ELSEVIER Mathematical Biosciences 157 (1999) 287–302

Mathematical Biosciences
an international journal

Non-linear analysis of the rhythmic activity in rodent brains [1]

Milkes Yone Alvarenga [a,d], Koichi Sameshima [d,*], Luiz Antonio Baccalá [b], Hyun Mo Yang [c]

[a] *Graduate Program, IMECC, State University of Campinas, Brazil*
[b] *Escola Politécnica, Electronics Engineering Department, University of São Paulo, Brazil*
[c] *IMECC, State University of Campinas, Brazil*
[d] *Discipline of Medical Informatics & Laboratory of Functional Neurosurgery, School of Medicine, University of São Paulo, Av. Dr. Arnaldo 455 CEP 01246-903 São Paulo, SP, Brazil*

Received 31 December 1997; received in revised form 1 July 1998; accepted 9 October 1998

Abstract

This paper discusses the employment of non-parametric non-linear prediction algorithms to investigate non-linear dynamics in the rhythmic brain activity of rats. Three algorithms (Sugihara–May Simplex, *K*-neighbour and Casdagli's) were tested yielding similar prediction results which – when subject to a suitable bootstrap based *t*-tests – revealed that the theta waves recorded in rat brains cannot have their intrinsic non-linearity dismissed at a significance of 0.05.

Keywords: Non–paramagnetic non-linear prediction; Theta waves; Surrogate methods

1. Introduction

Our research group has been engaged in the investigation of the mutual dynamical relations between several cortical and subcortical structures in the central nervous system of rats with focus on rhythmic brain activity, specially

* Corresponding author. E-mail: ksameshi@usp.br
[1] Financial support: PRONEX 41.96.0925.00, CNPq and FAPESP 96/12118-9.

0025-5564/99/$ – see front matter
PII: S0025-5564(98)10087-1

theta waves, that commonly emerge during oneiric episodes – believed by some as the equivalent of human dreams. By far, the most widely accepted approaches rely on linear methods [1,2] which are assumed true at least on a first approximation basis; this popularity justified by the clearer statistical issues involved in their estimation. The fact, however, that the neuron, the very unit of brain processing itself, has widely recognized non-linear behaviour leads to natural concerns about approaches involving linear approximations in the context of neural signal analysis.

As such, this paper describes our inquiry into objective methods for quantifying the departure from linearity in our observed signals. There are several ways to infer the degree of non-linearity of a given time series: some are based on higher order spectral analysis [3], others employ non-linear approximating functions [4] and yet others look into the qualitative dynamics of the data [5]. A close review of the literature soon convinced us that none of these former methods was completely satisfactory as all of them require fairly long observations for reliable estimation, particularly in noise contaminated contexts, a scenario hardly ever compatible with neuroelectric signals which, apart from added noise, can only be considered stationary over relatively short time spans.

A seemingly viable alternative appeared in Theiler's work [6] who proposed the use of 'bootstrap' as a means of overcoming the limitations due to small samples [7]. The approach's cornerstone is that any measure of dynamical non-linearity computed from an observed signal must be significantly altered whenever the signal data is 'linearized', i.e. when related – also called 'surrogate' – time series are randomly generated on the basis of preserving the original signal's second order statistics (spectrum).

Adequate use of Theiler's approach calls for the definition of some figure of merit capable of discriminating non-linearity. We elected predictability through non-parametric non-linear methods as our figure of merit. This choice was dictated by the method's dispensation of estimating model specific parameters. In addition, non-parametric non-linear prediction methods provide some idea of the dimensionality required when future parametric models come to be considered upon confirmation of evidence for an underlying non-linear dynamics.

There are many non-parametric non-linear prediction methods, so that from the beginning it was somewhat unclear whether any one such method would be preferable to others in some sense. Thus, in part, this paper also seeks to investigate and possibly suggest the most appropriate method for testing non-linearity.

Section 2 contains a description of the type of signal used together with a brief review of the prediction methods compared herein: Sugihara–May Simplex Method [8], *K*-neighbour [9] and Casdagli's method [10]. The section ends with a description of the non-linearity test procedures. Our results are summarized in Section 3 and more fully discussed in Section 4.

2. Material and methods

2.1. The signals

Theta waves in rodents constituted the raw material in our analysis. This type of signal is roughly characterized by an oscillatory pattern with some superposed noise (Fig. 1). Our time series – local field potentials records – sampled at 256 Hz were obtained during desynchronized sleep from visual cortex, thalamus and hippocampal fields through multiple bipolar microelectrodes implanted chronically in rats.

Each N-point time series $\mathbf{x}(k)$, $N = 2000$ or 3000, (see Table 1) was represented by the vector

$$\mathbf{x} = [x(1), \ldots, x(N)]^T = \begin{bmatrix} \mathbf{x}_p \\ \mathbf{x}_t \end{bmatrix}$$

partitioned for analysis into two other vectors of equal length $\mathbf{x}_p$ e $\mathbf{x}_t$ called respectively the *prediction set* and the *test set*.

Table 1 contains a summary of the time series used in this analysis.

2.2. Non-linear non-parametric prediction methods

Non-linear non-parametric methods rest on the notion that the topological features of the dynamics of a given system may be adequately captured by reconstructing a phase-space representation of a related time series under observation [11,12]. Thus, prior to non-linear forecasting, regardless of the method employed, one must construct a data structure – the so-called attractor – denoted $\mathscr{R}_{\tau,m}(\mathbf{x}_p)$ made up of vectors describing phase-space points of the form

$$\mathbf{x}(t) = [x(t), x(t-\tau), x(t-2\tau), \ldots, x(t-(m-1)\tau)]^T, \tag{1}$$

where τ is the so called *embedding lag*, i.e. the delay between points in the $\mathbf{x}(k)$ series and m – the *embedding dimension* [12]. Use of x_p emphasizes the exclusive use of the prediction set in attractor reconstruction.

