Bispectral analysis of traffic in high speed networks

Gy. Terdik, Z. Gál, S. Molnár*, E. Iglói

Center for Informatics and Computing,
L. Kossuth University, Debrecen H-4010, PF. 58, Hungary
E-mail: Terdik@cic.unideb.hu,

*High Speed Networks Lab.,
Department of Telecommunications & Telematics,
Technical University of Budapest, H-1117,
Pázmány Péter sétány 1/D, Budapest, Hungary,
E-Mail: molnar@bme-tel.ttt.bme.hu

Abstract

In this paper we report results regarding bispectral analysis of the longrange dependent ATM WAN traffic. Six different data sets were measured on 155 Mbps links of the SUNET by a custom-built tracing tool capable to record over 8 million consecutive cell arrivals. The complex fractal behavior of the ATM traffic claims utilization of the higher order spectra analysis. For each of the analyzed data sets was found that the Gamma distribution fits very well. The bispectrum was studied for extracting some additional information with respect to the long memory parameter. The nonlinearity ot the time series was also tested by the help of the bispectrum.

Keywords: ATM network traffic, multifractals, long range dependence, non-Gaussian time series, bispectrum, higher order cumulants, higher order spectra, linearity test

1 Introduction

In recent years we are observing a paradigm shift in the understanding of the nature of traffic in our modern packet networks including, e.g., ISDN, Ethernet, ATM and Internet. This is due to the identified fractal behavior of measured traffic [1]. The observed fractal property has opened a new venue in teletraffic research [1], [2]. However, fractal traffic modeling with the concepts of

heavy-tailedness, long-range dependence (LRD) and self-similarity is not well established yet in teletraffic theory moreover the practical application to packet network traffic is an open research topic. There are even new concepts beyond self-similarity introduced in recent years like multifractals which seems to be useful to capture some types of complex traffic behavior [2]. Different tests have been developed to detect heavy-tailedness [3], LRD and self-similarity [4]. These methods include, e.g., variance-time plot, R/S method, periodogram, Whittle-estimator, etc. The application of these statistical procedures to measured data (e.g., from Internet or ATM networks) suggests the presence of fractal properties [5], [6], [7]. However, the identification of fractal properties is not easy in practice and require a comprehensive and deep statistical analysis of the data in order to avoid misunderstandings (e.g., from non-stationarity effects [8]). Our aim is to provide an analysis of actual measured traffic from a different point of view by focusing on the higher-order statistics. The bispectral analysis is used for this purpose by which new information can be gained from a non-Gaussian process. The bispectral properties of data traffic with fractal property have not been investigated yet and our goal is that the results presented in this paper may support the understanding about the challenging fractal phenomenon. We used actual traffic taken from an ATM Wide Area Network in our analysis study. The measurements were taken by a custom-built measurement tool which was able to record 8 million consecutive cell arrivals. These long traffic traces enabled us to carry out a statistically reliable analysis.

The structure of the paper is the following. Section 2 is about the data. In section 3 the time series will be examined first in terms of extent of memory. No wonder that long range dependence is coming out considerably. Much rather surprising is the fact that the gamma distribution fits very well to all of the increments of the time series. This non-Gaussian feature makes it available to use bispectral analysis. So, not only classical spectral methods but the spectrum-bispectrum based least squares fitting method will be used to estimate the long memory parameter. Moreover, two different bispectral methods will be applied to test the linearity and significant nonlinearity will be found. Section 4 contains the conclusions.

We should like to state that this paper is addressed to the reader familiar with the most widely known notions of long range dependence. Some knowledge about higher order cumulants and higher order spectra is also required.

