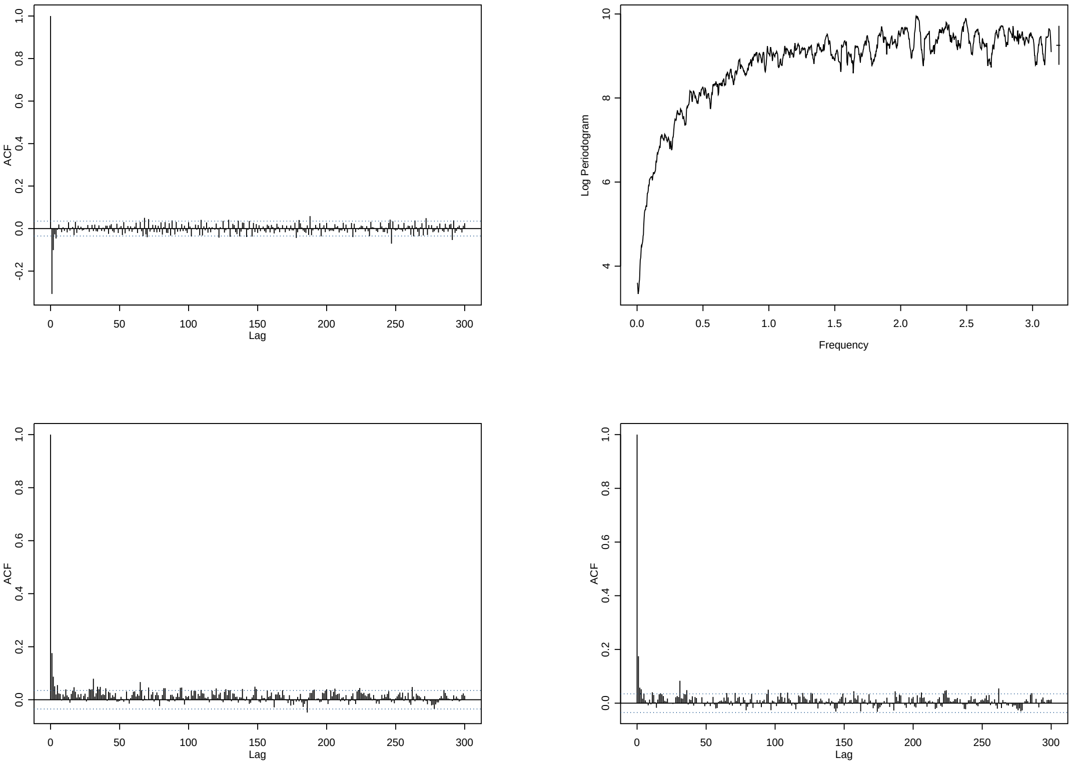

Figure 4: Sample autocorrelations (top-left) and smoothed log-periodogram (top-right) of the differences of **BC-pAug**. Bottom-left and bottom-right are the sample autocorrelations of the absolute and squared values of the differences, respectively.

Using Gaussian maximum likelihood in the statistical package `Splus`, we fitted the differences of **BC-pAug** to an ARMA(3,1) model. The orders $p = 3$ and $q = 1$ are chosen using the Akaike Information Criterion (see Brockwell and Davis [7], Section 9.2). However, orders $(p, q) = (1, 1)$ until $(5, 1)$ all result in reasonably good fits. The estimated values of the parameters are $\hat{\phi}_1 = 0.42$, $\hat{\phi}_2 = 0.05$, $\hat{\phi}_3 = 0.01$, $\hat{\theta}_1 = -0.94$ and $\hat{\sigma}^2_Z = 7703$. In Figure 5 the sample autocorrelations of the residuals and a qq-plot comparing the quantiles of the residuals with those from the density $f(x) = (\delta/2)\exp\{-\delta|x|\}$, with $\delta = 1/56$, are shown. From these plots it is reasonable to say that the residuals are uncorrelated and have roughly exponential tails. The value of the Portmanteau statistic $Q_Z(k)$ (see (3.3)) is smaller than the critical value $\chi^2_{0.95}(k-4)$ for lags k up to 300 (for $k = 300$ the P-value is still 0.12). The Bartlett statistic T^* (see (3.4)) has value 0.53, which is much smaller than the critical value of 1.36 at the 5 percent level. Hence, we may conclude that ARMA(3,1) is an appropriate model for the differences of **BC-pAug**. This implies, of course, that **BC-pAug** itself can be modeled by an ARIMA(3,1,1) process.

