EEG TIME SERIES ANALYSIS WITH EXPONENTIAL AUTOREGRESSIVE MODELLING

Tugce Balli and Ramaswamy Palaniappan

Department of Computing and Electronic Systems, University of Essex,
Wivenhoe Park, CO4 3SQ, United Kingdom

ABSTRACT

This paper proposes the use of exponential autoregressive (EAR) model for modelling of time series that are known to exhibit non-linear dynamics such as random fluctuations of amplitude and frequency. Biological signal (bio-signal) such as electroencephalogram (EEG) is known to exhibit non-linear dynamics. Such signals cannot be modelled with traditional linear modelling techniques like autoregressive (AR) models as these models are known to provide only an approximation to the underlying properties of the non-linear signals. In this study, the suitability of EAR models as compared to AR models is shown using EEG signals in addition to several non-linear benchmark time series data where improved signal to noise ratio (SNR) values are indicated by the EAR models. Overall, the results indicate that use of EAR modelling which has yet to be exploited for bio-signal time series analysis has the huge potential in the characterisation and classification of EEG signals.

Index Terms— Exponential autoregressive, Electroencephalogram, Non-linear time series analysis, Genetic algorithm

1. INTRODUCTION

Linear modelling techniques have been widely used for analysis of time series as they are simpler in principle as compared to their non-linear counterpart. Among these modelling techniques, autoregressive (AR) modelling is one of the most commonly used method [1,2]. Basically this approach assumes that the time series is a regressive process and estimates the model coefficients according to that assumption. However the time series with non-linear fluctuations cannot be modelled precisely with linear modelling techniques like AR. In these situations, a technique that can model the non-linear fluctuations, which will provide a better representation of time series, is required.

Haggan and Ozaki [3,4] introduced exponential autoregressive (EAR) model for modelling non-linear fluctuations in time series. They stated that the analysis of stochastic processes have been mostly done using some form of linear time series modelling and this can only provide an approximation to the underlying properties of the signals. Besides it is found that many signals exhibiting random vibrations display non-linear behaviour, hence a non-linear model that gives a good approximation to the underlying properties of a signal is required. Accordingly, Haggan and Ozaki proposed EAR model which exhibits certain features of random vibrations that does not occur in linear models namely the amplitude-dependent frequency, jump phenomena and limit cycle.

In this study, we set out to show the improved signal to noise ratio (SNR) of a common biological signal (bio-signal), namely electroencephalogram (EEG) signal plus several other standard time series data reconstructed using coefficients from EAR as compared to AR modelling techniques. Binary genetic algorithm (BGA) hybridised with recursive least squares (RLS) method has been used to obtain the non-linear and linear coefficients of the EAR model.

The rest of this paper is organised into three sections. In section 2, the EAR model and estimation of EAR model coefficients are discussed in detail. Section 3 presents the experimental results of fitting EAR and AR models to some of the benchmark non-linear time series and EEG signals. And in the final section, the conclusions and possible future improvements to this study are presented.

2. METHODOLOGY

In this section, conventional AR model, EAR model and method of using GA to obtain the non-linear EAR model coefficients will be discussed.

2.1 Autoregressive model

An autoregressive model of order p is defined by

$$x_t = -\sum_{k=1}^{p} a_k \cdot x_{t-k} + e_t \;, \qquad (1)$$

where $\boldsymbol{x_t}$ is the data at sampled point n, $\boldsymbol{a_k}$ are coefficients of the AR model and $\boldsymbol{e_t}$ is Gaussian white noise with mean zero.

The model order, $\boldsymbol{p}$ can be selected as order with minimum Akaike Information Criterion (AIC) [5], defined by

$$AIC = (N - p)\ln \sigma_e^2 + 2(2p+1), \qquad (2)$$

978-1-4244-1643-1/08/$25

with N as the length of time series, and σ_e^2 as the residual variance of the AR model which is given by

$$\sigma_e^2 = \frac{1}{N-p}\sum_{t=p+1}^{N}(x_t - \sum_{i=1}^{p} a_i \cdot x_{t-i})^2, \quad (3)$$

where AR coefficients are estimated by Burg's method [6].

