# Developed Algorithm for Supervising Identification of Non Linear Systems using Higher Order Statistics: Modeling Internet Traffic

J. Antari[1], A. El Khadimi[2], D. Mammas[3] and A. Zeroual[4]

[1]*Ibn Zohr University, Faculty Polydisciplinary of Taroudant, Department of Computer Engineering, Morocco*

[2]*Department of Systems and Communications, National Institute of Posts and Telecommunications (INPT), Rabat, Morocco*

[3]*Ibn Zohr University, EST Agadir, Morocco*

[4]*Cadi Ayyad University, Faculty of Sciences Semlalia, Team signal processing, BP 2390 Marrakesh, Morocco.*

*antari_j@yahoo.fr*

***Abstract***

*In this work, we use the formulas of statistic techniques for developing an algorithm based on third order moments and autocorrelation function. This algorithm permits to identify non linear system coefficients for recovering the real information from input-output systems. Simulation examples and comparison with other method in the literature are provided to verify the performance of the developed algorithm. The obtained results demonstrate the efficiency and the accuracy of the developed algorithm for non linear system identification under various values of signal to noise ratio (SNR) and different sample sizes N. To corroborate the theoretical results for a real process, we applied the developed algorithm to search a model able to represent the internet traffic data.*

***Keywords:*** *Autocorrelation function; higher order statistics; non linear systems; Simulation and modelling; Traffic internet data*

## 1. Introduction

In most case of system identification and modelling, the problems of the uncertainties persist always between the real systems and the evolution of obtained models. These uncertainties are due to the lacks of exhaustive knowledge on the performance of systems and the obtained models take in account that a part of the parameters influences the output evolution of models. Several models are identified in literature such as the finite impulse response systems (linear models) which are identified using different algorithms based on higher order cumulants of system output. These algorithms can be in general classified into three classes of solutions: closed form solutions, optimization-based solutions and linear algebra solutions [1-4]. The linear algebra solutions have received much attention because they have 'simpler' computation and are free of the problems of local extremes that often occur in the optimization solutions. Although the closed-form solutions have similar features, they usually do not smooth out the noise caused from observation and computation [5]. But the linear models are not efficient for representing and modelling all systems, because the majority of systems are represented by non linear models.

Nonlinear models appear to be powerful tools for modelling and representing physical systems with high precisions. Among these models, the polynomial models are more exploited to describe the input-output of systems and that can be represented by Volterra series [6], which has been used to describe a large class of nonlinear systems, and have been applied extensively in various engineering practice [7]. Furthermore, many important non-linear effects in engineering and science can be approximated by a Volterra series of second or third order, i.e., quadratic nonlinear systems or cubic nonlinear systems [8]. These classes are widely used in non linear filtering, communication, active noise control, biomedical engineering, signal processing and are identified using moment versions [9, 10, 11, 12]. These last techniques are more useful in many real applications where we are faced to truly non Gaussian signals. Indeed, the moments are applicable to non Gaussian signals and constitute a strong identification tool for most applications characterized by non Gaussian or non linear process [5, 13].

The identification algorithms in the literature permit to identify the system from their output (blind identification). But, it exists also the systems which can be identified from the input and output (supervised identification) for finding their parameters. From this reason we use, in this work, the formulas of statistic techniques for developing a supervised algorithm based on third order moments and autocorrelation function. This algorithm permits to identify non linear systems (quadratic non linear systems) coefficients and exploits q +1 equations for identifying q coefficients. In the last part of this work, we apply the developed algorithm to search for a model able to represent real data [14].

## 2. Higher Order Statistics and Non Linear Model

### 2.1 Higher Order Statistics

For zero-mean stationary process y(n), the moments up to order four are given by [15]:

$C_{1y} = E\{y(n)\}$,

$C_{2y}(m) \equiv R(m) = E\{y(n),y(n+m)\}$: Autocorrelation function

and $C_{3y}(m,k) = E\{y(n),y(n+m),y(n+k)\}$

Where E{.} : mathematical expectation.