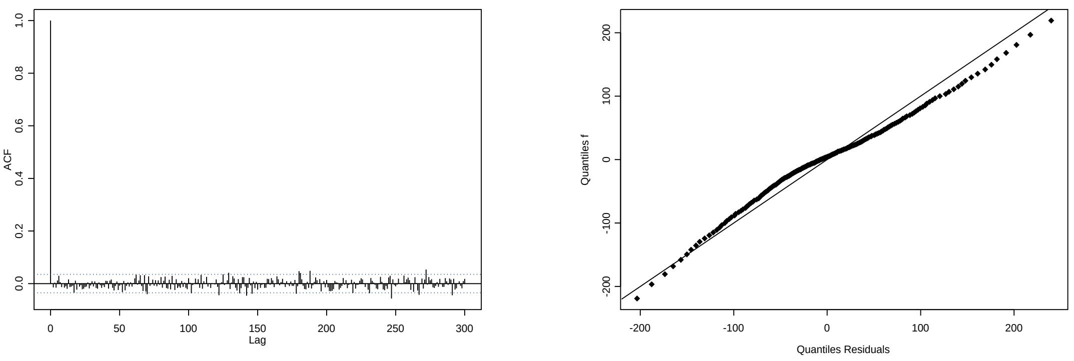

Figure 5: Sample autocorrelations (left) and a qq-plot (right) of the residuals from the ARIMA(3,1,1) fit to **BC-pAug**. In the qq-plot the quantiles of the residuals are compared to those of the density $f(x) = (\delta/2)\exp\{-\delta|x|\}$, with $\delta = 1/56$.

4 Analysis of other workload measurements

In this section we consider other workload series which are studied in Leland et al. [23] and Paxson and Floyd [27] as well as more recent traces from an ATM network (Gogl [17]) and the famous Nile River data from Hurst [19]. Each of these series is believed to exhibit LRD. Below, we investigate as to whether evidence for the alternative hypothesis of non-stationarities can be found. In the spirit of Section 3, our strategy will be to show that the data can be modeled by an appropriate ARIMA process. For all series the sample autocorrelations of the (first order) differences and its absolute and squared values are small at large lags.

Bellcore data

The Bellcore data included in the Internet Traffic Archive [20] consists of four traces: **BC-pAug** and **BC-pOct** contain internal Bellcore traffic and **BC-OctExt** and **BC-OctExt4** contain external traffic, i.e. packet streams between Bellcore and the rest of the world. Each trace consists of timestamps and packet sizes of the first 1 million packet arrivals of a longer trace started in either August or October 1989. Packet sizes are between 64 and 1518 bytes, as imposed by the Ethernet protocol. For a detailed description of the traces we refer to Fowler and Leland [15] and Leland et al. [23]. For an introduction into computer networks see Tanenbaum [33]. We created a workload series by computing either the number packets or the number of bytes that arrive per time unit, e.g. seconds. The traces of external traffic **BC-OctExt** and **BC-OctExt4** last for about 34 and 21 hours, respectively. They exhibit large peaks and **BC-OctExt** contains a clear diurnal cycle. Therefore, we do not regard these traces stationary (or ARIMA) and will not perform a detailed analysis.

Of the traces **BC-pAug** and **BC-pOct** we consider the packet and byte arrivals per second. The trace **BC-pOct** lasts for 1759 seconds. In Figure 6 **BC-pOct** is plotted (packets per second). There has been some dispute about the level shift that occurs between 1000 and 1200 seconds in the trace. According to Duffield et al. [12] this is a clear indication of non-stationarity. Their conclusion is based upon a comparison of the variance-time plots of the whole series and the segments 200-1000 and 1200-1759. In Table 3 the corresponding estimates of H can be found. As in the case of **BC-pAug** (see Table 2) the estimated values are smaller for the separate segments. However, Abry and Veitch [1], using their wavelet based estimator, find a consistent value of $H \approx 0.8$.