2.2 Exponential autoregressive model

The EAR model was originally defined by a second order autoregressive form by making coefficients amplitude dependent [3,4,7]:

$$x_t = (\varphi_1 + \pi_1 \cdot e^{-\gamma \cdot x_{t-1}^2}) \cdot x_{t-1} + (\varphi_2 + \pi_2 \cdot e^{-\gamma \cdot x_{t-1}^2}) \cdot x_{t-2} + e_t, \quad (4)$$

where $\boldsymbol{\varphi_1}$, $\boldsymbol{\varphi_2}$, $\boldsymbol{\pi_1}$, $\boldsymbol{\pi_2}$ and $\boldsymbol{\gamma}$ are EAR coefficients, $\boldsymbol{e_t}$ is Gaussian white noise with mean zero and $\boldsymbol{x_t}$ is data at sampled point t. This second order model can be extended to a general model of order p, defined by

$$x_t = \sum_{k=1}^{p}(\varphi_k + \pi_k \cdot e^{-\gamma x_{t-1}^2}) \cdot x_{t-k} + e_t \quad (5)$$

,

where $\boldsymbol{\varphi}$, $\boldsymbol{\pi}$ and $\boldsymbol{\gamma}$ are EAR coefficients, $\boldsymbol{p}$ is the model order, $\boldsymbol{e_t}$ is Gaussian white noise with mean zero and $\boldsymbol{x_t}$ is data at sampled point t.

Note that the non-linearity of the EAR method comes from the exponential term $e^{\gamma \cdot x(t-1)^2}$, which makes the series globally non-linear [7]. If non-linear parameter γ is set to zero, the equation will become an ordinary linear AR model with coefficients $a_p = \varphi_p + \pi_p$ such that

$$x_t = \sum_{k=1}^{p} a_k \cdot x_{t-k} + e_t. \quad (6)$$

As mentioned earlier there are certain features that does not occur in linear models namely the amplitude dependent frequency, jump phenomena and limit cycle theorem. For EAR model, Haggan and Ozaki had shown that amplitude dependent frequency and jump phenomena occur and besides these, there are some conditions that should be checked for the existence of limit cycle [3]. These are:

(1) All roots of $z^p - \varphi_1 \cdot z^{p-1} - ... - \varphi_p = 0$ should lie inside the unit circle;
(2) All roots of $z^p - (\varphi_1 + \pi_1) \cdot z^{p-1} - ... - (\varphi_p + \pi_p) = 0$ should not lie inside the unit circle;
(3) The condition $(1 - \sum_{i=1}^{p} \varphi_i)/(\sum_{i=1}^{p} \pi_i) > 1$ or <0 should hold for the existence of limit cycle.

The estimation of the $2p$+1 coefficients $\{\gamma, \varphi_i, \pi_i \mid i=1,2,...,p\}$ of the EAR model is a non-linear optimisation problem, hence is complicated with increasing model order. In order to achieve this task, BGA hybridised with RLS algorithm is used following a previous study [7]. The non-linear coefficient, γ, is determined with BGA, and once this coefficient is obtained the model becomes a linear regression problem in which the linear coefficients $\{\varphi_i, \pi_i \mid i=1,2,...,p\}$ are determined by RLS algorithm. Moreover the model order is selected as the order with minimum AIC value [5].

Genetic algorithms are search algorithms inspired by the natural selection and natural genetics which are used to solve optimisation problems. Initially there is a population of candidate solutions (chromosomes) to the optimisation problem and the solutions evolve toward better solutions according to the principles of natural selection (i.e. survival of the fittest) [8]. For the selection of the fittest chromosome, a fitness function that measures the performance of a chromosome in the population must be defined according to the optimisation problem to be solved.

