Turkish Journal of Electrical Engineering and Computer Sciences

Volume 18 | Number 1 Article 2

1-1-2010

# Determination of autoregressive model orders for seizure detection

SERAP AYDIN

Turk J Elec Eng & Comp Sci, Vol.18, No.1, 2010, © TÜBİTAK
doi:10.3906/elk-0906-83

# Determination of autoregressive model orders for seizure detection

**Serap AYDIN**
*Ondokuz Mayıs University, Faculty of Engineering*
*Department of Electrical and Electronics Engineering*
*Kurupelit, Samsun-TURKEY*
*e-mail: drserapaydin@hotmail.com*

**Abstract**

*In the present study, a step-wise least square estimation algorithm (SLSA), implemented in a Matlab package called as ARfit, has been newly applied to clinical data for estimation of the accurate Auto-Regressive (AR) model orders of both normal and ictal EEG series where the power spectral density (PSD) estimations are provided by the Burg Method. The ARfit module is found to be usefull in comparison to a large variety of traditional methods such as Forward Prediction Error (FPE), Akaike's Information Criteria (AIC), Minimum Description Lenght (MDL), and Criterion of Autoregressive Transfer function (CAT) for EEG discrimination.*

*According to tests, the FPE, AIC and CAT give the identical orders for both normal and epileptic series whereas the MDL produces lower orders. Considering the resulting PSD estimations, it can be said that the most descriptive orders are provided by the SLSA. In conclusion, the SLSA can mark the seizure, since the estimated AR model orders meet the EEG complexity/regularity such that the low orders indicate an increase of EEG regularity in seizure. Then, the SLSA is proposed to select the accurate AR orders of long EEG series in diagnose for many possible future applications. The SLSA implemented by ARfit module is found to be superior to traditional methods since it is not heuristic and it is less computational complex. In addition, the more reasonable orders can be provided by the SLSA.*

**Key Words:** *EEG, seizure, AR model, stepwise least square algorithm.*

## 1. Introduction

The activity of brain cells results in both cortical and intracortical measurable electrical activities, the so-called EEG. These meaningful signals have frequently been analyzed by visual inspection in diagnoses. However, computer based quantitative EEG analysis has become a focus of attention due to the fact that statistical signal processing tools provide the more reliable results in association with neuro-pathological conditions. In particular, spectral EEG analysis is object to the study of epilepsy characterized by seizures involving abnormal, rhythmic discharges of cortical neurons such that these synchronic neuro-electrical activities lead to uncontrollable high amplitude oscillations in ictal periods [1]. It is well known that EEG series originated by seizure have less regularity than normal EEG [2]. To obtain the signal regularity, several indicators such as Lyapunov exponents, Hurst exponents [3] and, entropies [4] have been applied to EEG series. Since AR model order can indicate the degree of signal regularity, both cortically recorded normal EEG series and intracortical ictal EEG series are analyzed by using the Burg Method (BM) based on the AR model.

Determination of the AR model order arises in many fields of signal processing applications to find the appropriate dimensionality which fits a given set of observations. In last biomedical applications, AR modeling has been used in quantitative EEG analysis [5–7]. In recent works, the coefficients of the determined AR model have been considered as features to identify and then classify the EEG series [8, 9]. However, large variety of methods has not been compared in selecting of an optimum order to represent the EEG signals yet. The selection of the accurate order plays the most important role in AR modeling of time series [10, 11].

The study of epilepsy is one of the important applications of EEG. Epilepsy is characterized by seizures and involves abnormal, rhythmic discharges of cortical neurons. Abnormal neuronal electrical activities occur synchronously, creating a large amplitude signal and leading to uncontrollable oscillations in ictal periods [3]. In the present study, methodologically selected AR model orders are assumed to be correlated with the degree of EEG complexity in case of epilepsy. Then, five different algorithms such as Forward Prediction Error (FPE), Akaike's Information Criteria (AIC), Minimum Description Length (MDL), Criterion of Autoregressive Transfer function (CAT) and the ARfit module driven by a stepwise least squares algorithm (SLSA) are attempted to select the AR model order in association with a high quality clinical data set where the Burg Method (BM) is used to compute the power spectral density (PSD) of EEG series of interest. Description of both data collection protocols and methods are given in the following section.