2 ATM WAN traffic measurements

The configuration of the measurement is given in Figure 1. As a business customer of Telia, the Swedish network operator, different parts of the Swedish University Network (SUNET) are attached to Telia's ATM wide area networks. The aggregated traffic on the SUNET was analyzed during the summer of 1996

in the framework of a common trial between the SUNET community and Telia Research. The LAN traffic of universities in the northern region, around Uppsala are connected to an FDDI backbone which is connected on R1, R2 routers and a 34 Mbps PDH link to the ATM backbone in Stockholm. This network joins the northern LANs of SUNET to the international Internet backbone and to the southern university networks around Göteborg. A CBR (Constant Bit Rate) connection with 38.16 Mbps cell rate was established on the SDH link between the routers R4 and R5 for the trial. The measurements reported here were performed on the connection between Uppsala and Göteborg. The ATM traffic streams were duplicated by means of optical splitters avoiding impacts on the original traffic flows. The duplicated traffic streams were routed on dedicated links to Telia Research in Haninge, where almost one hundred traffic traces were collected with more than 8 millions cell arrivals in each trace using a non-commercial custom built measurement instrument developed in the RACE Parasol project [9]. The measured connections used the Guaranteed Traffic Class of Telia thus the influence from other traffic in the network was negligible. The measured traffic was an aggregation of common Internet traffic including traffic generated by HTTP, FTP, telnet, chat, IPphone, etc.

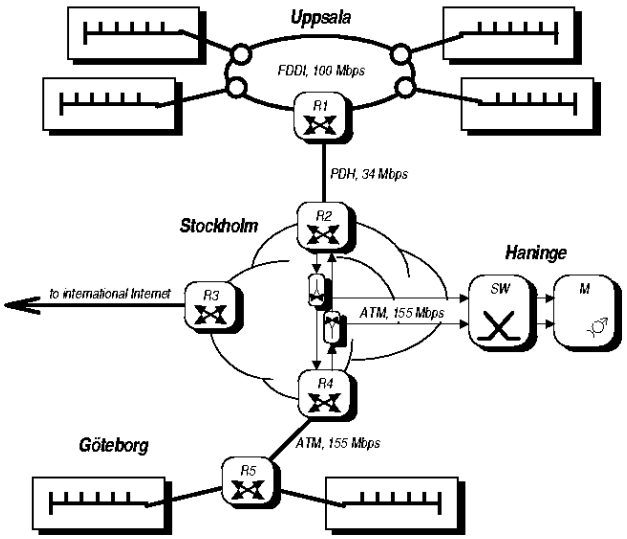

Figure 1. Topology of SUNET

We analyzed the time series of aggregated 155 Mbps SUNET ATM WAN traffic on the 20ms level. There are six data sets, Set1, Set2, ...,Set6. The data sets contain approximately from $4 \cdot 10^4$ to $6 \cdot 10^4$ observations each. The observations are the number of non-empty cells gone through the connection in consecutive 20ms time intervals. We shall denote these time series by Y_t.

3 Bispectral analysis of the ATM WAN traffic

It is well known that various types of high speed network traffic like ATM presents the property of long range dependence or long memory, see [10], [11] and [12]. This feature of the ATM traffic can be obviously observed in the following phase plot, see Figure 2. The points on the phase plot have coordinates (Y_t, Y_{t+1}). They are concentrated around the diagonal, meaning that large values of Y_t are usually followed by large ones and small values by small ones. For stationary processes this characteristic is referred to as pseudotrend and it reports the presence of long range dependence.