Initially a certain number of chromosomes are randomly generated to form the initial population that represents possible values of γ, the non-linear coefficient of the EAR model. In the next step, each chromosome in the population is converted into real values in the range of γ_{min} to γ_{max}. The range of values are selected according to problem under study, here we converted the chromosomes into real-valued ones in the range of 0 to 10. For each of the chromosomes converted into a real-value, the remaining coefficients $\boldsymbol{\varphi_1},\ldots, \boldsymbol{\varphi_p}, \boldsymbol{\pi_1},\ldots, \boldsymbol{\pi_p}$ with p as the model order are estimated using RLS algorithm.

In order to estimate the coefficients using RLS algorithm the EAR model can be defined by the following regression formula

$$x_t = Z^T(t)\theta + e_t, \quad (7)$$

where

$$Z(t) = [x_{t-1},..., x_{t-p}, e^{-\gamma(i)x_{t-1}^2} \cdot x_{t-1},..., e^{-\gamma(i)x_{t-1}^e} \cdot x_{t-p}]^T, \quad (8)$$

$$\theta = [\varphi_1,..., \varphi_p, \pi_1,...., \pi_p]. \quad (9)$$

Following this, the linear coefficients can be estimated by

$$\theta(t+1) = \theta(t) + K(t+1) + \varepsilon(t+1), \quad (10)$$

$$K(t+1) = P(t) \cdot Z(t+1) \cdot (1 + Z^T(t+1) \cdot P(t) \cdot Z(t+1))^{-1}, \quad (11)$$

$$P(t+1) = P(t) \cdot (1 - K(t+1) \cdot Z^T(t+1)), \quad (12)$$

$$\varepsilon(t+1) = x_{t+1} - Z^T(t+1) \cdot \theta(t). \quad (13)$$

with $\theta(N)$ as the final estimated coefficients. Once the linear and non-linear coefficients of EAR model are estimated, next step is to calculate the fitness value of each chromosome which is the residual variance of EAR model, defined by

$$J(\theta) = \sigma_e^2 = \frac{1}{N-p}\sum_{t=p+1}^{N}(x_t - \sum_{i=1}^{p}(\varphi_i + \pi_i \cdot e^{-\gamma \cdot x_{t-1}^2})x_{t-i})^2, \quad (14)$$

where θ=($\varphi_1,\ldots, \varphi_p, \pi_1,\ldots, \pi_p$), $\boldsymbol{p}$ is the model order, and $\boldsymbol{N}$ is the length of time series.

The next generations are created by selecting the chromosomes based on the fitness function. There are several selection algorithms for performing the selection process, some common examples are: elitist selection, roulette wheel selection, and rank based methods such as tournament selection [8]. The next step is to employ genetic operators such as crossover, mutation and inversion on the selected chromosomes [8]. This process of new population generation is repeated until the maximum number of generations is reached. Table 1 summarises the parameters of GA used in the experimental study.

Table 1: GA parameters

Coding of genes	Binary coding converted to real value in the range [0,10]
Fitness function	Residual variance (σ_e^2) of EAR model
Length of chromosomes	10
Population size	20
Selection (reproduction)	Tournament Selection (50% of the population) and Roulette Wheel Selection (50% of the population)
Crossover type, probability	Two point crossover, 0.5
Mutation type, probability	Randomly selected bits, 0.1
Inversion type, probability	Inversion between two randomly selected points, 0.01
Maximum Generations	50

Figure 1 shows the block diagram of the GA method used for the estimation of EAR coefficients.

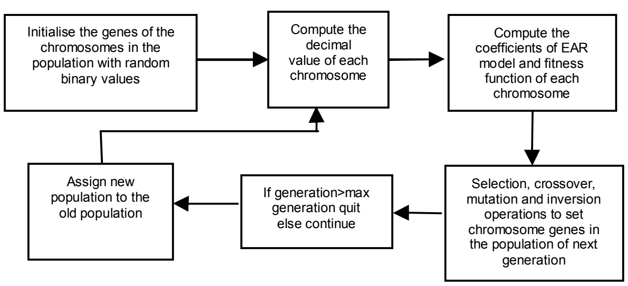

Figure 1: Block diagram on the use of GA

3. EXPERIMENT, RESULTS AND DISCUSSIONS

In this study, different examples of EEG and other non-linear time series are utilised to compare the performance of EAR and AR models. For performance comparison, the time series are reconstructed with corresponding coefficients (for both AR and EAR modelling techniques) and the signal to noise ratio (SNR) between original and reconstructed signals are compared.