### 2.2 Non Linear Model

The non linear model considered in this section is diagonal quadratic systems, which represent the particular case of the Volterra series, the choice of this type due to the simplicity for determine the parameters of the process. So, the diagonal quadratic system is defined as:

$$y(n) = \sum_{i=0}^{q} h_d(i)x^2(n-i) \quad \text{and} \quad z(n) = y(n) + \eta(n) \tag{1}$$

with:

- $x(n)$ is the input sequence

- $h_d$ and q are respectively the parameters and the order of the diagonal quadratic systems

- $y(n)$ represents the system output in noiseless case and $z(n)$ is the observed system output corrupted by additive Gaussian noise $\eta(n)$.

The following conditions are assumed to be satisfied:

A1: The model order q is supposed to be known,

A2: The input sequence $x(n)$ is independent and identically distributed (i.i.d) zero mean, the variance is $\sigma_x^2 \cong 1$, and non Gaussian,

A3: The system is causal and $h_d(0)=1$,

A4: The measurement noise sequence $\eta(n)$ is assumed to be zero mean, i.i.d, Gaussian, independent of $x(n)$ with unknown variance and the mth order cumulants are zero when $m > 2$.

## 3. Developed Algorithm and Simulation

### 3.1. Developed Algorithm and Proof

The starting point in this subsection is to describe all equations which linked the second order moments, the third order moments and the diagonal parameters of quadratic system using Leonov-Shiryayev formula [16]:

The second order moments of the $y(n)$ is described by the following expression:

$$C_2y(\tau)= \mathrm{Cum}\{y(n), y(n+ \tau)\} \tag{2}$$

$$C_{2y}(\tau) = (\gamma_{4x} - \gamma_{2x}^2)\sum_{i=0}^{q} h_d(i)h_d(i+\tau) \tag{3}$$

we suppose that: $\Gamma_x = \gamma_{4x} - \gamma_{2x}^2$

where: $\gamma ix$ ith order moments at origin, Cum(y) represents the moments of processes $y(n)$ and $\tau$ represents the time lag of random sequence.

In the same way we defined respectively the third order moments as follows:

The third order moment of the signal $y(n)$ is given by the following equation:

$$C3y(\tau 1, \tau 2)= \mathrm{Cum}\{y(n), y(n+ \tau 1), y(n + \tau 2)\} \tag{4}$$

$$C_{3y}(\tau_1, \tau_2) = (\gamma_{6x} - 3\gamma_{4x}\gamma_{2x} + 2\gamma_{2x}^3)\sum_{i=0}^{q} h_d(i)h_d(i+\tau_1)h_d(i+\tau_2) \tag{5}$$

we suppose that: $K_x = \gamma_{6x} - 3\gamma_{4x}\gamma_{2x} + 2\gamma_{2x}^3$

The Fourier transform of the Eqs. 3 and 4 are given by the following equations:

$$S_{2y}(w) = TF\{C2y(\tau)\} = \Gamma_x Hd(w)Hd(-w) \quad (6)$$

$$S_{3y}(w_1,w_2)=TF\{C_{3y}(\tau_1, \tau_2)\}=K_x.\ H_d(w_1)H_d(w_2)H_d(-w_1-w_2) \quad (7)$$

with $H(w) = \sum_{i=0}^{+\infty} h_d(i)\exp(-jw\, i)$

We use the Fourier transform of the Eqs. (6) and (7), we demonstrate the relationships that linked the spectra, bispectra and diagonal parameters hd(i).