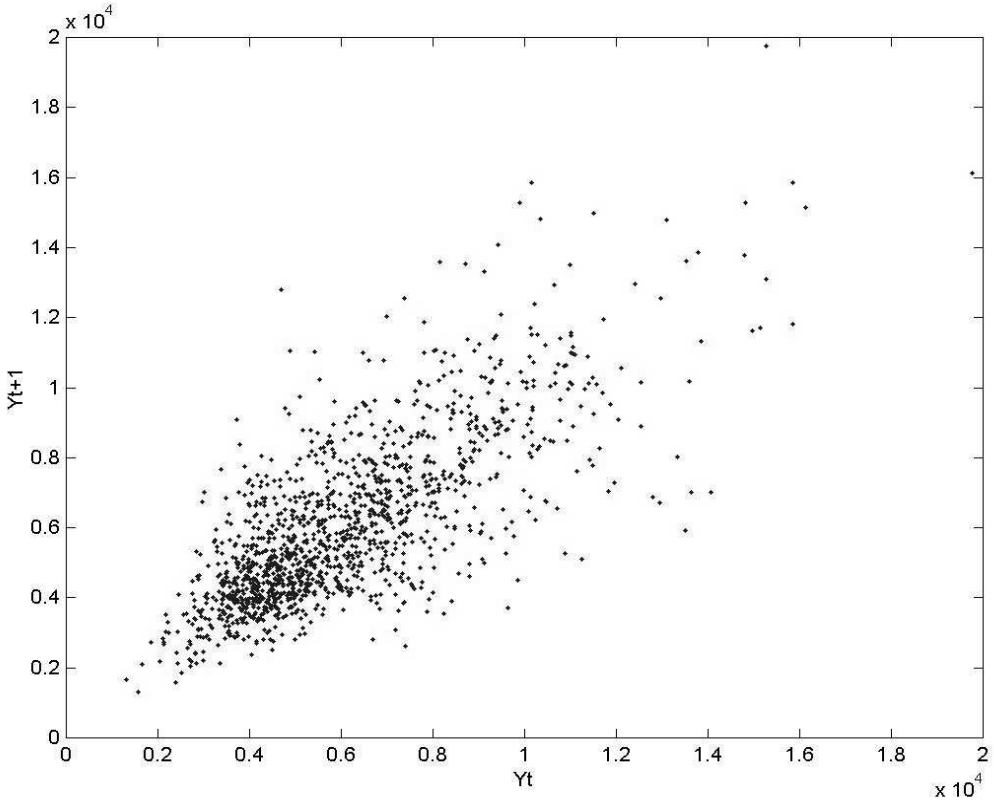

Figure 2. Phase plot of Set3

Before the definition of the long range dependence we give a brief account of the analysis of data in frequency domain.

3.1 Cumulants, spectra and estimation of bispectrum

For properties of cumulants and higher order spectra we refer to the books [13] and [14] and [15]. If a strictly stationary process Y_t has third order moments, then

not only the covariances $\mathrm{Cov}(Y_t, Y_{t+s})$ are invariant with respect to the time-shift but the third order cumulants $\mathrm{Cum}(Y_t, Y_{t+r}, Y_{t+s})$ as well. It is convenient to use the notations $z = e^{i2\pi\omega}$, $z_k = e^{i2\pi\omega_k}$ such that $z_{(1:k-1)} = (z_1, z_2, \ldots, z_{k-1})$ and $z_{(1:k-1)}^{-t_{(1:k-1)}} = \prod z_j^{-t_j} = e^{-i2\pi \sum t_j\omega_j}$. The Fourier transform of the third order cumulants is called a bispectrum. While the spectrum S_2 is real and nonnegative the bispectrum S_3 is complex valued and it has the following properties of symmetry,

$$\begin{aligned} S_{3,Y}\left(z_{(1:2)}\right) &= S_{3,Y}\left(z_2, z_1\right) = S_{3,Y}\left(z_1, z_3\right) = S_{3,Y}\left(z_3, z_1\right) \\ &= S_{3,Y}\left(z_3, z_2\right) = S_{3,Y}\left(z_2, z_3\right) = S_{3,Y}^*\left(z_{(1:2)}^{-1}\right), \end{aligned}$$

where $z_3 = (z_1 z_2)^{-1}$, and * denotes complex conjugate. These equations imply that there are twelve triangles of frequencies in the plane, each of which can be considered as the basic domain for the bispectrum because it is completely specified over the entire plane if it is determined over one of the twelve triangles. We shall fix the triangle with vertices $(0,0), (1/2,0), (1/3,1/3)$ as the basic domain for the bispectrum. The bicoherency is defined by

$$B\left(z_1, z_2\right) = \frac{\left|S_{3,Y}\left(z_1, z_2\right)\right|^2}{S_{2,Y}\left(z_1\right) S_{2,Y}\left(z_2\right) S_{2,Y}\left(z_3\right)}.$$

Let us denote the estimate of the bispectrum of Y_t by $S_{3Y}^T(z_1, z_2)$, where the length of our time series is T. The smoothed biperiodogram will be used for the estimation of the bispectrum of Y_t. Recall that the biperiodogram of Y_t is

$$I_{YYY}^T(z_1, z_2) = \frac{1}{T} d_T(z_1) d_T(z_2) d_T(z_3),$$

where

$$d_T(z) = \sum_{t=0}^{T-1} Y_t z^{-t}$$

is the discrete Fourier transform of the series Y_t.