In total, nine data sets have been used, the first data set is an artificial time series generated from a second order EAR model which is the same model as used earlier by Haggan and Ozaki [3] while the remaining five time series are some of well known non-linear benchmark data sets [9-12]. Apart from these, three EEG signals recorded from one healthy subject and one epilepsy patient during seizure and seizure free intervals have been modelled using EAR and AR models.

Data set 1: 1000 data points are generated from a second order EAR model which is defined by

$$x_t = (1.95 + 0.23 \cdot e^{-\gamma \cdot x_{t-1}^2}) \cdot x_{t-1} - (0.96 + 0.24 \cdot e^{-\gamma \cdot x_{t-1}^2}) \cdot x_{t-2} + e_t$$

where $\boldsymbol{e_t}$ is Gaussian white noise with mean 0 and variance 0.01. The model orders from 2 to 15 were used for estimating the EAR and AR model coefficients to determine the optimum model order.

The results indicated that the final model for AR method is the fourth order model which minimises the AIC with SNR of 45.52. While the final model for EAR method is the second order model which minimises the AIC and satisfies the limit cycle theorem with SNR of 53.38. The estimated linear coefficients of the EAR model are

Table 2: Linear coefficients of EAR model for data set 1

i	**1**	**2**
$\boldsymbol{\Phi_i}$	1.9505	-0.96057
$\boldsymbol{\Pi_i}$	0.23459	-0.24091

The estimated coefficients are very close to the actual coefficients of the time series. Therefore this data set validates the effectiveness of the estimation of EAR coefficients with the proposed BGA hybridised with RLS algorithm.

Data sets 2-6: Data sets 2-6 are well known non-linear time series namely LYNX, far-infrared laser, Sunspot, Mackey-Glass and Henon data sets [9-12]. The Figure 2 below shows the results of EAR and AR models applied to non-linear benchmark time series.

It is clear from the figure that the EAR model gives a better reconstruction of corresponding time series compared to AR model as it yields improved SNR values for different model orders. In other words EAR model gives an improved characterisation of the corresponding non-linear time series.

In order to further illustrate the performance comparison of AR and EAR methods, we have utilised three classes of EEG signals recorded from one healthy subject, one epilepsy patient during seizure free interval and during seizure interval. Each of these EEG signals consists of 4096 sample points. These signals are publicly available [13]. The results of the EAR and AR methods applied to these signals are shown in Figure 3. It can be seen that the SNR values obtained from the EAR method are improved compared to AR method indicating that EAR coefficients give better reconstruction and characterisation of EEG signals.

