Proceedings of the 28th IEEE
EMBS Annual International Conference
New York City, USA, Aug 30-Sept 3, 2006

Detecting Determinism in EEG Signals using Principal Component Analysis and Surrogate Data Testing

Amir H. Meghdadi, Reza Fazel-Rezai, *Senior Member, IEEE*, and Yahya Aghakhani

***Abstract*—A novel method is proposed here to determine whether a time series is deterministic even in the presence of noise. The method is the extension of an existing method based on smoothness analysis of the signal in state space with surrogate data testing. While classical measures fail to detect determinism when the time series is corrupted by noise, the proposed method can clearly distinguish between pure stochastic and originally deterministic but noisy time series.**

A set of measures is defined here named partial smoothness indexes corresponding to principal components of the time series in state space. It is shown that when the time series is not pure stochastic, at least one of the indexes reflects determinism.

The method is first successfully tested through simulation on a chaotic Lorenz time series contaminated with noise and then applied on EEG signals. Testing results on both our experimental recorded EEG signals and a benchmark EEG database verifies this hypothesis that EEG signals are deterministic in nature while contain some stochastic components as well.

I. Introduction

Electroencephalography (EEG) signal is a good source of information about the brain function. Characterizing and analysis of EEG signals is hence very important. Nonlinear dynamical analysis is an important approach for this problem, which uses the formalism of deterministic chaos for characterization and analysis of time series [1][2]. Recently, there has been much attention in implementation of recent progresses in this area in order to investigate neurophysiological signals such as EEG and MEG [3][4]. One important question in nonlinear analysis approach is the issue of determinism in time series. There are many classical methods for detecting nonlinearity and determinism in time series, which implement nonlinear chaotic parameters such as correlation dimension and Lyapunov exponents [1][2]. These methods have been empowered and modified by the advent of "surrogate data" testing [5] which can accurately detect real deterministic nature of time series [10][12][14][1]. This and many other approaches have been applied on EEG signals in order to answer this question whether the source of complex EEG signal is a nonlinear chaos (deterministic) or stochastic process.[13][14][3][6][7].

Manuscript received April 3, 2006. This work was supported in part by the University of Manitoba Research Grant Program (URGP).

A. H. Meghdadi, is a PhD student at the University of Manitoba, Winnipeg, MB R3T 5V6 Canada (e-mail: meghdadi@ ee.umanitoba.ca).

R. Fazel-Rezai is with the Department of Electrical and Computer Engineering, University of Manitoba, Winnipeg, MB R3T 5V6 Canada (corresponding author: phone: 1-204-474-9490; fax: 1-204-261-4639; email: fazel@ee.umanitoba.ca).

Y. Aghakhani is with the Department of Internal Medicine, Health Sciences Centre, University of Manitoba, Winnipeg, MB R3A 1R9 Canada (email: yaghakhani@exchange.hsc.mb.ca).

Although the issue of being chaotic or stochastic is still an open question [3], many of the reported studies in the literature concluded a non-deterministic nature for the EEG signals [14][13]. Non-linearity and determinism however, are reported to be present in special cases or particular brain states such as epileptic activities [8][9][13].

Nonetheless, the validity of the above methods is questionable when the time series is corrupted by noise. In [14], it is clearly shown that adding 10% white noise to a deterministic time series can results misleading false negative results and it is recommended that: "method should be used with much caution to examine noisy real systems"[14]. Consequently, it is not clear whether the EEG signals are really stochastic or highly noisy deterministic. The proposed method in this paper overcomes the above problem in detecting determinism in noisy time series and tries to answer the question about the real nature of EEG signal. The new method is an extension of the method used in [14] based on smoothness index. It is shown that our method is now able to detect any deterministic component in a highly corrupted and noisy time series.

The rest of the paper is as follows. In section II, the method of testing determinism and its extension is discussed. In section III, the method is tested by simulation on a standard chaotic time series and section IV demonstrates the positive results of detecting determinism both on experimental recorded EEG and a benchmark EEG signal. Finally, Section V presents a summary and conclusion.