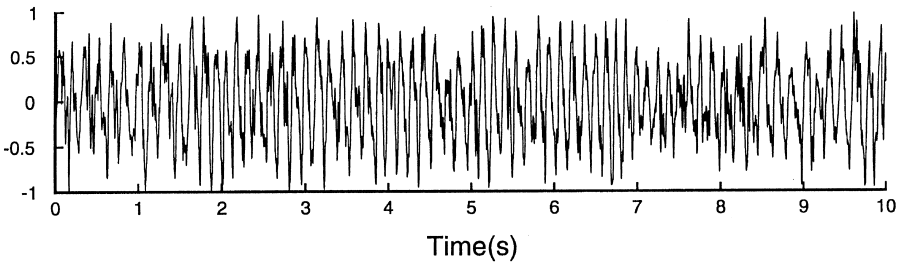

Fig. 1. Example of hippocampal local field potential record used here for non-linearity analysis.

Table 1
Distribution of the optimal parameters used in setting the non-linear non-parametric prediction algorithms

Record #	Brain area **	m_0	τ_0	K_0
1	CA1	8	9	20
2	CA3	8	9	20
3	CA1	10	9	20
4	CA3	10	9	40
5	CA1	8	10	20
6	CA1	11	9	50
7	CA3	6	9	20
8 *	CA1	8	9	20
9 *	CA3	8	9	20
10	CA3	5	10	50
11	CA1	8	10	20
12	CA2	8	9	20
13	CA3	12	10	20
14	SUB	8	9	70
15	CA3	8	10	20
16	SUB	8	9	70
17	CA3	11	9	20
18	CA3	8	9	50
19	VPL	8	9	20
20	A18	10	10	40

* $N = 3000$ for the Records 8 and 9, and $N = 2000$ for all other Records were used.
** CA1, CA2, CA3 and subiculum (SUB) are hippocampal regions; VPL–ventral posterolateral nucleus of thalamus; and A18 corresponds to visual cortex.

Using this same data structure, it is possible to produce an one-step-ahead forecast

$$\hat{x}(h) = \mathscr{P}[\tilde{\mathbf{x}}_{\mathrm{t}}(h), \mathscr{R}_{\mathrm{r,m}}(\mathbf{x}_{\mathrm{p}})],$$

where $\mathscr{P}$ is a suitable algorithm (functional) acting on

$$\tilde{\mathbf{x}}_{\mathrm{t}}(h) = [x_{\mathrm{t}}(h-1), x_{\mathrm{t}}(h-1-\tau), x_{\mathrm{t}}(h-1-2\tau), \ldots, x_{\mathrm{t}}(h-1-(m-1)\tau)]^T.$$

Essentially different prediction algorithms are represented by distinct $\mathscr{P}$ functionals. The first one such method is the Simplex Algorithm [8] which used $\mathscr{R}_{\mathrm{r,m}}(\mathbf{x}_{\mathrm{p}})$ information directly. The other two methods considered in this paper: K-neighbour Algorithm and Casdagli's Algorithm require an additional parameter K to specify the number of closest neighbouring phase-space points used in their forecast.

To apply these procedures one must first determine suitable values for τ and m (and K) prior to effective prediction which essentially amounts to an optimization problem involving each series under analysis and an appropriate 'loss' function reflecting predictability.

An estimate of the cross-correlation function between the predicted function and the test set (also known as Kravtsov's degree of predictability [13]),

$$\rho(\mathbf{x}_t, \hat{\mathbf{x}})(k), \tag{2}$$

proved to be a very useful 'loss' function. Here $\hat{\mathbf{x}}$ denotes the vector $[\hat{x}(1), \ldots, \hat{x}(N/2)]^T$.

In order to choose the best embedding dimension, a discrete one-dimensional parameter search on increasing m was performed using the Simplex Method for τ chosen – following Theiler [14] – as approximately the time of the first zero crossing in the estimated autocorrelation of $\mathbf{x}_p$. For the previously determined optimum embedding dimension m_0, the optimum value of K was found for each of the other two methods by a similar one-dimensional parameter search as described in [15]. Both optimizations were concerned with determining the first occurrence of local maxima for the average of (2) for $k = 1, 2, \ldots, 64$. This prediction lag span was chosen because $k = 64$ corresponds to at least two cycles of oscillatory activity.

Remark 1. *The choice of maximizing predictability for* $k = 1$ *as in* [15] *was dismissed because it systematically lead to* $m_0 = 3$ *which failed to yield the best overall performance for longer prediction lags* (*see, e.g., Fig. 2*).

2.3. The non-linearity tests

Since for optimum τ_0, m_0 (and K_0) Eq. (2) is a good indicator of a non-linear method's prediction accuracy, its value was used in testing whether the signal is significantly non-linear when compared to the prediction of $i = 1, \ldots, N_s$ 'surrogate' time series,

$$\mathbf{x}_s^{(i)} = \mathscr{S}(\mathbf{x}),$$

built randomly from the original time series, each surrogate preserving the second order statistics of the original series.