We suppose: $w = w_1 + w_2$, the Eq. (6) becomes:

$$S_{2y}(w_1 + w_2) = \Gamma x.\ H_d(w_1 + w_2)H_d(-w_1 - w_2) \quad (8)$$

From the Eqs. (7) and (8) we obtain the following equation:

$$S_{3y}(w_1,w_2)\ H_d(w_1 + w_2) = \mu\ H_d(w_1)H_d(w_2)\ S_{2y}(w_1 + w_2) \quad (9)$$

with $\mu = \left( K_x / \Gamma_x \right)$

The inverse Fourier transform of the Eq. (9) demonstrates that the 3rd order moments, the second order moments and the diagonal hd(i) parameters of quadratic systems are combined by the following equation:

$$\sum_{i=0}^{q} C_{3y}(\tau_1 - i, \tau_2 - i)h_d(i) = \mu.\sum_{i=0}^{q} C_{2y}(\tau_1 - i)h_d(i)h_d(\tau_2 - \tau_1 + i) \quad (10)$$

We use the AutoCorrelation Function (ACF) property of a stationary process, such as $C_{2y}(\tau) \neq 0$ only for $-q \leq \tau \leq q$ and vanishes elsewhere. In addition, we take $\tau_1 = -q$, the Eq. (10) takes the form:

$$\sum_{i=0}^{q} C_{3y}(-q - i, \tau_2 - i)h_d(i) = \mu.h_d(0)h_d(\tau_2 + q)C_{2y}(-q) \quad (11)$$

$$\sum_{i=1}^{q} C_{3y}(q + i, \tau_2 + q)h_d(i) + C_{3y}(q, q + \tau_2) = .h_d(\tau_2 + q)C_{2y}(q) \quad (12)$$

The considered system is causal. So, the interval of the $\tau_2$ is $\tau_2 = -q , \ldots, 0$

From equation (12) we obtain the following matrix:

$$\begin{pmatrix} C_{3y}(q+1,0) & C_{3y}(q+2,0) & \cdots & C_{3y}(2q,0) \\ C_{3y}(q+1,1)-F_x & C_{3y}(q+2,1) & \cdots & C_{3y}(2q,1) \\ . & C_{3y}(q+2,2)-F_x & \cdots & . \\ . & & \ddots & . \\ . & & \ddots & . \\ C_{3y}(q+1,q) & C_{3y}(q+2,q) & \cdots & C_{3y}(2q,q)-F_x \end{pmatrix} \times \begin{pmatrix} h_d(1) \\ . \\ . \\ h_d(i) \\ . \\ h_d(q) \end{pmatrix} = \begin{pmatrix} F_x - C_{3y}(q,0) \\ -C_{3y}(q,1) \\ . \\ . \\ . \\ -C_{3y}(q,q) \end{pmatrix} \quad (13)$$

with: $F_x = \mu . C_{2y}(q)$

From the matrix (13) we obtain the following equation:

$$M.\Theta = K \quad (14)$$

where M, Θ and K are the sizes respectively: (q+ 1).(q), (q).(1) and (q+1).(1). The least square (LS) solution of the system of equation (14) is:

$$\hat{\Theta}(i) = (M^T.M)^{-1}.M^T.K \quad i = 1\ldots\ldots q \quad (15)$$

where: $M^T$ represents the transpose of the matrix M.

The equation (15) permits to identify the diagonal parameters of quadratic systems (non blind identification) and allow also the modelling of the data process.

### 3.2 Simulation Results

We test the performance of the developed algorithm before to be applied to real data. For this motivation we use for example the simulation of the quadratic system given by the following equation:

$$z(n) = x^2(n) - 2.15\, x^2(n-1) + 1.20\, x^2(n-2) + \eta(n)$$

The simulation results are illustrated in table 1 and 2 using signal to noise ratio (SNR= 0dB and 15 dB) with different sample sizes (N= 300, 600, 900) and for 50 Monte-Carlo runs. The choice of these sizes is imposed in practice, in most cases, which we have a small data (to be modelled).

The SNR is defined by: SNR= 10.Log10( $\sigma_y^2 / \sigma_\eta^2$ ), $\sigma_y^2$ and $\sigma_\eta^2$ represent respectively the variance of the output system and noise signal.