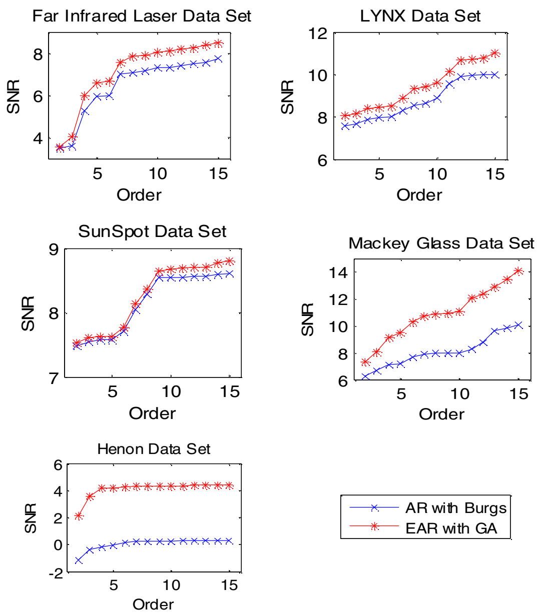

Figure 2: SNR values of AR and EAR methods for data sets 2-6

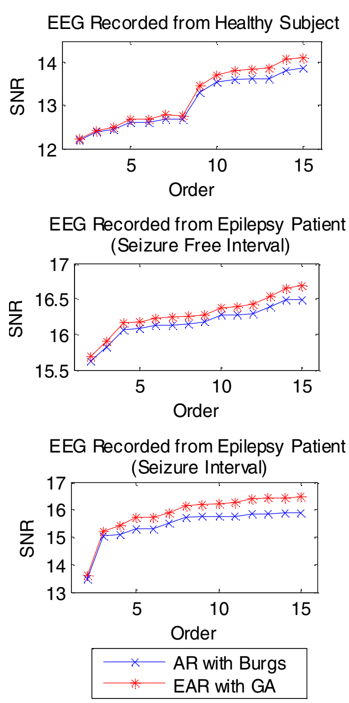

Figure 3: The SNR values of AR and EAR methods for EEG signals

4. CONCLUSION

In this study, the reconstruction (i.e. modelling) ability of EAR and AR models has been compared. The EAR model was developed using BGA and RLS algorithms while the AR model was developed using Burg's method. The overall results indicated that the EAR model provides a better characterisation for EEG signals and some examples of non-linear benchmark time series as it gave better SNR values than AR model.

For future work, the EAR model could also be used for many bio-signal applications such as mental condition monitoring (such as epilepsy) and mental task classification using EEG signals and diagnosis of certain diseases using ECG signals. As a further novel extension, the chromosomes with real values could be investigated where all of the coefficients are estimated using GA method. Apart from that the EAR model could be modelled with more than one exponential term, which would likely further improve the modelling performance.

5. REFERENCES

[1] Z.A. Keirn, and J.I. Aunon, "A new mode of communication between man and his surroundings", *IEEE Trans. Biomed. Eng.*, vol.37, no.12, pp. 1209 – 1214, 1990.

[2] D. Ge, N. Srinivasan, and S.M. Krishnan, "Cardiac arrhythmia classification using autoregressive modelling", *Biomed. Eng. Online*, 2002. <http://www.biomedical-engineering-online.com/ content/1/1/5>

[3] V. Haggan, and T. Ozaki, "Modelling nonlinear random vibrations using an amplitude-dependent autoregressive time series model", *Biometrika*, vol. 16, no.1, pp. 189-196, 1981.

[4] H. Kantz, and T. Schreiber. *Nonlinear Time Series Analysis*, 2nd ed., Cambridge, 2004.

[5] H. Akaike, "A new look at the statistical model identification", *IEEE Trans. on Automatic Control*, vol. 19, no.6, pp. 716–723, 1974.

[6] R. Shiavi, *Introduction to Applied Statistical Signal Analysis*, 2nd ed., Academic Press, 1999.

[7] Z. Shi, and H. Aoyama, "Estimation of exponential autoregressive time series model by using genetic algorithm", *Journal of Sound and Vibration*, vol. 205, no.3, pp. 309-321, 1997.

[8] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley, 1989.

[9] Datasets for Time-Series Analysis, Uni. Carlos III of Madrid, <http://tracer.uc3m.es/tws/TimeSeriesWeb/index.html>

[10] N. Gershenfeld, Santa Fe Time Series Competition Data, <http://www-psych.stanford.edu/~andreas/TimeSeries/ SantaFe.html>

[11] E. A. Wan, Time Series Data, Oregon Health & Science Uni., <http://www.cse.ogi.edu/%7Ericwan/data.html>

[12] M. Henon, "A two-dimensional mapping with a strange attractor", *Commun. Math. Phys.*, vol. 50, pp.69-77, 1976.

[13] EEG Time Series Download Page, <http://www.epileptologie-bonn.de/front_content.php? idcat=193&lang=3&changelang=3>