The rationale of this constraint is that linear models of a time series, whether parametric or not, are completely defined by the series second order statistics. In practice, $\mathscr{S}$ was implemented as an operator consisting of the computation of the Discrete Time Fourier Transform $X(f) = \mathscr{F}(\mathbf{x})$, followed by phase randomization performed remembering that $X(f)$ is in general a complex number

$$X(f) = A(f)[\cos\varphi(f) + \mathrm{i}\,\sin\varphi(f)],$$

where $A(f) = |X(f)|$ is the sole provider of second order statistics information. Surrogates differ only through having new phases $\varphi'(f)$ chosen randomly from an uniform distribution under the constraint that $\varphi'(f) = \varphi'(-f)$ so that the resulting spectrum

$$X'(f) = A(f)[\cos\varphi'(f) + \mathrm{i}\,\sin\,\varphi'(f)] \tag{3}$$

leads to each surrogate $x_s^{(i)}$ time series upon anti-transformation $\mathscr{F}^{-1}X'(f)$ for each randomly chosen $\varphi'(f)$, the generation of each surrogate requiring $N/2$ independent uniformly distributed random numbers between $[0,\ 2\pi]$ [16].

In constructing surrogates, one often recommends data rescaling prior to phase randomization [16]. This step 'gaussianizes' the data's first order statistics to allow distinguishing static and memoryless non-linearities from dynamic ones. In this study, both unscaled and scaled surrogates were used and compared.

Once generated, each surrogate was partitioned in two $\mathbf{x}_s^{(i)} = [\mathbf{x}_p^{(i)}\ \mathbf{x}_t^{(i)}]^T$ and underwent the same non-linear prediction procedure:

$$\hat{x}^{(i)}(h) = \mathscr{P}[\tilde{\mathbf{x}}_t^{(i)}, \mathscr{R}_{\tau_0,m_0}(\mathbf{x}_p^{(i)})] \tag{4}$$

with the associate figure of merit $\rho(\mathbf{x}_t^{(i)}, \hat{\mathbf{x}}^{(i)})\ (k)$ to whose value we associated Fisher's classical transformation of the correlation coefficient [17]:

$$z(\mathbf{x}_t^{(i)}, \hat{\mathbf{x}}^{(i)})\ (k) = \frac{1}{2}\ln\left[\frac{1+\rho(\mathbf{x}_t^{(i)}, \hat{\mathbf{x}}^{(i)})\ (k)}{1-\rho(\mathbf{x}_t^{(i)}, \hat{\mathbf{x}}^{(i)})\ (k)}\right] \tag{5}$$

that lead to the following sample distribution moments

$$\mu_{N_s}(k) = \frac{1}{N_s}\sum_{i=1}^{N_s} z(\mathbf{x}_t^{(i)}, \hat{\mathbf{x}}^{(i)})\ (k)$$

and

$$S_{N_s}^2(k) = \frac{1}{N_s}\sum_{i=1}^{N_s}[z(\mathbf{x}_t^{(i)}, \hat{\mathbf{x}}^{(i)})(k) - \mu_{N_s}(k)]^2$$

against which to test

$$z(k) = \frac{1}{2}\ln\left[\frac{1+\rho(k)}{1-\rho(k)}\right]$$

for the original series through an one sided t-test

$$t_{N_s-1}(k) = \frac{z(k) - \mu_{N_s}(k)}{\frac{s_{N_s}(k)}{\sqrt{N_s}}},$$

where

$$\rho(k) := \rho(\mathbf{x}_t, \hat{\mathbf{x}}_h)(k) = \frac{\sum_i [x_t(i) - \mu_t][\hat{x}(i+k) - \hat{\mu}_h]}{\sqrt{\sum_i [x_t(i) - \mu_t]^2 \sum_i [\hat{x}(i) - \hat{\mu}_h]^2}}$$

in which μ_t and $\hat{\mu}$ are respectively the means of $\mathbf{x}_t$ and $\hat{\mathbf{x}}$.

This t-test to be considered meaningful in rejecting the hypothesis of non-linearity at the α level via

$$t_{N_s-1}(k) > t_{N_s-1,\alpha}$$

requires that the population distribution of the surrogate $z(\mathbf{x}_t^{(i)}, \hat{\mathbf{x}}^{(i)})(k)$ to be approximately normal. In our approach, normality of $z(\mathbf{x}_t^{(i)}, \hat{\mathbf{x}}^{(i)})(k)$ was tested via the Kolmogorov–Smirnov (KS) procedure at a significance of 0.05 for each lag k with $N_s = 100$.

For the sake of comparison we also computed Theiler's $\Delta S(k)$ 'significance' measure [6]:

$$S(k) = \frac{|\mu_H - \mu_D|}{\sigma_H}, \tag{6}$$

where, in our case, $\mu_H = \mu_{N_s}(k), \mu_D = z(k)$ and $\sigma_H = s_{N_s}(k)$ followed by the accompanying error bar:

$$\Delta S(k) = \sqrt{\frac{((1 + 2S^2(k))}{N_s}} \tag{7}$$

where (7) is deemed significant for $S > 10$ and requires no presumption as to the underlying distribution of $z(\mathbf{x}_t^{(i)}, \hat{\mathbf{x}}^{(i)})(k)$.