# 2. Materials and methods

## 2.1. Data collection

The publicly available data described in [12] was downloaded from http://www.meb.uni-bonn.de/epileptologie/cms/front_content.php?idcat=193. In this section, only a short description is presented but refer to reference [12] for further details. The complete data set consists of five subsets each containing 100 single-channel EEG signals of 23.6 sec. Each segment has been selected and cut out from the continuous multichannel EEG recordings after visual inspection for artifacts originated by muscle activity or eye movements. Among them, three sets denoted by SetA, SetD, SetE are analyzed in the present study. SetA was taken from the surface EEG recordings of five healthy subjects relaxed in an awaken state with eyes open where the standardized electrode placement technique was used. SetD was measured from five patients in the epileptogenic zone during seizure free intervals. SetE consists of epileptic EEG signals collected from five different epileptic patients, recorded during the occurrence of the epileptic seizures from intracranial electrodes. SetA was recorded extra-cranially, whereas both SetD and SetE were intracranial records.

The depth electrodes were implanted symmetrically into the hippocampal formations and strip electrodes were implanted onto the lateral and basal regions (middle and bottom) of the neocortex. The epileptic EEG segments were selected from all the recording sites exhibiting ictal activity in SetE. These EEG signals were recorded with 128-channel amplifier system, using an average common reference. After a 12-bit analog-to-digital conversion, the data was written continuously onto the disk of a data acquisition computer system at a sampling rate of 173.61 Hz with bandpass filter settings at 0.53–40 Hz (12 dB/octave).

## 2.2. Methods

In general, EEG series are generally characterized by their PSD estimations since they are stochastic signals. In particular, intracortical EEG measurements are analyzed by using parametric predictors instead of non-parametric methods to obtain higher frequency resolution. It is theoretically stated that nonparametric PSD estimations suffer from spectral leakage effects originated from inherent windowing where the autocorrelation function is assumed to be zero outside the window [10, 11].

Among the parametric and nonparametric PSD predictors, the BM is found to be useful to observe the EEG synchronization in seizure [13]. In fact, AR model having less coefficients is the mostly practiced model among the linear models consisting of AR, Moving Average (MA) and ARMA for representation of a narrow spectrum [10,11]. Clearly, the AR model, which is a causal all-pole model driven by a white noise, provides us to track changes in the source of EEG [10].

### 2.2.1. Burg method

The BM assumes that the observed data denoted by $x(n)$ is output of a linear system characterized by a transfer function, $H(z)$. Then, $x(n)$ meets an AR model with order $p$ of the form

$$x(n) = -\sum_{i=1}^{p} a_p(i)x(n-i). \tag{1}$$

Here, $H(z) = 1/A(z)$ where $A(z) = 1 + \sum_{i=1}^{p} a_p(i)z^{-i}$ satisfies the Levinson Durbin recursion algorithm. Then, the PSD of the data is computed as

$$\hat{P}(f) = \frac{\hat{e}_t^2}{\left|1 + \sum_{i=1}^{p} a_p(i)e^{-j2\pi f}\right|^2}, \tag{2}$$

where the least square error denoted by $\hat{e}_t^2$ (sum of the forward and backward errors) is minimized [9]. Generally speaking, the BM is superior to nonparameric methods due to the following properties: (A) It does not apply window to data, and then it does not dependent on the unrealistic assumption (the autocorrelation sequence is not assumed to be zero outside the window), (B) Its frequency resolution is high. (C) Both forward and backward prediction errors are minimized in least squares sense, (D) It always yields a stable AR model, (E) It is computationally efficient [11]. The BM results a stable AR model with high frequency resolution. In addition, it has no spectral leakage problem [11]. Besides those important advantages, selection of the AR model highly affects the performance of the BM [10, 11]. The model order must be selected correctly to obtain true spectra. In past years, many techniques consisting of FPE, AIC, CAT and MDL were proposed on selection of $p$ [10,11]. In recent years, ARfit module has been proposed to predict the optimum model order where the errors for both

FPE and AIC are minimized together [14]. Those traditional order estimation methods and the ARfit algorithm are briefly presented into the following subsections.

#### 2.2.2. The traditional order selection algorithms

Selection of an appropriate AR model order plays an important role in many fields of acoustics, seismology, physics and biomedical engineering. In general, the order of $x(n)$ is determined by the minimum of the following criteria:

$$FPE(p) = \frac{N+p}{N-p}\hat{\sigma}^2,$$

$$AIC(p) = N\log(\hat{\sigma}^2) + 2p,$$

$$CAT(p) = \frac{1}{N^2}\sum_{k=1}^{p}\frac{N-k}{\hat{\sigma}^2} - \frac{N-p}{N\hat{\sigma}^2}$$

$$MDL(p) = -\log(\hat{\sigma}^2) + \frac{1}{2}p\log(N),$$

where $\hat{\sigma}^2$ is estimated prediction error variance of the white noise at order $p$ and $N$ is the number of samples in $x(n)$ [15].