II. A Method of Determinism for Time Series and Its Extension using Principal Component Analysis

A. Determinism based on smoothness and surrogate data

A method for determinism in short time series base on "surrogate testing" [5][14] was used and implemented in MATLAB©. The method is based on the assumption that smoothness of the state space trajectory implies determinism in time series [14]. Moreover a surrogate data analysis [5] is used to compare the smoothness of the signal and its surrogate. The method is briefly described here. For further information refer to [14]. If $x(t)$ is a time series, it is first transformed into a multidimensional state space as follows:

$$\vec{X}(t) = [x(t),\, x(t+T),\, \cdots,\, x(t+(d-1)T)] \quad (1)$$

while T is the time delay between the samples of the time series and d is the embedding dimension. The tangent vector of the trajectory in the state space is defined as:

$$\vec{Y}(t) = \vec{X}(t+1) - \vec{X}(t) \quad (2)$$

1-4244-0033-3/06/$20.00 ©2006 IEEE.

The angles between successive tangent vectors are computed as follows where R is the cosine function of the angles.

$$R(t) = \frac{\vec{Y}(t+1) \cdot \vec{Y}(t)}{\left|\vec{Y}(t+1)\right| \cdot \left|\vec{Y}(t)\right|} \tag{3}$$

A second order difference plot (SODP) can be also defined as a graphical representation of the rate of variability by plotting $R_{n+2} - R_{n+1}$ verses $R_{n+1} - R_n$. Finally, the central tendency measure (CTM) given by the following formula indicates a quantitative estimate for the variability in the SODP.

$$CTM = \frac{\sum_{n=1}^{N-2} \sqrt{(R_{n+2} - R_{n+1})^2 + (R_{n+1} - R_n)^2}}{N-2} \tag{4}$$

The index of smoothness (S) of the trajectory is then defined as the ratio of the CTM of the angle variations corresponding to the original time series to the average CTM value of the angle variations corresponding to its surrogate data.

$$S = \frac{CTM_O}{CTM_S} \tag{5}$$

The surrogate data is a random time series with the same statistical components and is used here to distinguish between real deterministic and stochastic time series. Time series are classified into deterministic and stochastic based on the value of smoothness index S. Deterministic time series are shown to have lower smoothness indexes of about 0.3 and smaller while stochastic time series have index values of more than about 0.7 [14]. Here, we call this margin (between 0.3 and 0.7) "margin of smoothness". In order for a time series to be classified accurately as deterministic or stochastic, the smoothness index should be well above or below this margin.

The ability of the above method in detecting determinism of the time series has been shown in the literature [10][14]. However, the main disadvantage of the method is its poor robustness to added noise. In the presence of noise, the smoothness index increases very rapidly with increasing noise level. Consequently, the time series will be falsely classified as stochastic and the real deterministic nature of the time series will not be detected.

In the next part, we proposed an extension to this method. The extended method provides a set of smoothness indexes called "partial smoothness index". It is shown here that even when the noise level is very high, at least one of the partial smoothness indexes is below the smoothness margin and hence reveals a source of determinism in the time series.

B. Partial Smoothness Index based on Principle Component Analysis

Incorporating principal component analysis (and SVD in particular) in time delay embedding method has been discussed in the literature. It has been also shown that singular values can reflect valuable information about the time series and its determinism [12]. In this paper, a novel concept of Partial Smooth Index (S_p) is proposed based on principle component analysis (PCA) of the signal in the state space. The objective is to distinguish between a deterministic signal that may be corrupted by noise and a real stochastic signal that is random in nature. The procedure for obtaining the partial smoothness indexes is described as follows.

The signal is first transformed into state space similar to the conventional method but using a larger embedding dimension named d_p,

$$\vec{X}(t) = [x(t), x(t+T), \cdots, x(t+(d_p-1)T)] \tag{6}$$

Matrix representation of the reconstructed attractor in state space is in the following form:

$$XD = \begin{bmatrix} x(1) & x(2) & \cdots & x(N-d_p+1) \\ x(2) & x(3) & \cdots & x(N-d_p+2) \\ & & & \vdots \\ x(d_p) & x(d_{p+1}) & & x(N) \end{bmatrix} = \left[\vec{X}(1)^T \quad \vec{X}(2)^T \quad \cdots \quad \vec{X}(N-d_p+1)^T \right] \tag{7}$$

A principal component analysis (PCA) is then applied on the input vectors in XD where input data is transformed such that the elements of the input vectors will be uncorrelated. The size of the input vectors may be reduced by retaining only those components that contribute more than a specified fraction of the total variation in the data set. Transformed vectors may be in lower dimension and are used to build the principle components of the signal in the original one-dimensional space. The smoothness index as defined in equation (5) is then computed for each of the resulting signals and named "partial smoothness index" S_p corresponding to the p^{th} principle component. If the time series signal has deterministic components, it is expected that some partial smoothness indexes reflect this smoothness.