Remark 2. *Use of the same value τ_0 as for the original series in (4) was dictated by $\mathscr{S}'$s lack of effect on the auto-correlation properties that were used in estimating τ for the original time series. Also the embedding dimension was borrowed from its estimate from the original series as the method's rationale of bootstrapping lies in looking for the attractor structure disruptions caused by phase randomization in (3).*

Remark 3. *It may be interesting to note that we also performed an analogous t-test for the untransformed $\rho(\mathbf{x}_t^{(i)}, \hat{\mathbf{x}}^{(i)})(k)$ variable obtaining results similar to the ones described in the next section. Fisher's transformation did not improve much on the already approximately normal surrogate ρ variables.*

3. Results

Following the optimization procedure briefly exemplified in Fig. 2, we obtained m_0, τ_0 and K_0 summarized in Table 1 whose modes were given by $m_0 = 8, \tau_0 = 9$ and $K_0 = 20$. In all methods, similar performance was obtained (Fig. 3), with Casdagli's method standing as the worst performer for our data. Since the K-neighbour algorithm consistently achieved slightly better prediction, specially for larger k, only this method's results are presented in the forthcoming illustrations.

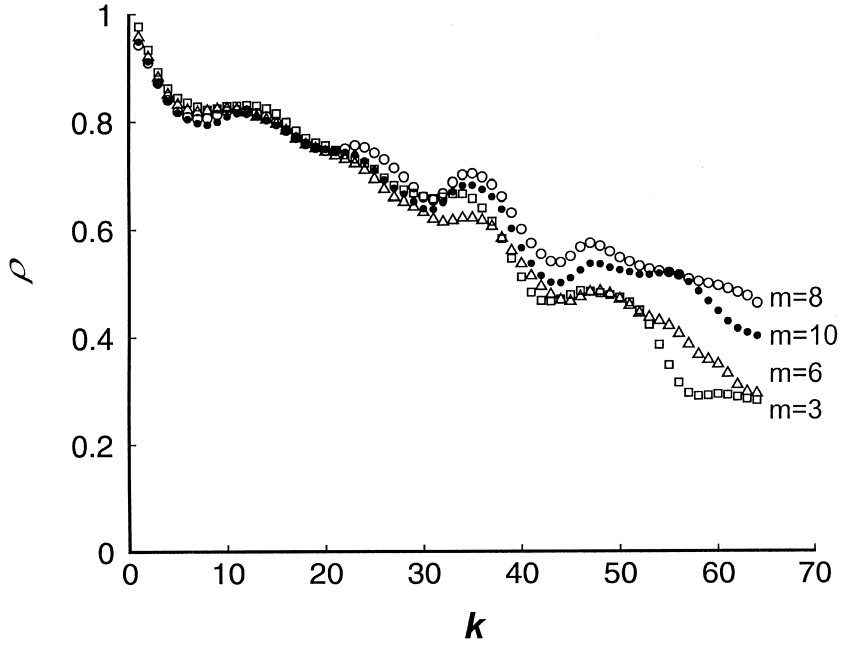

Fig. 2. Results of the determination of the most appropriate embedding dimension m_0 using the Simplex Method. The overall most appropriate dimension $m_0 = 8(\circ)$ was obtained. Other prediction results are also displayed for $m = 3(\square)$, $m = 6(\triangle)$ and $m = 10(\bullet)$.

In none of the Kolmogorov–Smirnov tests, illustrated in one example in Fig. 4(A) and (B) (respectively for unscaled and scaled surrogates), was it possible to reject normality for the $z(\mathbf{x}_t^{(i)}, \hat{\mathbf{x}}^{(i)})(k)$ distributions, lending credence to the corresponding t-tests, illustrated graphically in Fig. 5(A) and (B), with help of a confidence interval

$$\mu_{N_s}(k) \pm t_{N_s-1,\alpha} \frac{S_{N_s}(k)}{\sqrt{N_s}} \tag{8}$$

built around the means respectively of unscaled and scaled surrogate distributions, both with similar behaviour. In all of the time series analyzed, only one series presented evidence for partial non-linearity hypothesis rejection (Fig. 6(A) and (B)) involving large prediction lags ($k > 28$).

A comparison through Theiler's method furnished S values around 2 that also help reject the linear hypothesis with the proviso of a normal surrogate distribution for most series as in Fig. 7(A) (using the scaled surrogates tested previously in Fig. 5(B)). Much smaller values for S (Fig. 7(B)) were attained for the series that had its non-linearity rejected in Fig. 6 by the t-test.

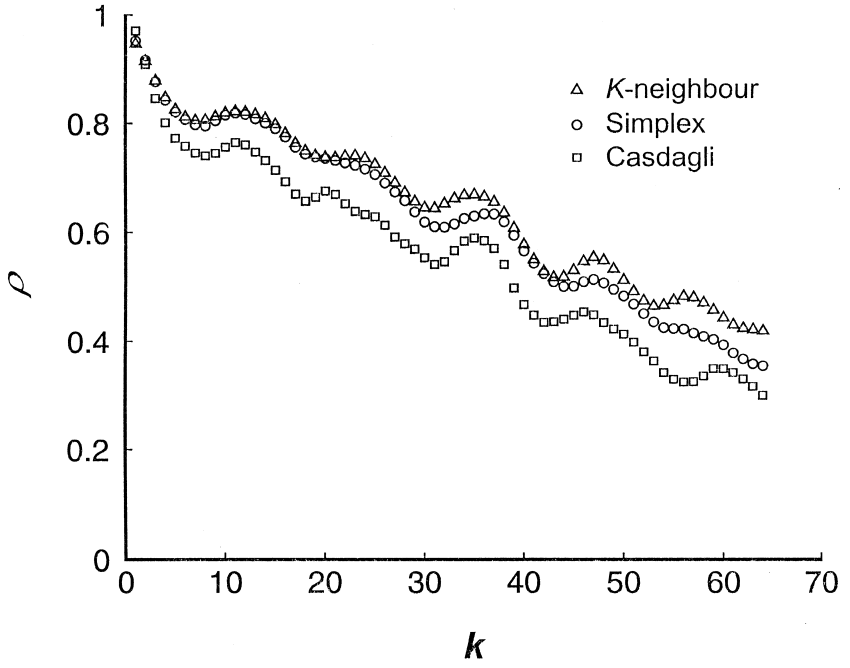

Fig. 3. Displays a comparison between the prediction of the three methods of non-parametric non-linear prediction considered in this paper exemplified through Record 1 in Table 1 ($m_0 = 8$, $\tau_0 = 9$ and $K_0 = 20$). The best performance was attained by the K-neighbour's method ($\triangle$) whereas the worst overall performer was Casdagli's method ($\square$) with the Simplex Method in between ($\circ$).