#### 2.2.3. The stepwise least square algorithm

In recent years, the ARfit module that implements the SLSA has been proposed for efficient estimation of multivariate AR (MVAR) model parameters to observe the changes in dynamic systems with respect to high dimensional data [16]. In an AR process, a univariate time series is characterized by modeling the current value of the variable as a weighted linear sum of its previous values [17]. In fact, an AR model extends to multiple time series characterized by modeling the current values of all variables as a linear sum of their previous values in a MVAR process. Both models are considered for linear systems which are described by the principle of superposition such that the input related responses are independent of each other.

A time series is assumed to be modeled by a MVAR model with order of $p_{\min}, ..., p_{\max}$ in the algorithm of ARfit [16]. A publicly available package called as Algorithm808 implements the SLSA with ARfit module [14]. The SLSA described in detail in the reference [16], is performed by *Algorithm 808* where the relevant results are limited by an approximate (95%) confidence interval [14].

The ARfit module uses a SLSA to compute the parameters of estimated AR($p$) models of successive orders $p_{\min} < p < p_{\max}$ via a regularized QR factorization of the data as proposed in [18]. In this assessment, both the effect of rounding errors and the data errors are reduced by implementing the singular value decomposition (SVD) based regularization techniques introduced by Hansen [19]. It was stated that the SVD based regularization approaches gives reliable estimations in brain activities [20].

## 3. Results

Each data set consists of five subsets, each containing 20 single epoches collected from a single volunteer. Patients are referred by numbers as $k = 1, ..., 5$. For each data set, the following steps are satisfied:

1. For each patient, AR model orders denoted by p are computed in association with 20 epoches. (i.e., compute $p_k(i)$ for $i = 1, ..., 20)$;

2. For each patient, average AR order is calculated as $m_k = \sum_{i=1}^{20} \frac{p_k(i)}{20}$,

3. For five patients, the mean of the average AR order is calculated as $a = \sum_{k=1}^{5} \frac{m_k}{5}$,

4. The PSD estimations are plotted with respect to the specified mean orders denoted by $a$ in Table 1.

**Table 1.** The predicted AR model orders for three data sets.

| | FPE, AIC, CAT | | | MDL | | | ARfit | | |
|---|---|---|---|---|---|---|---|---|---|
| | SetA | SetD | SetE | SetA | SetD | SetE | SetA | SetD | SetE |
| $m_1$ | 49 | 35 | 28 | 25 | 17 | 14 | 41 | 34 | 23 |
| $m_2$ | 46 | 33 | 27 | 23 | 15 | 13 | 42 | 36 | 23 |
| $m_3$ | 43 | 33 | 25 | 21 | 15 | 12 | 43 | 36 | 22 |
| $m_4$ | 42 | 33 | 24 | 19 | 13 | 12 | 40 | 34 | 26 |
| $\mathrm{m}_5$ | 43 | 34 | 24 | 19 | 13 | 12 | 43 | 35 | 24 |
| $a$ | 45 | 34 | 26 | 21 | 15 | 12 | 42 | 35 | 24 |

FPE, AIC, CAT and ARfit provide the same average orders for each data set. Also, the results provided by the MDL are almost half of these orders commonly estimated by using the other methods. However, both cases show that each set meets a different AR model order due to their different EEG characteristics depending on the healthy conditions.

Among the data sets, the highest AR order, 45, is estimated for SetA, whereas the lowest order, 26, is estimated for SetE. For SetD, the estimated AR model order, 34, is middle of them. It can be said that the lower order is estimated as healthy condition becomes worse when the order is predicted methodologically by using any of the algorithms consisting of FPE, AIC, CAT and ARfit.

To show the importance of the selected AR order on PSD estimations, PSD estimations of data sets with respect to specific orders are given in the figures. Figure 1 consists of three PSD estimations of SetA for three orders. It can be said that the higher frequency resolution in estimations is obtained as the AR model order is high like 45. Except the MDL, the performed algorithms provide the high orders. Besides, the highest frequency resolution of PSD estimations is observed when the ARfit is used.

Figure 2 and Figure 3 show the PSD estimations of SetD and SetE, respectively. For both data sets, the most useful results are originated by the AR orders selected by using any of the algorithms consisting of FPE, AIC, CAT and ARfit. The AR orders obtained by using the MDL do not provide the high frequency resolution for all data sets.

## 4. Discussion and conclusion

Both cortical normal EEG measurements and intracortical epileptic EEG series, in addition to intracortical ictal records, are analyzed in the present study. Several traditional methods and the ARfit are implemented in Matlab to estimate the optimum AR model orders of these diagnostic records. Among them, the ARfit algorithm is found to be reliable and superior.