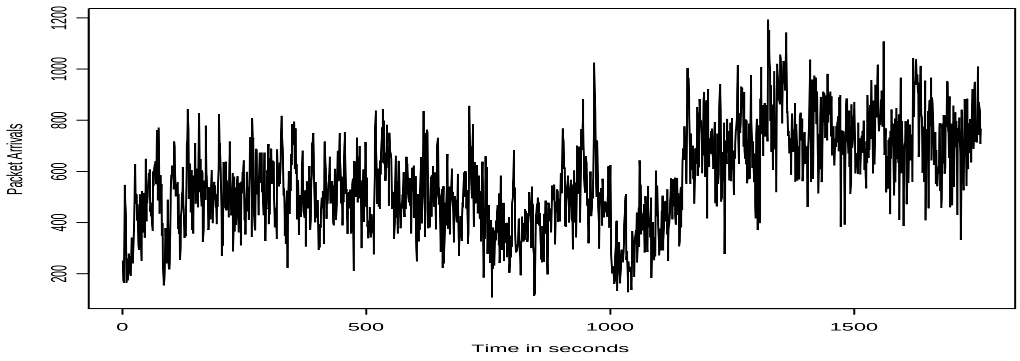

Figure 6: Packet arrivals per second in **BC-pOct**.

part	log-log variance-time	log-log R/S
1-1759	0.96	0.83
200-1000	0.80	0.62
1200-1759	0.64	0.52

Table 3: Estimates of the Hurst parameter H for two segments of **BC-pOct**.

In Table 4 the estimated ARIMA models for the series **BC-pAug** and **BC-pOct** can be found. Since the estimated values of the parameters ϕ_j, $j \geq 2$, were smaller than 0.1 in absolute value, we have only included $\hat{\phi}_1$. Notice that the values of $\hat{\phi}_1$ and $\hat{\theta}_1$ are very similar for the four workload series. The fit is evaluated by the Portmanteau and Bartlett goodness-of-fit tests. The number of lags k included in the Portmanteau statistic $Q_Z(k)$ is chosen such that for larger lags the test fails at the 5 percent level (or k has reached $N/10$, beyond which the sample autocorrelations are less reliable). As we see, for both packets and bytes per second an appropriate ARIMA model can be found.

Wide Area Traffic

In their study of wide area traffic Paxson and Floyd [27] use traces from the Digital Equipment Corporation (**DEC-PKT**) and the Lawrence Berkeley Laboratory (**LBL-PKT**), obtained in March 1995 and January 1994, respectively. Each of these traces consists of all wide area packets during one hour. Here, we consider the number of packet arrivals per second in **DEC-PKT-1** through **DEC-PKT-4** and **LBL-PKT-4** and **LBL-PKT-5**. Also, two hours' worth of TCP traffic between the Lawrence Berkeley Laboratory and the rest of the world, **LBL-TCP-3**, is included. The data can be obtained through the Internet Traffic Archive [20]. We will refer to **DEC-PKT**, **LBL-PKT** and **LBL-TCP-3** as **DEC**, **LBL** and **LBL-3**, respectively. Paxson and Floyd [27] find evidence for LRD in these series using a variance-time plot.

In Figure 7 the sample autocorrelations and periodogram of **DEC-1** are shown. From the two peaks in the periodogram (approximately at frequencies 0.2 and 0.4) and the periodic behavior of the autocorrelations we may conclude that periodicities are present in **DEC-1**. Hence, the assumption of stationarity becomes rather doubtful. The same holds for **DEC-n**, $n = 2, 3, 4$. From Table 4 we can conclude that the fit of an estimated ARIMA model and these series is not very good. Especially the Portmanteau test gives poor results. In our opinion, however, Figure 7 shows that non-stationarity rather than LRD is present in **DEC**.

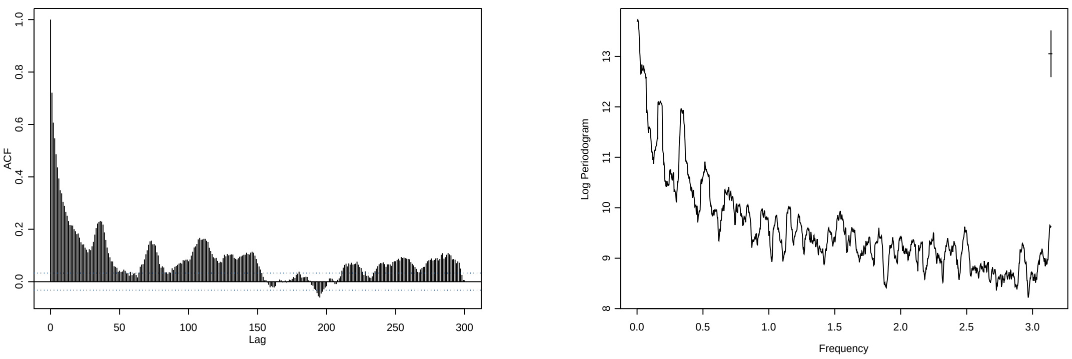

Figure 7: Sample autocorrelations and smoothed log-periodogram of **DEC-1**.