Let $W(\omega_1, \omega_2)$ be a nonnegative bounded continuous weight function such that

$$\iint_{-\infty}^{+\infty} W(\omega_1, \omega_2) d\omega_1 d\omega_2 = 1.$$

We choose $W(\omega_1, \omega_2)$ to be 0 for $\mid \omega_j \mid > 1/2, \quad j = 1, 2$ and also $W(\omega_1, \omega_2)$ to satisfy the symmetry conditions

$$\begin{aligned} W(\omega_1, \omega_2) &= W(\omega_2, \omega_1) \\ &= W(\omega_1, \omega_3), \end{aligned}$$

where $\omega_3 = -\omega_1 - \omega_2 \bmod (1)$ and therefore $\omega_3 \in [-1/2, 1/2]$. Now the smoothed biperiodogram is defined by

$$S_{3,Y}^T(\lambda_1, \lambda_2) = \frac{1}{T^2 b_T^2} \sum_{s_1, s_2 = 1}^{T-1} W_1(s_1, s_2) I_{eee}^T\left(\frac{s_1}{T}, \frac{s_2}{T}\right),$$

where b_T denotes a sequence of scale parameters such that $b_T > 0$, $b_T \to 0$, $b_T^2 T \to \infty$, as $T \to \infty$ and $W_1(s_1, s_2) = W(b_T^{-1}(\lambda_1 - \frac{s_1}{T}), b_T^{-1}(\lambda_2 - \frac{s_2}{T}))$. The estimate of $S_{3,Y}^T$ actually involves a weighting of $(2Tb_T+1)^2$ biperiodogram ordinates in the neighborhood of (λ_1, λ_2).

3.2 Long range dependence

The stationary process Y_t is called long range dependent with long memory parameter H if its spectrum $S_{2,Y}(\omega)$ behaves around zero like $c\,|\omega|^{1-2H}$, with some constant $c > 0$. There are several methods for the estimation of the long memory parameter H which is called also as Hurst parameter. These are the Whittle method, the Geweke–Porter-Hudak method, the R/S method, see [4], the variance-time method and its modification, the IDC method [8]. We used the Whittle and the Geweke–Porter-Hudak (G-P-H) methods, and we obtained the following estimated values $\widehat{H}$ for the long memory parameter H.

Setname	$\widehat{H}$	
	Whittle	G-P-H
Set1	.82	.88
Set2	.89	.86
Set3	.89	.91
Set4	.80	.79
Set5	.88	.94
Set6	.82	.81

We did not want to fit a completely specified model and were interested only in the parameter H, so we used the first fifth part of the frequency domain $(0, 1/2)$, where the behavior of the spectrum is determined mostly by the long memory effect, i.e., $S_2(\omega) \simeq c\omega^{1-2H}$, for ω in the neighborhood of zero.

As it is well known (see, e.g.[4]), the applied two spectral methods utilizes only the information given by the second order moment structure of the data. A stationary Gaussian process is characterized totally by the mean and the second order moments, or equivalently, by the spectrum. The spectrum of a non-Gaussian process does not contain all the information about the parameters in general, therefore it is necessary to use the third and higher order properties, i.e., the bispectrum and higher order spectra for the identification. It turns out, that the information given by the spectrum and the bispectrum describes some classes of nonlinear models for example bilinear ones, see Terdik [16]. Therefore the non-Gaussian method of estimation based on the second and third order spectra, see Brillinger [17], can be used for the estimation of the parameters of a non-Gaussian process. We found that the estimated bispectra of all data sets Set1, Set2, ...,Set6 are different from zero, see Figure 3.