**Table 1. Estimated Parameters for SNR=0dB**

**True values: $h_d(1)$= -2.15 and $h_d(2)$= 1.20.**

| N | Algorithms | $\hat{h}_d(1) \pm$ std | $\hat{h}_d(2) \pm$ std | NMSE |
|---|---|---|---|---|
| 300 | Developed algorithm | -2.1496 ± 0.0441 | 1.2019 ± 0.0375 | $1.9252\ 10^{-4}$ |
| | RLS method | -2.1396 ± 0.0489 | 1.1952 ± 0.0298 | $2.3030\ 10^{-4}$ |
| 600 | Developed algorithm | -2.1581 ± 0.0354 | 1.2013 ± 0.0205 | $2.8051\ 10^{-5}$ |
| | RLS method | -2.1488 ± 0.0311 | 1.1985 ± 0.0215 | $3.4846\ 10^{-5}$ |
| 900 | Developed algorithm | -2.1507 ± 0.0263 | 1.2025 ± 0.0184 | $1.0866\ 10^{-5}$ |
| | RLS method | -2.1438 ± 0.0240 | 1.2011 ± 0.0182 | $1.8290\ 10^{-5}$ |

**Table 2. Estimated Parameters for SNR=15dB**

**True values: $h_d(1)$= -2.15 and $h_d(2)$= 1.20.**

| N | Algorithms | $\hat{h}_d(1) \pm$ std | $\hat{h}_d(2) \pm$ std | NMSE |
|---|---|---|---|---|
| 300 | Developed algorithm | -2.1483 ± 0.0081 | 1.2011 ± 0.0052 | $1.5800\ 10^{-6}$ |
| | RLS method | -2.1521 ± 0.0083 | 1.1998 ± 0.0061 | $1.8664\ 10^{-6}$ |
| 600 | Developed algorithm | -2.1489 ± 0.0050 | 1.1991 ± 0.0032 | $1.0228\ 10^{-6}$ |
| | RLS method | -2.1497 ± 0.0062 | 1.2007 ± 0.0040 | $1.3978\ 10^{-6}$ |
| 900 | Developed algorithm | -2.1503 ± 0.0015 | 1.2000 ± 0.0009 | $1.2299\ 10^{-7}$ |
| | RLS method | -2.1502 ± 0.0041 | 1.2003 ± 0.0030 | $1.4235\ 10^{-7}$ |

From Tables 1 and 2 we can note that:

The NMSE values using the developed algorithm are lower than those obtained by the RLS method. Indeed, the estimates of diagonal parameters ($\hat{h}_d(1)$ and $\hat{h}_d(2)$) are approximately closer to real values for developed algorithm and the figures 1 and 2 confirm the superiority of this algorithm. The obtained values of the standard deviation (std) demonstrate the small fluctuation around the mean parameters. So, the developed algorithm has a property which is very interest in signal processing, mainly in signal detection in noisy environment.

In the following section we have used the developed algorithm to modelling the data of video packets transmission [14].