4. Discussion

With regard to our aim of gauging possible differences among prediction methods, our results (Fig. 3) indicate that the methods addressed here are essentially equivalent in so far as their discriminating ability for detecting the presence of non-linearity. Obviously, thanks to the lower computational complexity ($O(N^2)$ compared to more than $O(N^3)$ for the Simplex Method and $O(N^4)$ for Casdagli's method [2]), the K-neighbour should be the method of choice, moreover so because, it gave slightly superior performance. Interestingly, the most elaborate method (Casdagli's) provided the least reliable predictions in the range $1 \leqslant k \leqslant 64$ (see Fig. 8 for a typical example). This can be

[2] These computational complexity figures refer specifically to our implementations of the methods in MATLAB (MathWorks, Inc, USA).

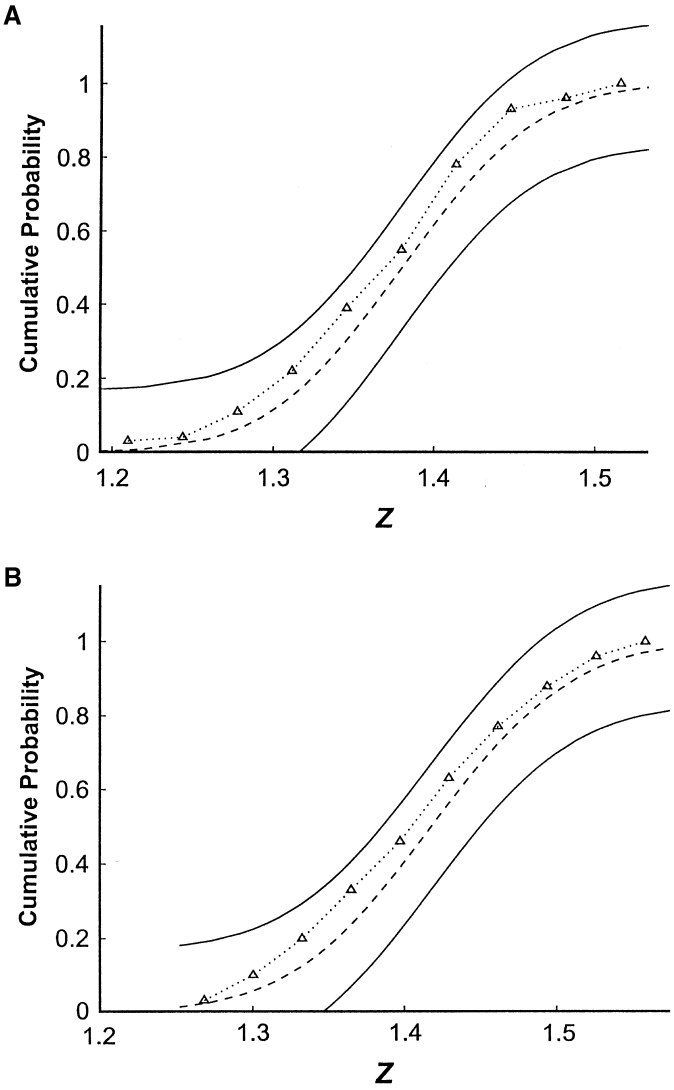

Fig. 4. Depicts the typical results of the Kolmogorov–Smirnov (KS) test for the $z(\mathbf{x}_t^{(i)}, \hat{\mathbf{x}}^{(i)})(k = 1)$ exemplified for Record 1 in Table 1 ($m_0 = 8$, $\tau_0 = 9$ and $K_0 = 20$), where the hypothesis of normality cannot be rejected at a significance of 0.05. The KS test for normality could not be rejected for any of the Records considered. KS tests (not shown) were also applied to $\rho(\mathbf{x}_t^{(i)}, \hat{\mathbf{x}}^{(i)})(k)$ leading also to the inability to reject their normality at the same level.

understood thanks to the method's design to be optimal for $k = 1$, and to the very poor condition number (on the order of 1500) of the matrices involved in its computation.