For **LBL-4** and **LBL-5** an appropriate ARIMA model can be found, although the order is higher than for the series analyzed so far. The fit of the estimated ARIMA model and **LBL-3** is less good. For $k < 200$ the Portmanteau test fails, but for $k \in [201, 402]$ the P-value is larger than 0.05. The Bartlett statistic is close to the critical value of 1.36. Considerably better results are obtained if only half of **LBL-3** is considered (which is still 3600 seconds long). In Table 4 the first and second half of **LBL-3** are denoted by **LBL-3(1)** and **LBL-3(2)**, respectively.

Munich University ATM data

Gogl [17] considers the traffic through an ATM link connecting the Technical University of Munich to the German Broadband Research Network. In both directions, called RX and TX, the number of passing ATM cells is measured in time units of 2 seconds. Here we consider two series of such measurements, **ATM-RX** and **ATM-TX**, both from March 1998. The length of each sequence is 43200, corresponding to 24 hours. Since a clear diurnal cycle is present we restricted the analysis to the busy hours: 17:00 until 18:00 for **ATM-RX** and 12:30 until 13:30 for **ATM-TX**. From Table 4 we see that both series (also called **ATM-RX** and **ATM-TX**) can be modeled by an ARIMA process.

Nile River data

The last time series we consider is also the most famous one: the yearly minimal water levels of the Nile River at the Roda Gauge near Cairo for the years $622 - 1281$. In 1951 Hurst [19] was the first to observe the LRD-like effect in an R/S plot of this series. We obtained the data from Beran [3], Section 12.2. Again, Table 4 shows a nice fit with the estimated ARIMA model.

trace	(p,q)	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\sigma}_Z^2$	**Portmanteau**	**Bartlett**
BC-pAug (pa)	$(3,1)$	0.42	$*$	-0.94	-	$7.7 \cdot 10^3$	0.12 ($k = 300$)	0.53
BC-pAug (by)	$(2,1)$	0.38	$*$	-0.94	-	$4.6 \cdot 10^9$	0.15 ($k = 75$)	0.52
BC-pOct (pa)	$(4,1)$	0.41	$*$	-0.95	-	$12.8 \cdot 10^3$	0.08 ($k = 55$)	0.54
BC-pOct (by)	$(4,1)$	0.35	$*$	-0.95	-	$8.3 \cdot 10^9$	0.19 ($k = 50$)	0.56
DEC-1 (pa)	$(3,1)$	0.49	$*$	-0.93	-	$18.2 \cdot 10^3$	0.06 ($k = 12$)	0.66
DEC-2 (pa)	$(3,1)$	0.46	$*$	-0.96	-	$21.8 \cdot 10^3$	0.11 ($k = 8$)	0.71
DEC-3 (pa)	$(4,1)$	0.47	$*$	-0.96	-	$22.5 \cdot 10^3$	0.14 ($k = 23$)	0.48
DEC-4 (pa)	$(3,1)$	0.38	$*$	-0.93	-	$25.7 \cdot 10^3$	0.11 ($k = 15$)	1.15
LBL-3 (pa)	$(4,1)$	0.47	$*$	-0.94	-	$4.2 \cdot 10^3$	0.06 ($k = 402$)	1.24
LBL-3(1) (pa)	$(3,1)$	0.43	$*$	-0.93	-	$4.3 \cdot 10^3$	0.15 ($k = 359$)	0.75
LBL-3(2) (pa)	$(4,1)$	0.43	$*$	-0.91	-	$4.0 \cdot 10^3$	0.88 ($k = 359$)	0.73
LBL-4 (pa)	$(5,2)$	-0.28	0.30	0.29	0.62	$5.0 \cdot 10^3$	0.28 ($k = 359$)	0.59
LBL-5 (pa)	$(7,2)$	-0.26	0.45	0.22	0.72	$3.5 \cdot 10^3$	0.23 ($k = 359$)	0.51
ATM-RX (ce)	$(2,1)$	0.34	$*$	-0.84	-	$3.8 \cdot 10^7$	0.55 ($k = 179$)	0.50
ATM-TX (ce)	$(5,1)$	0.46	$*$	-0.92	-	$4.1 \cdot 10^7$	0.82 ($k = 179$)	0.39
Nile	$(1,1)$	0.35	-	-0.91	-	$4.9 \cdot 10^3$	0.47 ($k = 66$)	0.56