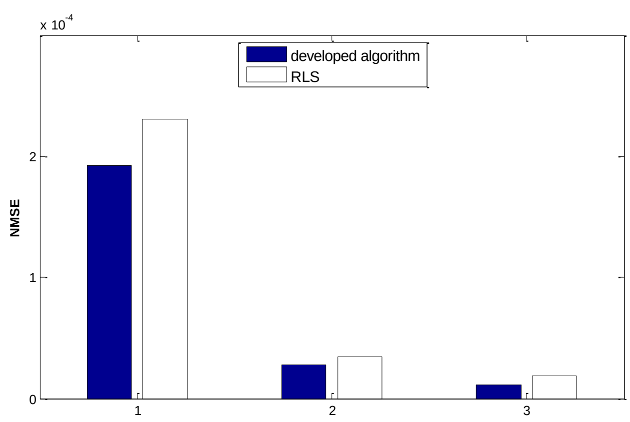


**Figure 1. Values of NMSE for SNR= 0dB**

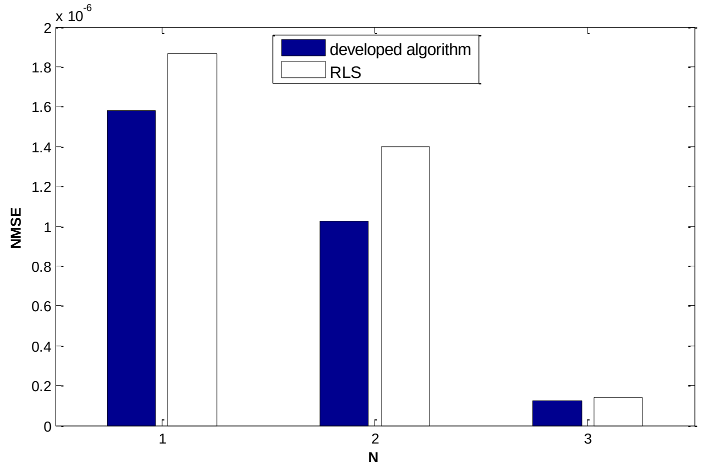


**Figure 2. Values of NMSE for SNR= 15dB**

## 4. Data Analysis

In this part, we analyse the data video packets transmission for the network management. For this reason we select about the average 600 packets in two sides namely A and B [14]. We note that there are two parameters T1 and T2, which represent respectively the input and output of the data in each side.

The Figure 3 demonstrates that the data are non Gaussian process. The last is essential in the modelling when we use the statistics methods. But, it’s necessary to verify the stationary of the process in the following part.

The Figure 4 plots the evolutions of the data in sides A and B which are characterized by a non stationary behaviour. In order to make this phenomenon in evidence, we plot the AutoCorrelation function (ACF) of the process (Figure 5). The low decrease of the ACF and the periodic phenomenon confirm that the process is non stationary. And as, our contributions are interest to develop the models based on the HOS techniques. So, it is necessary to transform these process to the stationary process using a low pass filter for eliminate the low frequency, which are responsible to the non stationary phenomenon [17].

The Partial AutoCorrelation function PACF (Figure 7) of the transformed data (Figure 6) demonstrates the stationary phenomenon of the process. Indeed, the PACF decreases rapidly and is inside the confidence interval of 95%. So, these results may prove the applicability of the HOS techniques for the non-blind identification of diagonal quadratic systems which permit to modelling the video-packets data for the network management.