III. SIMULATION RESULTS

Superiority of the proposed method in detecting original deterministic nature of noisy signals is shown here through simulation. A nonlinear chaotic Lorenz time series is considered and has been corrupted by noise. The method is applied with the parameters shown in Table 1.

TABLE 1
PARAMETERS OF THE METHOD

Parameter	Value
Embedding Dimension	$d=7$
Time Delay	$T=15$
Embedding Dimension (partial)	$d_p=110$

Fig. 1 shows the conventional smoothness index of the time series versus the amount of additive noise in solid line. Added noise is normal noise with zero mean and the standard deviation compared to the standard deviation of the time series. It is clear that even a small amount of noise with relative standard deviation of 0.01, can have a significant misleading increasing effect in the value of smoothness index. The minimum value among the partial smoothness indexes is also plotted at the same time with dashed line. It can be shown that this index consistently remains below the smoothness margin. This means that the time series has major deterministic components. Only when the standard deviation of the added noise is about the same level of signal, the smoothness index starts to increase.

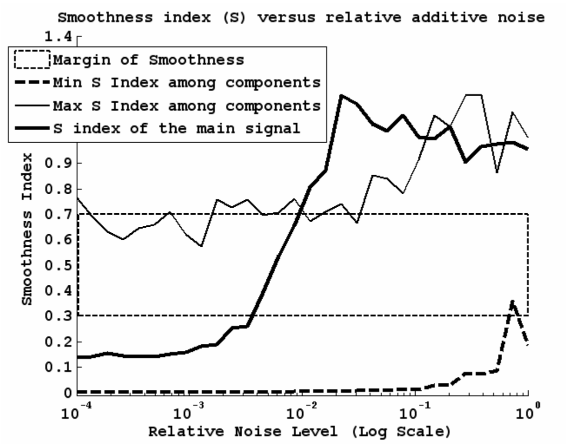

Fig. 1. Maximum and minimum partial smoothness index for different noise levels compared to conventional smoothness index

The partial smoothness indexes can be computed for all the principle components. Fig. 2 demonstrates the values of the first 8 major partial smoothness indexes corresponding to the first 8 major components. It has been computed for the worst case when the signal is added with a noise in the same level as the signal. It can be shown that there is one partial smoothness index corresponding to the fourth component, with a value well lower than smoothness margin. This implies that the signal is not pure noise and there is a deterministic source in the signal. In order to validate the above conclusion, partial smoothness indexes are also computed for pure noise in a random time series. Fig. 3 demonstrates the values of partial smoothness indexes with all the values above the smoothness margin. This shows that this signal is pure stochastic.

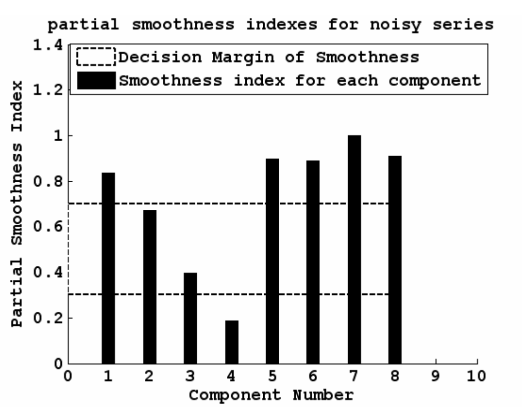

Fig. 2. Values of the partial smoothness indexes for different components.