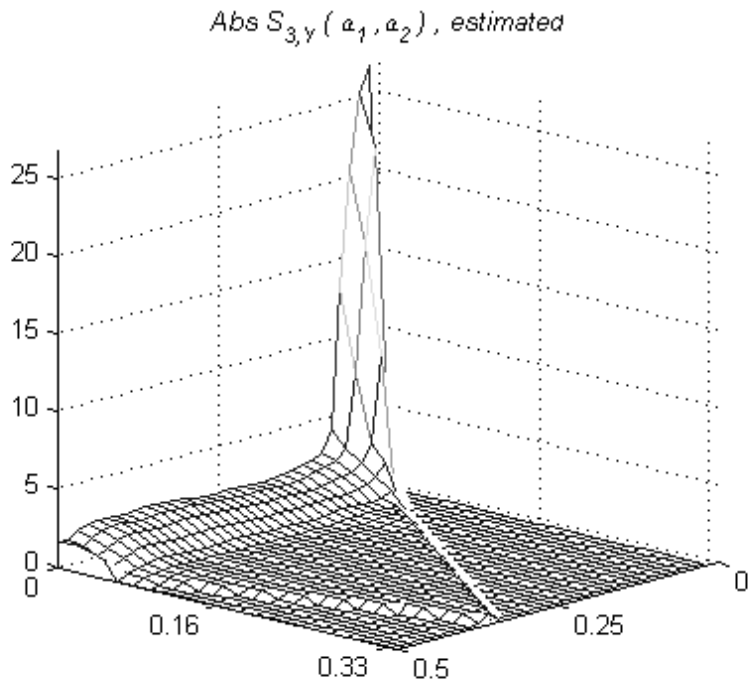

Figure 3. Estimated bispectrum of Set3

Thus, the bispectrum is available for extracting some additional information with respect to the long memory parameter. This can be explained by the fact that the estimated spectrum and the bispectrum are always asymptotically independent, see [13]. There are two situations when the bispectrum can not be applied for parameter estimation. The one is when it is zero and the other is when the sixth order moment of the distribution of the time series is infinite. The first case, i.e., the zero bispectrum can be ruled out since the estimated skewness and the histograms of the data sets show that the distributions of data sets are far away from Gaussian. By our experience based on all the data sets there is a recognizable convergence to the fractional Brownian motion as the aggregation level or time scale increases but the speed of the convergence is very slow. The convergence of the marginal distribution to the Gaussian one is slower than that it is in the case of independent observations. This is a consequence of the strong dependence structure since the Berry–Esséen bound does not necessarily applies to long memory processes. So, even the 20ms time scale aggregated data proved to be non-Gaussian and considerably positively skewed. The positive skewness comes out in the phase plot in Figure 2 and on the following histogram, too, see Figure 4. We did not have to be afraid of the second case either, since the distributions have no heavy tails. Actually we found that the gamma distribution fits very well at all aggregation levels for all data sets. The estimated and the

theoretical density function and characteristic function show the exact fitting of the marginal distributions to the gamma, see Figure 4 and Figure 5.