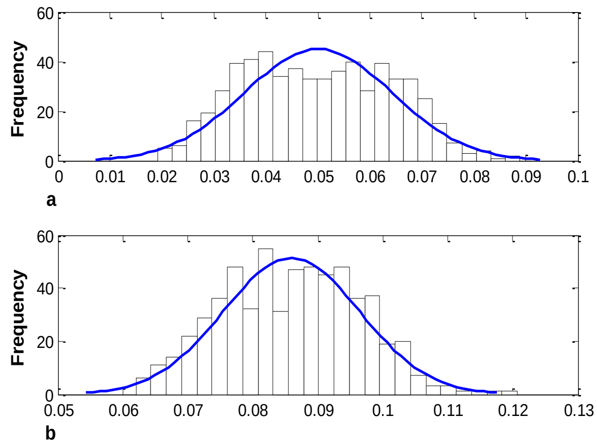


**Figure 3. Frequency Distribution of the Video-packets (a for side A, b for side B)**

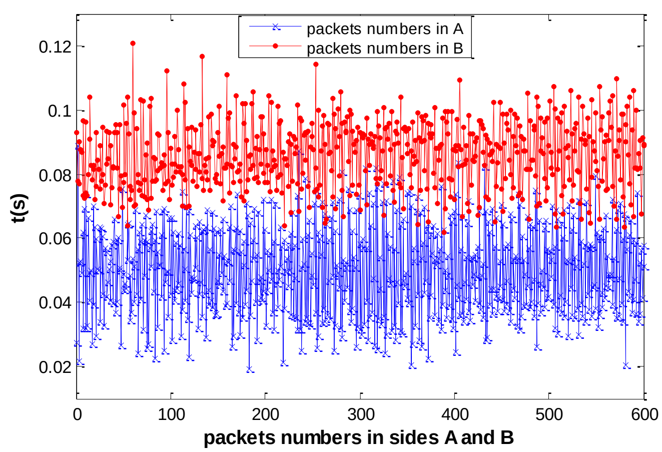


**Figure 4. Evolution of the Data in Sides A and B**

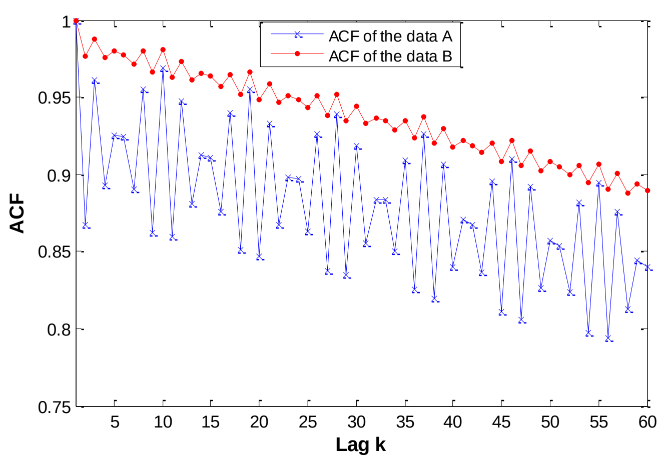


**Figure 5. ACF of the Data A and B**

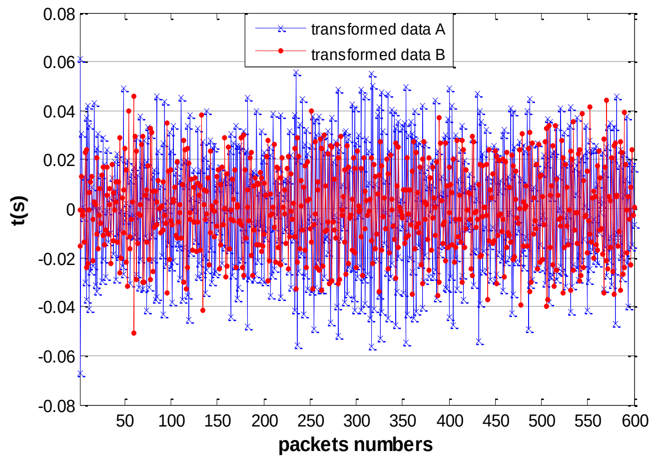


**Figure 6. Transformed Data in Sides A and B**

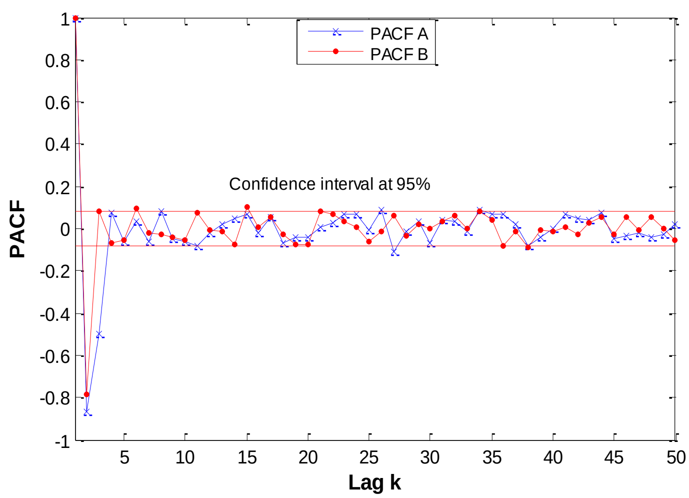


**Figure 7. PACF of the Transformed Data in Sides A and B**

## 5. Identification Model

In this paragraph, using previous algorithm, we develop the model describe the data of video-packets transmission for the network management.