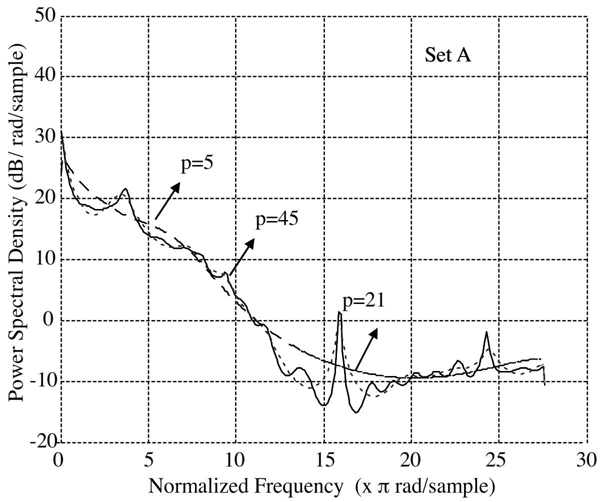


**Figure 1.** PSD estimations for SetA with respect to particular AR model orders.

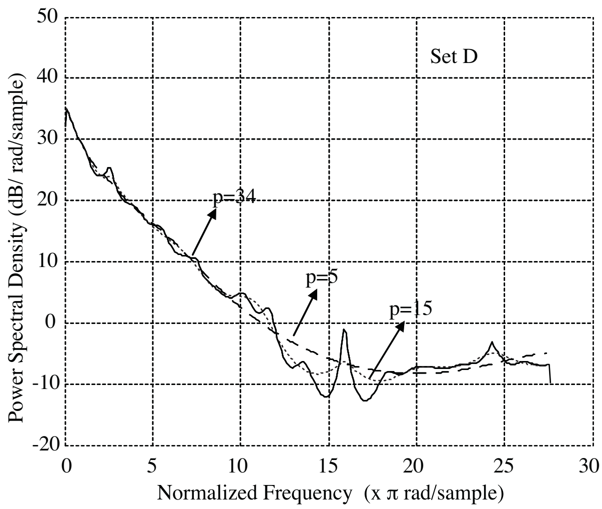


**Figure 2.** PSD estimations for SetD with respect to particular AR model orders.

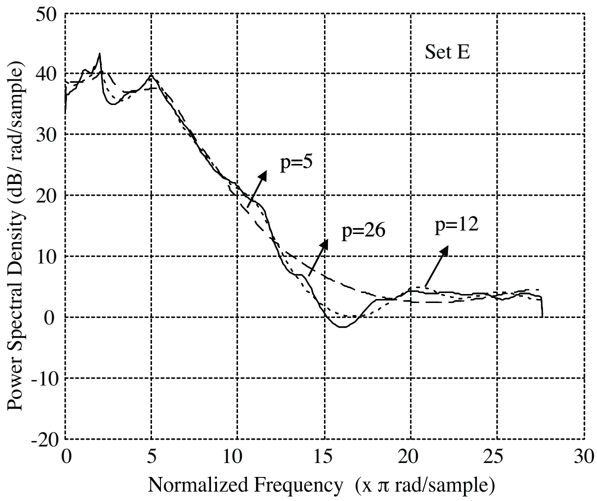


**Figure 3.** PSD estimations for SetE with respect to particular AR model orders.

In literature, it was stated that the changes on the time series components such as oscillation periods and damping times can be characterized by MVAR models with respect to their SVD pairs [14]. Then, the ARfit module is developed to detect these changes. The current results show that the electrophysiological variations on EEG series can also be identified by using the ARfit. In other words, the meaningful sharp oscillations in EEG can be detected owing to the implementation of the ARfit module. Moreover, neither spurious peaks in the spectrum (in case of too high order), nor loss of spectral detail (in case of excessively low order) are encountered in the assessment of the ARfit.

Also, regarding as the PSD estimations, it can be said that the useful AR model orders can also be estimated by using the algorithms of FPE, AIC and CAT. Nevertheless, FPE, AIC and CAT are known to be heuristic and more subjective choices in many applications [21]. However, the ARfit is not heuristic and it is considerable less computational complex such that the optimum model can be estimated about $p_{\max} - p_{\min} + 1$

times faster than with those traditional algorithms that require $p_{\max} - p_{\min} + 1$ separate QR factorizations.