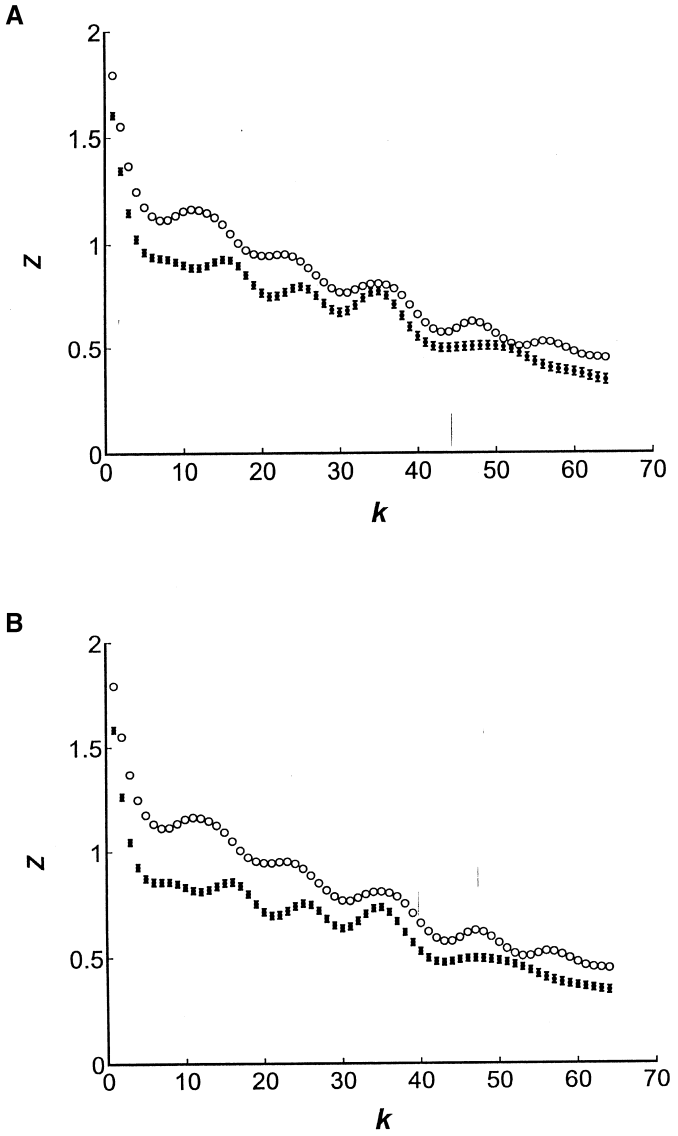

Fig. 5. Plots the series predictability criterion, $z(k)$, ($\circ$) after Fisher's transformation (5) against the mean of $N_s = 100$ for both unscaled (A) and scaled (B) surrogates ($\bullet$). Also shown are confidence intervals in Eq. (8) for the proposed t-test ($\alpha = 0.05$) which reject the linearity hypothesis for all prediction lag k values in Record 2 in Table 1 ($m_0 = 8$, $\tau_0 = 9$ and $K_0 = 20$). The K-neighbour Algorithm was used for all predictions.

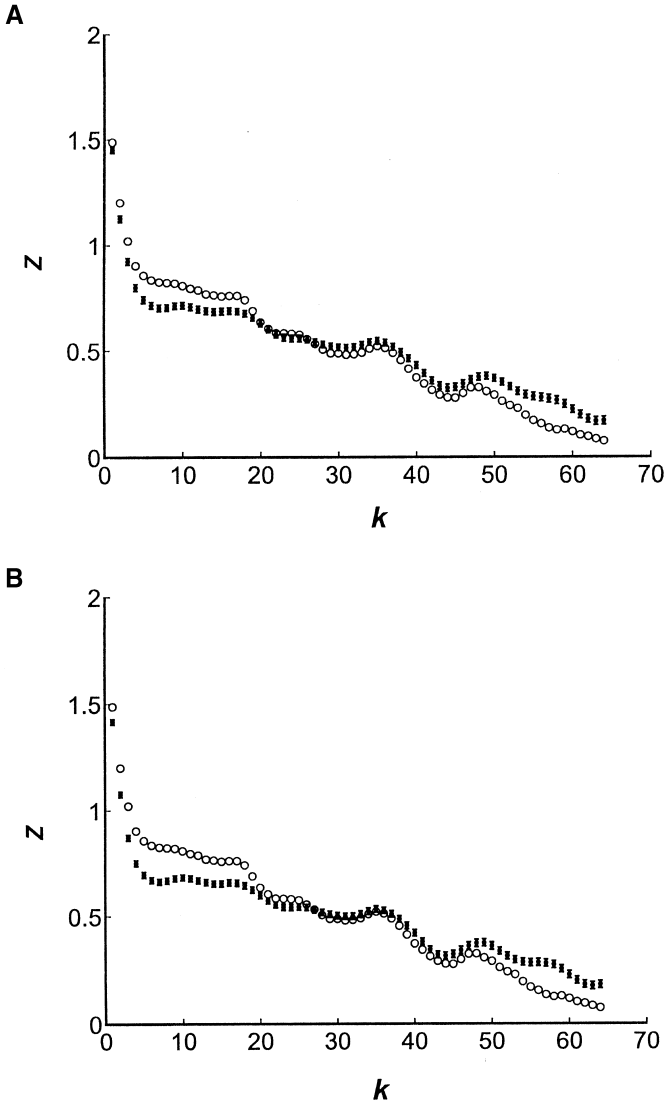

Fig. 6. Plots the series predictability criterion, $z(k)$, ($\circ$) after Fisher's transformation (5) against the mean of $N_s = 100$ for both unscaled (A) and scaled (B) surrogates ($\bullet$). Also shown are confidence intervals in equation (8) for the proposed t-test ($\alpha = 0.05$) for Record 16 in Table 1 ($m_0 = 8$, $\tau_0 = 9$ and $K_0 = 70$). In this case linearity is only partially rejected for prediction lags k below 28. The K-neighbour Algorithm was used for all predictions. In all 20 records studied, this was the only case of inability to reject non-linearity.