The Figure 8 shows the real and the estimated data of the video-packets transmission using developed algorithm with the input data is T1 in sides A and B. The analysis of this figure demonstrates that, the obtained model using developed algorithm is very similar to the real data and with root mean square error of 0.0034.

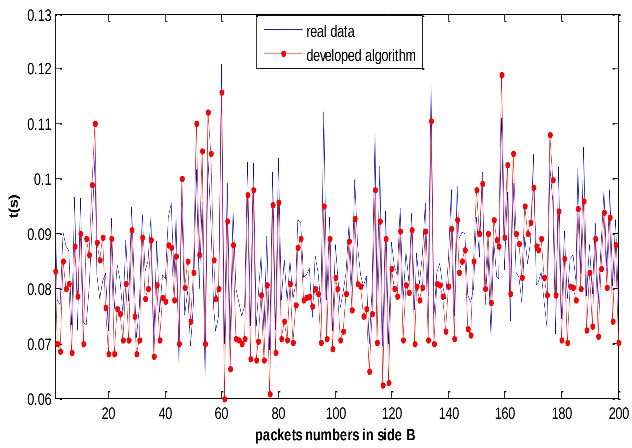


**Figure 8. Comparison of Different Models**

## 6. Diagnostic Checking

Typically the goodness of fit of statistical models to a set of data is judged by the ACF residual data. So, we plot in figure 9 the ACF residual data [19] for demonstrate the previous results. From these results, we observe the ACF residual data (Figure 9) are inside in the confidence interval (±0.0816), this implies that the residual values are uncorrelated and describe by a white noise. So, the developed algorithm can be used to model the data video-packets transmission for the network management.

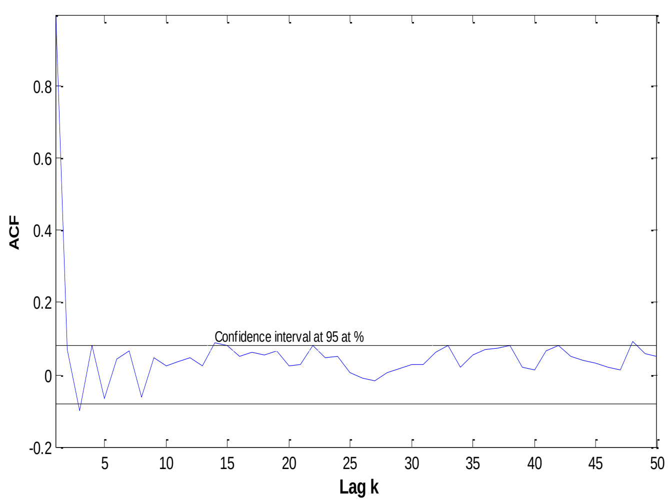


**Figure 9. ACF of the Residual Time Series (for side B)**

## 7. Conclusion

In this work, we have developed a supervised algorithm based on the third order moments for identifying parameters of quadratic systems excited by non Gaussian and independent identically distributed signals. The simulation results and the comparison with RLS method using different SNR and sample sizes show that the developed algorithm is adequate for identifying quadratic systems.

We have presented also the simulation of the data video-packets transmission in networks from video server. The obtained results show that, the sequences of generated values have the same statistical characteristics as the real data and the developed model provide satisfactory performances. So, we conclude that the quadratic systems can be used as alternative methods to generate the video-packets transmission at different times.

In the perspective we will be to use the developed model for forecasting internet traffic.