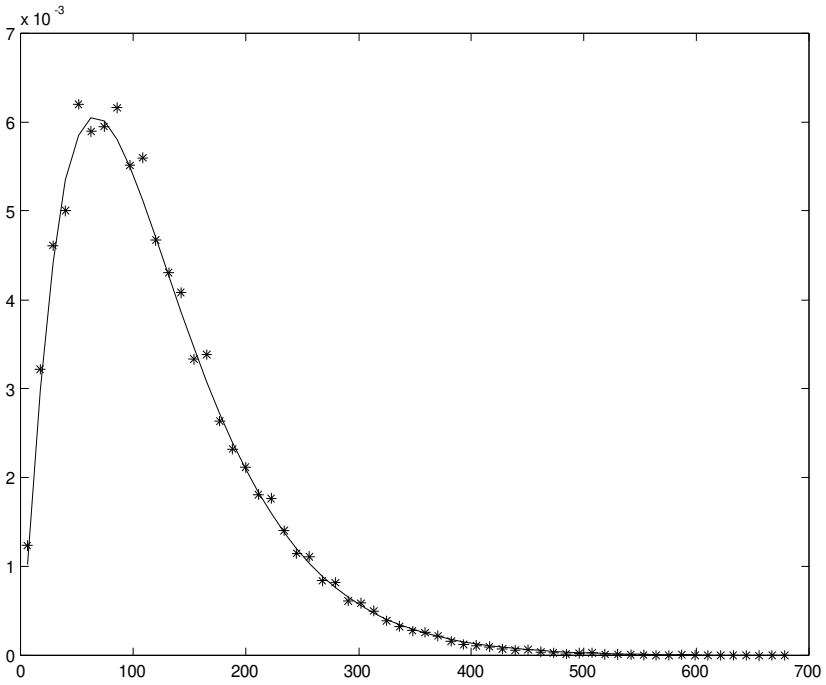

Figure 4. Histogram of data Set3 and the gamma density function with the estimated parameters

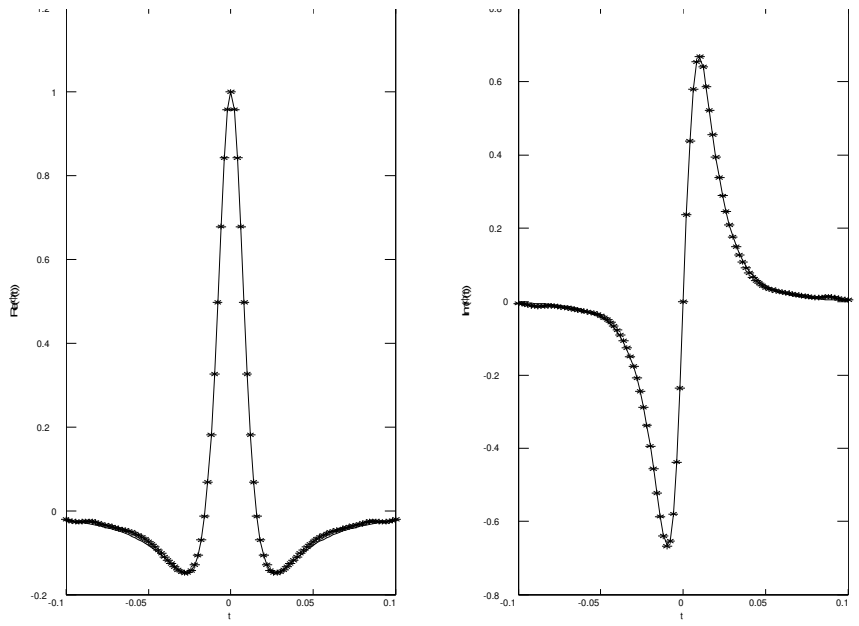

Figure 5. Real and imaginary parts of the estimated and the theoretical characteristic functions for Set3

The following estimation method arising from Brillinger [17] and worked out in [18], [15], is the improved version of the method of squared relative errors. The object function to minimize with respect to the model parameter θ is

$$Q_T(\theta) = \frac{1}{|\Lambda|} \sum_{\omega \in \Lambda} \left(\frac{S_2(\omega;\theta) - S_2^T(\omega)}{S_2(\omega;\theta)} \right)^2 + \frac{1}{|\Delta|} \sum_{(\omega_1,\omega_2)\in\Delta} \frac{\left| S_3(\omega_1,\omega_2;\theta) - S_3^T(\omega_1,\omega_2) \right|^2}{S_2(\omega_1;\theta) S_2(\omega_2;\theta) S_2(\omega_1+\omega_2;\theta)}, \tag{1}$$

where the length T data set is segmented to smaller pieces of length v, Λ denotes the set of the Fourier frequencies in the interval $(0, 1/2)$, Δ is the set of the Fourier frequency pairs in the triangle with vertices $(0,0), (1/2,0), (1/3,1/3)$. We applied this method without the knowledge of the exact form of the model because we wanted to estimate H alone. Thus, small parts in the neighborhoods of zero of the frequency sets was used as Λ and Δ above, because near zero the behavior of the spectrum and the bispectrum are determined mostly by the long memory effect. The values of parameter $\theta = H$ estimated by this method were more stable than the ones estimated by minimizing only the first term of (1). The stability is meant with respect to variations of both the technical parameters of

the procedure, and the aggregation level of the data.