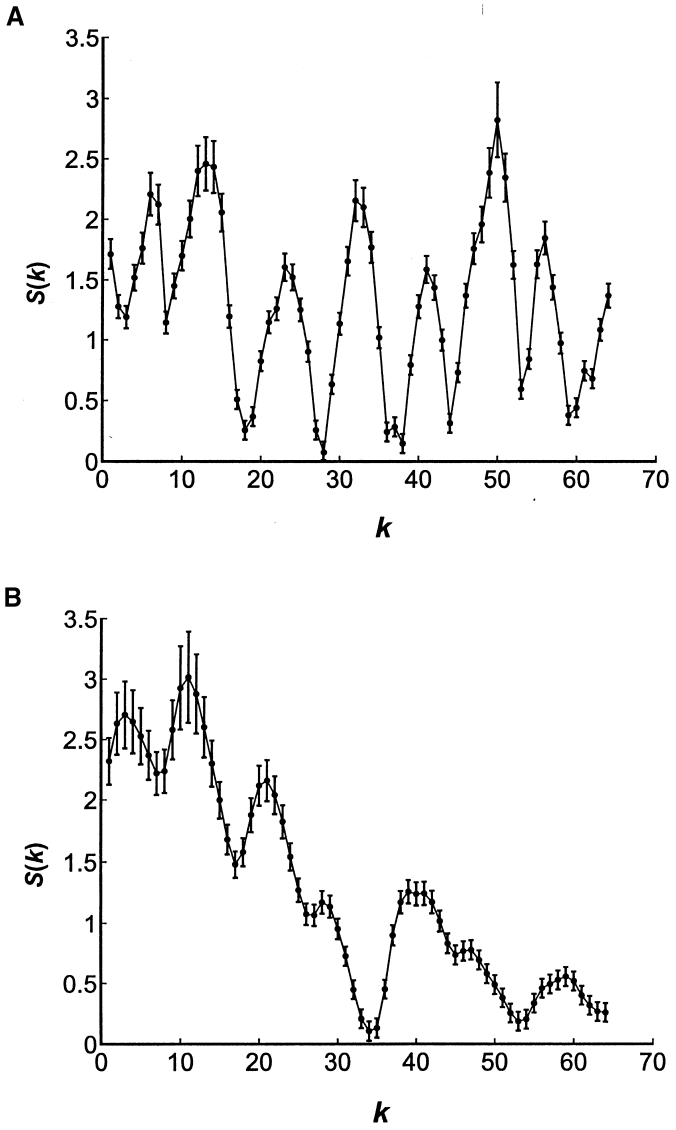

Fig. 7. Displays the results of Theiler's S (6) for scaled surrogates of Record 2 (A) and Record 16 (B) in Table 1 ($m_0 = 8$, $\tau_0 = 9$ and $N_s = 100$ with $K_0 = 20$ and $K_0 = 70$, respectively).

The pattern of slow (less than exponential) decay in predictability of $\rho(k)$ (Figs. 2, 3 and 8) is characteristic of deterministic systems with the possible involvement of strange attractors [8], this alone being suggestive of inherent dynamical non-linearities. The sharp decrease in $\rho(k)$ for

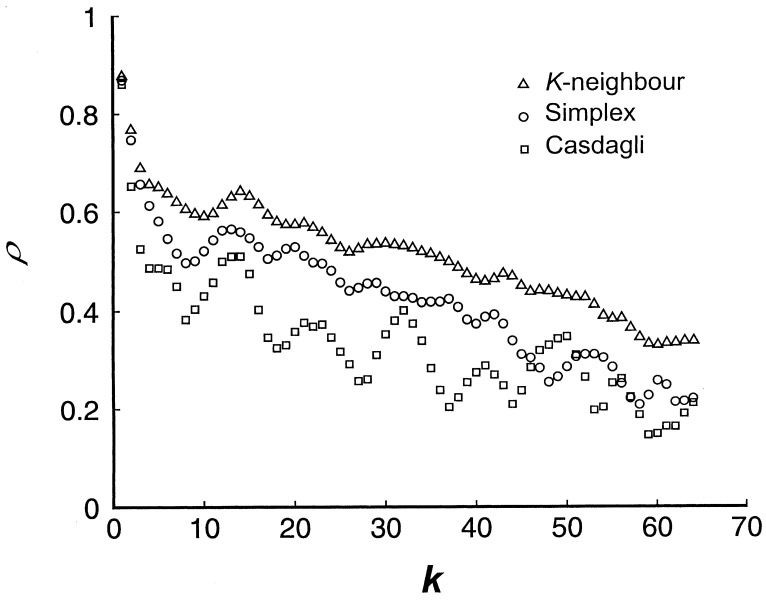

Fig. 8. Shows further comparison of the prediction algorithms emphasizing Casdagli's predictor large oscillatory behaviour for our data (Record 14 in Table 1: $m_0 = 8$, $\tau_0 = 9$ and $K_0 = 70$).

$1 \leqslant k \leqslant 3$, on the other hand, points to the presence of small amounts of added noise [6].

Another interesting feature concerns the virtual lack of difference in the results for scaled and unscaled surrogates which point to the presence of no significant static non-linear transformation of an essentially linear dynamics. Hence differences in prediction between the original series and their surrogates are most likely due to dynamic causes.