## References

[1] J. K. Tugnait, “Approaches to FIR system identification with noisy data using higher order statistics”, IEEE Trans. on ASSP, vol. 38, **(1990)**, pp.1307-1317.
[2] R. Pan and C. L. Nikias, “The complex spectrum of higher order cumulants and non-minimum phase system identification”, IEEE Trans. on ASSP, vol. 36, **(1998)**, pp. 186-205.
[3] S. Safi and A. Zeroual, “Modelling solar data using high order statistics”, AMSE Advances in Modelling and Analysis, vol. 6, **(2001)**, pp. 1-16.
[4] J. Liu and Z. He, “Blind identification and equalization using higher-order cumulants and ICA algorithms”, Proc. of the International Conference Neural Networks and Brain (ICNN and B’98), Beijing, **(1998)** October.
[5] S. Safi and A. Zeroual, “Blind parametric identification of non-Gaussian FIR systems using higher order cumulants”, Int. Journal of Systems Science, vol. 35, **(2004)**, pp. 855–867.
[6] A. Y. Kibangou, G. Favier and M. M. Hassani, “Laguerre-Volterra filters optimization based on Laguerre spectra”, EURASIP Journal on Applied Signal Processing, vol. 17, **(2005)**, pp. 2874-2887.
[7] X. Jing, Z. Lang and S. A. Billings, “Determination of the analytical parametric relationship for output spectrum of Volterra systems based on its parametric characteristics”, J. Math. Anal. Appl., vol. 351, **(2009)**, pp. 694–706.
[8] H. -Z. Tan and Z. -Y. Mao, “Blind identifiability of quadratic non-linear systems in higher-order statistics domain”, Int. J. Adapt. Control Signal Process. vol. 12, (1998), pp. 567-577.
[9] H. -Z. Tan and T. Aboulnasr, “Tom-based blind identification of nonlinear Volterra systems”, IEEE Trans. on Instrumentation and Measurement, vol. 55, **(2006)**, pp. 300-310.
[10] H. -Z. Tan and T. W. S. Chow, “Blind identification of quadratic nonlinear models using neural networks with higher order cumulants”, IEEE Trans. on Industrial Electronics, vol. 47, **(2000)**, pp. 687- 696.
[11] J. Antari, S. Chabaa, A. Zeroual, R. Iqdour and S. Safi, “Identification of quadratic systems using high order cumulants and neural networks: Application to model the delay of video-packets transmission”, Journal of Applied Soft Computing- ASOC, Elsevier, vol. 11, no. 1, **(2011)**, pp. 1-10.
[12] J. Antari and A. Zeroual, “Blind identification of diagonal kernels of non linear process using cumulants: Application to modelling internet traffic”, Int. Journal on Information and Communication Technologies-IJICT, vol. 1, **(2008)**, pp. 37-42.
[13] G. B. Giannakis and J. M. Mendel, “Identification of non-minimum phase systems using higher order statistics”, IEEE Trans. ASSP, vol. 37, **(1989)**, pp. 360- 377.
[14] M. Masugi and T. Takuma, “Using a Volterra system model to analyze nonlinear response in video-packet transmission over IP networks”, Communications in Nonlinear Science and Numerical Simulation, vol. 12, **(2007)**, pp. 411–421.
[15] S. Safi, A. Zeroual and M. M. Hassani, “Modelling solar half-hour data using fourth order cumulants”, Int. J. of Solar Energy, vol. 22, **(2002)**, pp. 67-81.
[16] J. L. Lacoume, P. O. Amblard and P. Comon, “Statistiques d’ordre supérieur pour le traitement du signal”, Published by Masson, Paris, **(1997)**.
[17] S. Safi, A. Zeroual and M. M. Hassani, “Prediction of global daily radiation using higher order statistics”, Renewable Energy, vol. 27, **(2002)**, pp. 647- 666.

[18] M. Iqbal, “An introduction to solar radiation”, Academinc press, **(1993)**.
[19] P. J. Brokwell and R. A. Davis, “Time series theory and methods”, 2nd ed., Springer-Verlag, **(1991)**.