	Set1	Set2	Set3	Set4	Set5	Set6
$\widehat{H}$	0.88	0.82	0.89	0.81	0.84	0.89

3.3 Linearity test and bispectrum

Let us consider a linear filter of the form

$$Y_t = \sum_{m=-\infty}^{\infty} a_m e_{t-m},$$

such that along with the process e_t the process Y_t be 3^{rd} order stationary also. Define the transfer function of this linear operator by

$$\alpha(z) = \sum_{m=-\infty}^{\infty} a_m z^{-m}.$$

It is known that for a linear process the spectral density of n^{th} $(n = 2, 3)$ order is given by

$$S_{n,Y}\left(z_{(1:n-1)}\right) = \prod_{1}^{n} a\left(z_k^{-1}\right) S_{n,e}\left(z_{(1:n-1)}\right), \tag{2}$$

where $\prod_{j=1}^{n} z_j = 1$. In particular, when e_t is i.i.d. (independent and identically distributed) then all spectra of e_t are constants and the Y_t is called linear. The smoothed biperiodogram is used for test of Rao-Hinich, [19]-[20], and for test by Terdik-Máth, see [21]. In view of the spectra of a linear process, see Eq. (2), it is clear that the bicoherency of a linear process is constant. In case the series is not Gaussian Subba Rao and Gabr [19] and Hinich [20] built a linearity test to check the hypothesis that the innovation sequence $\{e_t\}$ is independent, examining whether the bicoherency of Y_t is constant. While the previous method tests strong linearity, the following one tests weak linearity. We shall call a model to be weakly linear if the linear predictor is optimal. This means that the innovation sequence e_t fulfils the following conditions:

$$\forall t \in \mathbb{Z}: \quad \mathrm{E}\left(e_t \mid \mathcal{F}_t\right) = 0, \quad \mathrm{E}\left(e_t^2 \mid \mathcal{F}_t\right) = \mathrm{E}\, e_t^2 = \sigma^2, \tag{3}$$

where $\mathcal{F}_t$ is the Borel field generated by $e_s,\ s < t$, which is a stronger assumption than the whiteness of the noise e_t, but weaker than the independence of the series e_t. In view of our definition of weak linearity, testing the linearity is equivalent to checking the optimality of the linear predictor. The alternative is that the optimal predictor is of a quadratic type. The test is based on an additive property of the bispectrum of the innovation process, see Terdik and Máth [21]. We performed these tests and found the observed errors of the first kind of Rao–Hinich test

are all zero and the test of Terdik–Máth found only .01 for the Set2 and zeros otherwise. The conclusion is that all the data sets are significantly nonlinear, i.e., an adequate model for the ATM traffic could be chaotic.

It is interesting to note that the imaginary part of the estimated bispectrum is zero for each of the six ATM data sets. This is equivalent to the property $\mathrm{Cum}(Y_t, Y_{t+r}, Y_{t+s}) = \mathrm{Cum}(Y_t, Y_{t-r}, Y_{t-s})$, $r, s \in \mathbb{Z}$, which means something like third order moment reversibility, e.g., white noise sequences do not necessarily have this property but independent stationary sequences do have, and the Gaussianness of a stationary sequence also involves third order moment reversibility. A possible physical explanation of the reversibility is due to the symmetrical congestion control mechanisms of the TCP layer and the fixed ATM cell size.

4 Conclusions

It has been found that the time series of data sets under examination are long range dependent, nonlinear and gamma distributed. The long memory parameters were estimated by spectrum-bispectrum based least squares fitting and were found to be more stable than the ones arising from simply spectral methods. Two different bispectral methods were applied to test the linearity and all of the six data sets were found to be significantly nonlinear. The zero imaginary parts of the bispectra show a noticeable feature, namely the reversibility of the ATM traffic. An interesting problem is to find the exact form of a model with the above mentioned properties. So, to construct a stationary increment process, all the increment processes of which are gamma distributed, long range dependent, nonlinear and third order moment reversible, is one of our future research topics.