Table 4: Parameter estimates and goodness-of-fit of estimated ARIMA(p,1,q) models for various workload series; (pa)=packets per second, (by)=bytes per second, (ce)=cells per 2 seconds. For $q = 1$, the estimates $\hat{\phi}_j$, $j \geq 2$, are all smaller than 0.1 in absolute value (in this case a $*$ represents $\hat{\phi}_2$). The value of the Bartlett test statistic T^* (see (3.4)) is given, as well as the P-value of the Portmanteau test (see (3.3)). The number of lags k included in the Portmanteau statistic $Q_Z(k)$ is chosen such that for larger lags the test fails (or k has reached $N/10$, beyond which the sample autocorrelations are less reliable). Both tests are performed at the 5 percent level, the critical value of T^* being 1.36.

5 Conclusion

In this paper we analyzed workload measurements of computer networks, which are believed to exhibit LRD. Theoretically, a stochastic process with LRD must be stationary. In practice, however, stationarity and non-stationarity cannot be distinguished without additional information or assumptions on the underlying probability laws. It is known that non-stationarities like shifts in the mean and a slowly decreasing trend can also cause LRD-like features in the observed series. We concluded that most of the workload series can be modeled by a non-stationary ARIMA(p,1,q) process.

This raises the question as to whether it is reasonable to assume that there is LRD in the workload measurements. A computer network is a rather complicated environment in which sending hosts compete for bandwidth (i.e. they are not independent), congestion control algorithms adapt sending rates to measured queue lengths and delays and routers try to find the fastest route for packets to travel. Workload characteristics may also change when a new application is started or a running one is terminated (this can be seen in the Bellcore data, in which different packet sizes are linked to different applications: the arrival process of packets with the same size does not look very stationary). Finally, the number of active connections has an influence on traffic characteristics, see e.g. Cao et al. [8, 9]. These observations support the opinion that stationarity of the workload in a computer network, even for short time periods, is rather unlikely.

Acknowledgment. I would like to thank Thomas Mikosch for his statistical comments and suggestions, Helmut Gogl for kindly supplying his ATM data, Sietse Achterop for information concerning Ethernet, Patrice Abry for sending me the wavelet papers and Titus van 't Veer and Marius de Vink for helping to transfer the data sets to my computer.

References

[1] Abry, P. and Veitch, D. (1998) Wavelet analysis of long-range dependent traffic. *IEEE Transactions on Information Theory* **44**, 2–15.

[2] Bartlett, M.S. (1955) *An Introduction to Stochastic Processes with Special Reference to Methods and Applications, First Edition.* Cambridge University Press.

[3] Beran, J. (1994) *Statistics for Long-Memory Processes.* Chapman & Hall, New York.

[4] Beran, J., Sherman, R., Taqqu, M.S. and Willinger, W. (1995) Long-range dependence in variable-bit-rate video traffic. *IEEE Transactions on Communications* **43**, 1566–1579.

[5] Bhattacharya, R.N., Gupta, V.K. and Waymire, E. (1983) The Hurst effect under trends. *Journal of Applied Probability* **20**, 649–662.

[6] Boes, D.C. and Salas, J.D. (1978) Nonstationarity of the mean and the Hurst phenomenon. *Water Resources Research* **14**, 135–143.

[7] Brockwell, P.J. and Davis, R.A. (1991) *Time Series: Theory and Methods, Second Edition.* Springer-Verlag, New York.

[8] Cao, J., Cleveland, W.S., Lin, D. and Sun, D.X. (2001) On the nonstationarity of Internet traffic. *Proceedings of the ACM SIGMETRICS 2001 Conference*, MIT, Cambridge (U.S.A.), 102–112.

[9] Cao, J., Cleveland, W.S., Lin, D. and Sun, D.X. (2001) The effect of statistical multiplexing on Internet packet traffic: theory and empirical study. Technical Report, Bell Labs, Murray Hill, New Jersey.

[10] Cox, D.R. (1984) Long-range dependence: a review. In: *Statistics: An Appraisal*, David, H.A., David, H.T., editors, Iowa State University Press.

[11] Crovella, M. and Bestavros, A. (1996) Self-similarity in world wide web traffic, evidence and possible causes. *Performance Evaluation Review* **24**, 160–169.

[12] Duffield, N.G., Lewis, J.T., O'Connell, N., Russell, R. and Toomey, F. (1994) Statistical issues raised by the Bellcore data. *Proceedings of the 11th IEE UK Teletraffic Symposium*, Cambridge.