The other criterion so called the MDL can not produce the adequate orders. In selecting of AR model order, the methods of FPE and AIC minimize the average error variance for a one-step and an information theoretical function, respectively [21]. The methods of FPE and AIC do not yield consistent estimates of the model order as the length of the time series increases whereas both are asymptotically equivalent [21]. The MDL criterion, also called the Bayesian information criterion, uses a penalty function which provides consistent estimation of the model order [22].

In the ARfit module, the both the effect of rounding errors and data errors are minimized in the SLSA in association with determined approximate confidence interval [16]. The SLSA is stated as a numerically stable procedure in reference [23]. The using of the SLSA provides to obtain a more reliable residual noise variance. In fact, ARfit solves a regularized estimation problem with respect to an ill-conditioned moment matrix weighted with a regularization parameter. In summary, ARfit module is proposed as very useful, fast and efficient tool in brain activities to estimate a reliable AR model order. The results show that the estimated AR model orders can be used as markers to support the clinical findings in diagnose when ARfit is used.

# References

[1] Guyton A.C: Text Book of Medical Physiology, Saunders, sec.2, 968, (1986).

[2] John M., Peter J.: Epilepsy as a Dynamic Disease, Bio and Med Physics, ISBN: 978-3-540-42762-9, (2003).

[3] Kannathala N., Acharyab U.R., et al: Characterization of EEG-A comparative study, Comp Meth and Prog in Biomed, 80, 17- 23 (2005).

[4] Kannathala N., Choob M.L., et al: Entropies for detection of epilepsy in EEG, Comp Meth Prog in Biomed, 80, 187–194 (2005).

[5] Cromwell D.H., Jones R.H., Kapuniai L.E: Autoregressive representation of infant EEG for the purpose of hypothesis testing and classification, EEG and Clin. Neurophysiology, 43, 317–324 (1977).

[6] Jansen B.H., Boume J.R., Ward J.W: Autoregressive estimation of short segment spectra for computerized EEG analysis, IEEE Transaction on BME, 28(9), 630–637 (1981).

[7] Wright J., Kydd R., Sergejew A: Autoregression models of EEG, Biol. Cybern. 62, 201–210 (1990).

[8] Wang T.W., Guohui W., Huanqing F: Autoregressive model order property for sleep EEG, J of Biomedical Eng, 21(3), 394-6 (2004).

[9] Mohammadi G., Shoushtari P. et al: Person identification by using AR model for EEG signals, Proceed of World Academy of Science, Eng and Techn, 11, 281–285 (2006).

[10] Proakis J.G., Manolakis D.G: Digital Signal Processing, Prentice Hall, 3rd Ed., Chp.12, p.925–956 (1996).

[11] Manolakis D.G, Ingle V.K., Kogon S.M: Statistical and Adaptive Signal Processing, Artech House Press, Ch.1, Ch.9 (2005).

[12] Andrzejak R.G., Lehnertz K., Mormann F., et al: Indications of nonlinear deterministic and finite dimensional structures in time seires of brain electrical activity: Dependence on recording region and brain state. Physical Review E, 64, 061907 (2001).

[13] Aydın S: Comparison of power spectrum predictors in computing coherence functions for intracortical EEG signals, Annals of Biomedical Engineering, 37(1), 192–200 (2009).

[14] Schneider T., Neumaier A: Algorithm 808: ARfit-A Matlab package for the estimation of parameters and eigenmodes of multivariate autoregressive models. ACM Trans Math Softw, 27, 58–65 (2001).

[15] Kay S.M: Modern Spectral Estimation, Ch.7, Prentice-Hall, Englewood-Cliffs N.J.(1988).

[16] Neumaier A., Schneider T.: Estimation of parameters and eigenmodes of multivariate autoregressive models, ACM Trans on Math Softw, 27(1), 27–57 (2001).

[17] Chatfield C.: The Analysis of Time Series, Chapman and Hall (1996).

[18] Lütkepohl H.: Introduction to Multiple Time Series Analysis, 2nd ed, Cp.3, Springer-Verlag, Germany (1993).

[19] Hansen P.C.: Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion. SIAM, Philadelphia (1997).

[20] Aydın S.: Tikhonov regularized solutions for improvement of signal-to-noise-ratio in case of auditory-evoked potentials, Medical and Biological Engineering and Computing, 46(10), 1051–1056 (2008).

[21] Kashyap R.L: Inconsistency of the AIC rule for estimating the order of autoregressive Models, IEEE Trans on Auto. Cont., AC-25, 996–998 (1980).

[22] Schwarz G., Estimation of the dimension of a model, Annals of Statistics, 6, 461–464 (1978).

[23] Björck, A.: Numerical Methods for Least Squares Problems, SIAM, Philadelphia (1996).