ACKNOWLEDGMENTS

1. The reported ATM WAN measurements were carried out by the CARAT laboratory of Telia Research in Sweden. We would like to thank Nils Björkman, Urban Hansson, Alexander Latour-Henner and Aziz Miah for the measurements.
2. This research is partially supported by the Hungarian National Science Foundation OTKA No. T 032658.
3. The authors are grateful to the referee for valuable remarks and suggestions.

References

[1] W. Willinger, M. S. Taqqu, and A. Erramilli. A bibliographical guide to self-similar traffic and performance modeling for modern high speed networks. In F.P. Kelly, S. Zachary, and I. Ziedins, editors, *Stochastics Networks: Theory and Applications*, pages 339–366. Oxford University Press, 1996.

[2] S. Molnár and I. Maricza. (Eds.). Source characterization in broadband networks. Technical report, COST 257 Interim Report, January 1999.

[3] S. I. Resnick. Heavy tail modeling and teletraffic data. *The Annals of Statistics*, 25(5):1805–1869, 1997.

[4] J. Beran. *Statistics for Long-Memory Processes*, volume 61 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, 1994.

[5] 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. *Computer Communication Review*, 25:100–113, 1995.

[6] S. Molnár and A. Vidács. On modelling and shaping self-similar ATM traffic. In *Proceedings, International Teletraffic Congress, ITC15, Washington, D.C., USA*, 1997.

[7] S. Molnár, T. D. Dang, and A. Vidács. Heavy-tailedness, long-range dependence and self-similarity in data traffic. In *Proceedings, 7th IEEE International Conference on Telecommunication Systems Modeling and Analysis, Nashville, Tennessee, USA*, March 18-21 1999.

[8] S. Molnár, A. Vidács, and A. A. Nilsson. Bottlenecks on the way towards fractal characterization of network traffic: Estimation and interpretation of the Hurst parameter. In *International Conference of the Performance and Management of Complex Communication Networks*, Tsukuba, Japan, 17-21 November 1997. PMCCN'97.

[9] O. Melteig. Introduction to the PARASOL project. In *The 9th Nordic teletraffic Seminar, Norwegian Telecom, Research Department, August, 1997.*

[10] J. Beran, R. Sherman, M. S. Taqqu, and W. Willinger. Long-range dependence in variable-bit-rate video traffic. *IEEE Transactions on Communications*, 43(2/3/4):1566–1579, 1995.

[11] M. E. Crovella and A. Bestavros. Explaining world wide web traffic self-similarity. 1995.

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

[13] D. R. Brillinger. *Time Series: Data Analysis and Theory.* Mc Graw Hill, New York, expanded edition, 1981.

[14] M. Rosenblatt. *Stationary sequences and random fields.* Birkhauser, Boston, 1985.

[15] Gy. Terdik. *Bilinear Stochastic Models and Related Problems of Nonlinear Time Series Analysis; A Frequency Domain Approach*, volume 142 of *Lecture Notes in Statistics*. Springer Verlag, New York, 1999.

[16] Gy. Terdik. On problem of identification for stochastic bilinear systems. *SAMS*, 17:85–102, 1995.

[17] D. R. Brillinger. Fourier inference: Some methods for the analysis of array and nongaussian series data. *Water Resources Bull.*, 21:743–756, 1985.

[18] N. Leonenko, A. Sikorskii, and Gy. Terdik. On spectral and bispectral estimator of the parameter of nonGaussian data. *Random Oper. and Stoch. Equ. (ROSE)*, 6(2):159–182, 1998. Correction: ROSE(1999), vol. 7, p. 107.

[19] T. Subba Rao and M. M. Gabr. A test for linearity of stationary time series. *Journal of Time Series Analysis*, pages 145–158, 1980. JTSA.

[20] M. J. Hinich. Testing for Gaussianity and linearity of a stationary time series. *Journal of Time Series Analysis*, 3:169–176, 1982. JTSA.

[21] Gy. Terdik and J. Máth. A new test of linearity for time series based on its bispectrum. *Journal of Time Series*, 19(6):737–749, 1998.