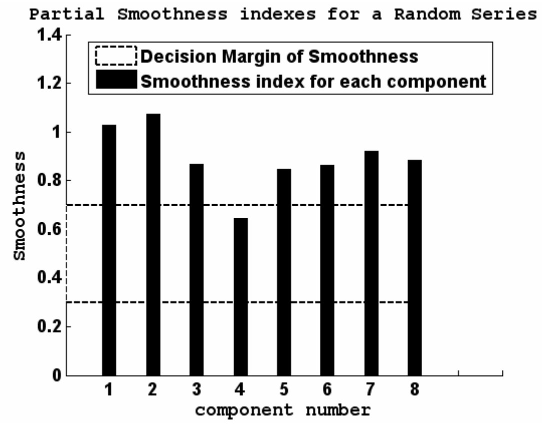

Fig. 3. Partial smoothness indexes of a uniform random time series

IV. Experimental Results on EEG signals

A. EEG data acquisition

Scalp EEGs were recorded using a digital EEG amplifier (EEG32 XLTEK, Canada). Monopolar electrodes in a standard 10-20 system are used. Subject was a 31 years old male in relaxing state with eyes open during a two minutes experiment. Five EEG channels (C3, C4, Cz, Fz, Pz) with a sampling rate of 499.9 Hz are recorded. Time series of the recorded signals are imported into MATLAB® and band passed filtered between 1 and 70 Hz. A notch filter is also used to remove 60Hz power line noise. Small artifact-free segments (2 second ~ 1000 samples) of the EEG signals were selected and tested for determinism using the above methods for all the channels.

Fig. 4 shows the partial smoothness indexes for channel Cz. The smoothness indexes were computed for the first 12 major components of the state space matrix. It can be shown that partial smoothness indexes corresponding to the components 1,5,6,7 have small values implying determinism of the EEG signal in essence.

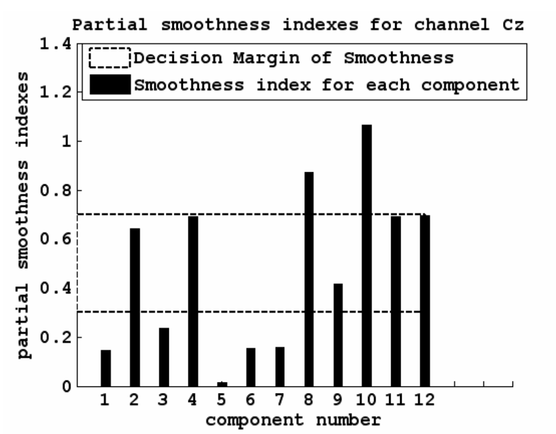

Fig. 4. Partial smoothness indexes relating to the first 12 major components computed for channel Cz.

The above analysis was performed for different channels of the recorded EEG signals. The results are summarized in Table 2. In the first row, all the conventional smoothness indexes are shown with large values and hence do not suggest any determinism in any channel.

In the next row, the minimum partial smoothness indexes among all the first 12 components are shown to be very small. Consequently, the above algorithm implies deterministic nature for all the channels.

TABLE 2
GENERAL AND PARTIAL SMOOTHNESS INDEX FOR DIFFERENT CHANNELS OF THE RECORDED EEG

Channels / S index	C_3	C_4	C_z	F_z	P_z
S Determinism	*1.26* *No*	*1.05* *No*	*1.33* *No*	*1.27* *No*	*1.04* *No*
$Min_p(S_p)$ Determinism	*0.05* *Yes*	*0.07* *Yes*	*0.05* *Yes*	*0.17* *Yes*	*0.04* *Yes*

B. Experiments on other EEG data sets

This method also applied on sample EEG segments available from Bonn University [9] with the same conclusion as in the recorded experimental data. This dataset is chosen here as a benchmark to verify the result of the proposed algorithms with a standard EEG data which has been used by many other authors in the literature.

The data set consists of EEG signals with a sampling rate of 173.61 Hz and filtered by the acquisition system in the range of 0.5 Hz and 85 Hz. Table 3 demonstrates the general and minimum partial smoothness indexes for 5 different EEG segments. The general smoothness indexes are shown in the first row. They are large enough and hence fail to detect any determinism. In the second row, it is shown that all the signals have at least one very small partial smoothness index, which implies determinism of the EEG signal.