As to the degree of non-linearity in our series vis-à-vis the proposed linearity hypothesis rejection criterion involving the t-test (appropriate for the approximately normal character of the surrogate distributions) we must conclude that one cannot dismiss non-linearity as playing a role in the generating mechanisms of the time series. To some extent this result may seem conflicting with Theiler's proposal that non-linearity is assured only for S on the order of 10 [6]. The reasoning behind Theiler's claim must be taken with some care as it completely ignores the nature of the underlying distribution of surrogates and may be understood as based upon distribution independent arguments for setting probability bounds such as Chebyshev's inequality, wherefrom very loose confidence intervals can be inferred anyway. When the approximate normality of $z(\mathbf{x}_t^{(i)}, \hat{\mathbf{x}}^{(i)})(k)$ is taken into account, S values around 2 correspond to normal 0.05-percentiles, lending further support to the t-tests results.

Some very recent studies have raised concerns over conclusions drawn from the use of surrogates whose reliability is guaranteed only for some random processes [18]. In fact, in a recent proposal, Theiler and Prichard [19] suggest a procedure for experimentally gauging the actual levels of false positives in Theiler's method. Use of this computationally burdensome refinement, mainly involving computing surrogates of surrogates, brought into evidence a sensitivity of many non-linearity testing statistics to outliers which can decrease test power. Specifically it was observed that for some statistics, edge effects due to the Fourier transformations used in surrogate generation can increase the rate of false positives. Following a concern raised in [20], we investigated the effect of discarding some points (10,50,100) at the edges of our surrogates prior to prediction. We observed only marginal perturbations on the resulting mean and variance. Though more research is needed, we speculate over a possibly reduced sensitivity of Kravtsov's ρ to these edge effects when compared to other more extensively investigated statistics.

The main import of these observations to our current research [1] will be that of forcing us upgrade our current methods to study non-linear effects in addition to investigating dynamical relationships between diverse areas of the brain by linear multivariate time series methods. This may, for instance, allow us to explain some of the regularities observed in the present study, such as the roughly constant values of $m_0 = 8$ and τ_0 (around 9 and 10 samples) for all the theta wave time series considered. In our mind this observation alone already suggests some emergent collective non-linear behaviour of these diverse structures may be at play.

Acknowledgements

We should like to thank A.C. Valle, G. Ballester and C. Timo-Iaria for kindly providing the data for this work. We also thank the anonymous referees for their interesting suggestions.

References

[1] L.A. Baccala, K. Sameshima, Directed Coherence: a tool for exploring functional interactions among brain structures, in: M.A.L. Nicolelis (Ed.), Methods for Simultaneous Neuronal Ensemble Recordings, CRC, Boca Raton, FL, 1998, in press.
[2] S.L. Bressler, R. Coppola, R. Nakamura, Episodic multiregional cortical coherence at multiple frequencies during visual task performance, Nature 366 (1993) 153.
[3] T. Subba-Rao, M. Gabr, An Introduction to Bispectral Analysis and Bilinear Time Series Models, Lecture Notes in Statistics, vol. 24, Springer, Berlin, 1984.
[4] M.C. Mozer, Neural net architectures for temporal sequence processing, in: A.S. Weigend, N.A. Gershenfeld, (Eds.), Time Series Prediction, Forecasting the Future and Understanding the Past, Addison-Wesley, Reading, MA, 1994, p. 243.

[5] J. Fell, J. Röschke, C. Schäffner, Surrogate data analysis of sleep electroencephalograms reveals evidence for non-linearity, Biol. Cybern. 75 (1996) 85.
[6] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, J.D. Farmer, Testing for non-linearity in time series – the method of surrogate data, Physica D 58 (1992) 77.
[7] B. Efron, R.J. Tibshirani, An Introduction to the Bootstrap, Chapman and Hall, New York, 1993.
[8] G. Sugihara, R.M. May, Non-linear forecasting as a way of distinguishing chaos from measurement error in time series, Nature 344 (1990) 734.
[9] D. Kaplan, L. Glass, Understanding non-linear dynamics, Springer, New York, 1995.
[10] M. Casdagli, Non-linear prediction of chaotic time series, Physica D 35 (1989) 335.
[11] N.H. Packard, J.P. Crutchfield, J.D. Farmer, R.S. Shaw, Geometry from a time series, Phys. Rev. Lett. 45 (1980) 712.
[12] F. Takens, Detecting strange attractors in turbulence, in: D.A. Rand, L.S. Young, (Eds.), Lecture Notes in Mathematics, Springer, Berlin, 1981, p. 366.
[13] W.W. Taylor, Quantifying predictability for applications in signal separation, in: S. Haykin, (Ed.), SPIE, vol. 1565, Adaptive Signal Processing, 1991, p. 267.
[14] J. Theiler, Spurious dimension from correlation algorithms applied to limited time series data, Phys. Rev. A 34 (1986) 2427.
[15] M. Casdagli, Chaos and deterministic versus stochastic non-linear modeling, J.R. Statist. Soc. B 54 (1991) 303.
[16] J. Theiler, B. Galdrikian, A. Longtin, S. Eubank, Using surrogate data to detect non-linearity in time series, in: M. Casdagli, S. Eubank, (Eds.), Non-linear Modeling and Forecasting, Addison-Wesley, Redwood City, 1992, p. 303.
[17] A. Stuart, J.K. Ord, Kendall's Advanced Theory of Statistics, Griffin, London, vol. 1, 1987.
[18] K. Chan, On the validity of the method of surrogate data, Fields Inst. Commun. 11 (1997) 77.
[19] J. Theiler, D. Prichard, Using "surrogate data" to calibrate the actual rate of false positives in tests for non-linearity in time series, Fields Inst. Commun. 11 (1997) 99.
[20] J. Theiler, P.E. Rapp, Re-examination of the evidence for low-dimensional, non-linear structure in the human electroencephalogram, Electroenceph. clin. Neurophysiol. 98 (1996) 213.