[13] Embrechts, P., Klüppelberg, C. and Mikosch T. (1997) *Modelling Extremal Events for Insurance and Finance.* Springer-Verlag, Berlin.

[14] Feller, W. (1951) The asymptotic distribution of the range of sums of independent random variables. *Annals of Mathematical Statistics* **22**, 427–432.

[15] Fowler, H.J. and Leland, W.E. (1991) Local area network traffic characteristics, with implications for broadband network congestion management. *IEEE Journal on Selected Areas in Communications* **9**, 1139–1149.

[16] Geweke, J. and Porter-Hudak, S. (1983) The estimation and application of long memory time series models. *Journal of Time Series Analysis* **4**, 221–238.

[17] Gogl, H. (1998) Traffic statistics and application profiles in ATM networks (final report). *Preprint.* Institut für Informatik, Technische Universität München.

[18] Heath, D., Resnick, S.I. and Samorodnitsky, G. (1998) Heavy tails and long-range dependence in ON/OFF processes and associated fluid models. *Mathematics of Operations Research* **23**, 145–165.

[19] Hurst, H.E. (1951) Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers* **116**, 770–799.

[20] Internet Traffic Archive. `http://www.acm.org/sigcomm/ITA`.

[21] Klemeš, V. (1974) The Hurst phenomenon: a puzzle? *Water Resources Research* **10**, 675–688.

[22] Künsch, H. (1986) Discrimination between monotonic trends and long-range dependence. *Journal of Applied Probability* **23**, 1025–1030.

[23] Leland, W., Taqqu, M.S., Willinger, W. and Wilson, D. (1994) On the self-similar nature of Ethernet traffic (extended version). *IEEE/ACM Transactions on Networking* **2**, 1–15.

[24] Mandelbrot, B.B. (1975) Limit theorems on the self-normalized range for weakly and strongly dependent processes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **31**, 271–285.

[25] Mikosch, T. and Stărică, C. (1999) Change of structure in financial time series, long range dependence and the GARCH model. *Preprint*, University of Groningen.

[26] Park, K., Kim, G. and Crovella, M. (1996) On the relationship between file sizes, transport protocols, and self-similar network traffic. *Proceedings of the IEEE International Conference on Network Protocols*, 171–180.

[27] Paxson, V. and Floyd, S. (1995) Wide area traffic: the failure of Poisson modeling. *IEEE/ACM Transactions on Networking* **3**, 226–244.

[28] Priestley, M.B. (1981) *Spectral Analysis and Time Series, Volume 1.* Academic Press, London.

[29] Robinson, P.M. (1995) Log-periodogram regression of time series with long range dependence. *Annals of Statistics* **23**, 1048–1072.

[30] Samorodnitsky, G. and Taqqu, M.S. (1994) *Stable Non-Gaussian Random Processes.* Chapman & Hall, New York.

[31] Shorack, G.R. and Wellner, J.A. (1986) *Empirical Processes with Applications to Statistics.* Wiley, New York.

[32] Stegeman, A.W. (1998) *Modeling traffic in high-speed networks by ON/OFF models.* Masters Thesis, Department of Mathematics, University of Groningen.

[33] Tanenbaum, A.S. (1996) *Computer Networks, Third Edition.* Prentice Hall, New Jersey.

[34] Taqqu, M.S. (1975) Weak convergence to fractional Brownian motion and to the Rosenblatt process. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **31**, 287–302.

[35] Taqqu, M.S. and Teverovsky, V. (1995) Estimators for long-range dependence: an empirical study. *Fractals* **3**, 785–798.

[36] Taqqu, M.S. and Teverovsky, V. (1996) On estimating the intensity of long-range dependence in finite and infinite variance time series. In: *A Practical Guide To Heavy Tails: Statistical Techniques and Applications*, Adler, R., Feldman, R. and Taqqu, M.S., editors, Birkhauser, Boston.

[37] Teverovsky, V. and Taqqu, M.S. (1997) Testing for long-range dependence in the presence of shifting means or a slowly decaying trend, using a variance-type estimator. *Journal of Time Series Analysis* **18**, 279–304.

[38] Willinger, W., Taqqu, M.S., Sherman, R. and Wilson, D.V. (1995) Self-similarity through high-variability: statistical analysis of Ethernet LAN traffic at the source level (extended version). *Computer Communications Review* **25**, 100–113.