TABLE 3
GENERAL AND PARTIAL SMOOTHNESS INDEX FOR DIFFERENT CHANNELS OF THE AVAILLABLE EEG DATASET (UNIVERSITY OF BONN)

Segment / S index	F001	F002	F003	F004	F005
S Determinism	*1.10* *No*	*1.78* *No*	*1.14* *No*	*0.92* *No*	*1.15* *No*
$Min_p(S_p)$ Determinism	*0.021* *Yes*	*0.003* *Yes*	*0.010* *Yes*	*0.012* *Yes*	*0.027* *Yes*

V. CONCLUSION

The EEG signals are usually considered as stochastic by many of the existing methods in detecting nonlinearity and determinism. However, using the proposed method, we showed that deterministic components in EEG signals can be detected. This suggests a deterministic nature for EEG signal which is corrupted by stochastic noise. The existence of this deterministic component is extremely important in nearly all the applications of electroencephalography from clinical diagnosis applications to brain computer interface (BCI) design.

ACKNOWLEDGMENT

The authors would like to thank the technical support of the Electrophysiology Laboratory at the Health Sciences Centre, Winnipeg, Manitoba, Canada. The financial support from University of Manitoba Research Program (URGP) is also acknowledged.

REFERENCES

[1] "Nonlinear biomedical signal processing" edited by Metin Akay, IEEE Press, New York, 2000.

[2] Garnett P. Williams, "Chaos theory tamed", Joseph Henry Press, Washington, D.C, 1997.

[3] C.J. Stam, "Nonlinear dynamical analysis of EEG and MEG: Review of an emerging field", *Clinical Neurophysiology*, Vol. 116, Issue 10, Oct 2005, pp. 2266-2301.

[4] E. Pereda, R. Q. Quiroga and J. Bhattacharya, "Nonlinear multivariate analysis of neurophysiological signals", *Progress in Neurobiology*, Volume 77, Issues 1-2, September-October 2005, pp. 1-37.

[5] T. Schreiber and A. Schmitz, "Surrogate time series", *Physica D: Nonlinear Phenomena*, Vol. 142, Issues 3-4, Aug 2000, pp. 346-382.

[6] H. Korn and P.hilippe Faure, "Is there chaos in the brain? II. Experimental evidence and related models", *Comptes Rendus Biologies*, Volume 326, Issue 9, September 2003, pp. 787-840.

[7] S. N. Sarbadhikari and K. Chakrabarty, "Chaos in the brain: a short review alluding to epilepsy, depression, exercise and lateralization", *Medical Engineering & Physics*, Volume 23, Issue 7, September 2001, pp. 447-457.

[8] R. G. Andrzejak, G. Widman, K. Lehnertz, C. Rieke, P. David and C. E. Elger, "The epileptic process as nonlinear deterministic dynamics in a stochastic environment: an evaluation on mesial temporal lobe epilepsy", *Epilepsy Research*, Volume 44, Issues 2-3, May 2001, pp. 129-140.

[9] R. G. Andrzejak, K. Lehnertz, C. Rieke, F. Mormann, P. David, C. E. Elger "Indications of nonlinear deterministic and finite dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state", *Phys. Rev. E*, 64, 061907, 2001.

[10] M. Small, C. K. Tse, "Detecting determinism in time series: the method of surrogate data," *Circuits and Systems I: Fundamental Theory and Applications, IEEE Transactions on*, vol.50, no.5, pp. 663- 672, May 2003.

[11] N. Kannathal, U. Rajendra Acharya, C.M. Lim and P.K. Sadasivan, "Characterization of EEG--A comparative study", *Computer Methods and Programs in Biomedicine*, Volume 80, Issue 1, October 2005, pp. 17-23.

[12] J. Bhattacharya and P. P. Kanjilal, "On the detection of determinism in a time series", *Physica D: Nonlinear Phenomena*, Volume 132, Issues 1-2, 15 July 1999, pp. 100-110.

[13] R. Stepien, "Testing for non-linearity in EEG signal of healthy subjects", *Acta Neurobiol. Exp*, 2002, Vol. 62, pp. 277-281.

[14] J. Jeong, J.C. Gore, B. S. Peterson, "A method for determinism in short time series and its application to stationary EEG", *Biomedical Engineering, IEEE Transactions on*, Volume 49, Issue 11, Nov. 2002 pp. 1374– 1379.