Biol. Cybern. 84, 309–321 (2001)

Biological Cybernetics

Variability of the electric organ discharge interval duration in resting *Gymnotus carapo*

Alberto Capurro[1], Andre Longtin[2], Epifanio Bagarinao[3], Shunsuke Sato[3], Omar Macadar[1], K. Pakdaman[3,4]

[1] Institute of Biological Research Clemente Estable, Neurophysiology Department, Faculty of Science, Avenida Italia 3318, Montevideo 11600, Uruguay
[2] Physics Department, University of Ottawa, 150 Louis Pasteur, Ottawa, Ont., Canada K1N 6N5
[3] Division of Biophysical Engineering, Department of Systems and Human Science, Graduate School of Engineering Science, Osaka University, Toyonaka 560-8531, Osaka, Japan
[4] Inserm U444, Faculté de Médecine, Saint-Antoine, 27, rue Chaligny 75571, Paris Cedex 12, France

Received: 14 March 2000 / Accepted in revised form: 9 October 2000

Abstract. We recorded the electric organ discharges of resting *Gymnotus carapo* specimens. We analyzed the time series formed by the sequence of interdischarge intervals. Nonlinear prediction, false nearest neighbor analyses, and comparison between the performance of nonlinear and linear autoregressive models fitted to the data indicated that nonlinear correlations between intervals were absent, or were present to a minor extent only. Following these analyses, we showed that linear autoregressive models with combined Gaussian and shot noise reproduced the variability and correlations of the resting discharge pattern. We discuss the implications of our findings for the mechanisms underlying the timing of electric organ discharge generation. We also argue that autoregressive models can be used to evaluate the changes arising during a wide variety of behaviors, such as the modification in the discharge intervals during interaction between fish pairs.

1 Introduction

The fish *Gymnotus carapo* emits brief electric pulses in water. These pulses are produced by the electric organ, a biological "battery" commanded by a neuronal pacemaker located in the hindbrain. The electric organ discharge (EOD) consists of such pulses separated by comparatively longer intervals of silence (for review, see Bastian 1996). In resting *Gymnotus carapo*, the duration of the inter-EOD intervals displays low variability in the absence of stimulation (coefficient of variation ranging from 0.01 to 0.03). This regularity is intermediate between the highly regular EOD of South American wave-type species and the very irregular EOD of African mormyrids (Bullock 1969; Russell et al. 1974). The purpose of our study was to qualitatively and quantitatively analyze the resting discharge train in *Gymnotus carapo*.

The EOD of *Gymnotus carapo* is controlled by the electromotor pathway which subserves the coordination of the electric organ activation (for review, see Lorenzo 1990; Caputi 1999). Its pacemaker nucleus activates a group of relay neurons. These excite the spinal electromotorneurons, which in turn activate the electrocytes of the electric organ. Our companion study of *Gymnotus carapo* (Capurro et al. 1999b) has revealed that the variability of delay intervals between pacemaker discharges and their associated EOD pulses is one order of magnitude smaller than the resting variability of either the pacemaker or the EOD intervals. Further, the fluctuations of this delay series are uncorrelated, and the linear correlation between pacemaker and EOD interval series is 0.99, even during large interval decreases during "novelty responses" (NRs). A major consequence of this finding is that the variability in the inter-EOD intervals, and in particular its serial correlation, closely reflect those of the pacemaker. This implies that the inter-EOD intervals recorded in water provide a "low distortion window" for observing the intervals between pacemaker discharges, and that a statistical model for one can be used for the other.

The pacemaker nucleus consists of approximately 70–100 neurons. The field potential that they produce lasts around 1.2 ms. Interestingly, an intracellular spike lasts around 2 ms (Bennett et al. 1967; Lorenzo 1990), suggesting a high degree of synchronization between the neurons. It is thought that their strong electrical coupling to one another allows them to act as one large single neuron, or "syncitium", with several electrotonic compartments, each having its own axon (Bennett et al. 1967). The ionic currents underlying pacemaker activity has been studied in a related species (Dye 1991), but not yet in *Gymnotus carapo*.

In the wave-type fish *Eigenmannia* and *Apteronotus*, the pacemaker receives input from a few hundred neurons in the diencephalic pre-pacemaker structures. Some of these neurons fire at rest, and their changes in firing

Correspondence to: A. Capurro
(Tel.: +598-2-4875532, Fax: +598-2-4875548
e-mail: acapurro@iibce.edu.uy)

rate mediate the jamming avoidance behavior (Heiligenberg 1991). The pre-pacemaker neurons receive input from several higher-order sensory nuclei as well as from motor systems and forebrain structures. In *Gymnotus carapo*, pre-pacemaker neurons have not yet been identified, although preliminary results (D. Lorenzo, personal communication) have found two small groups of neurons retrogradely labeled after injection of biocytine in the pacemaker nucleus, as in *Eigenmannia*. Studying the pre-pacemaker structures may thus provide the key to the origin of inter-EOD interval variability.

Previous studies of the inter-EOD intervals variability have dealt mainly with the mean and standard deviation of the intervals, as well as with the shape of the interval histogram (Bullock 1969; Westby 1975; Kramer 1990; Capurro 1994; Moortgat et al. 1998; Capurro et al. 1999a). We take these analyses further in the case of *Gymnotus carapo*, by formulating a statistical model that accounts for the correlation between successive intervals. More precisely, nonlinear local interval prediction (Longtin 1993; Schiff et al. 1994), false nearest-neighbor analysis (Kennel et al. 1992), and a systematic comparison between the performance of fitted nonlinear and linear autoregressive (AR) models (Barahona and Poon 1996) reveal that the correlations in the data are essentially linear.

AR models are finite-order parametric models used to parsimoniously "describe" the linear temporal properties of time series in many disciplines, since the model parameters determine the autocorrelation function. These models are linear difference equations relating the "state" at time n (here: the nth interval fluctuation) to a finite number of previous states. They incorporate a zero-mean noise term that assumes a different value at each n; successive values of this noise are uncorrelated ("white noise") and distributed identically, usually in a Gaussian manner. Examples of the use of AR models in biological systems can be found in the analysis of electrophysiological data, in particular EEG measurements (see, e.g., Gersh 1970; Franaszczuk et al. 1985). Multivariate AR time series models were also used in the kinematic analysis of antennal scanning movements in two species of millipede (Giszter et al. 1984), in measuring different levels of synchronization of observed activities between channels (Fukunishi and Murai 1995; Franaszczuk and Bergey 1999), in detecting directionality of neural interactions (Bernasconi and Koenig 1999), and in parametric spectral analysis of cortical event-related potentials (Ding et al. 2000).

Our study shows that linear ARs generally provide better one-step prediction than nonlinear ARs for our data, and further, that linear ARs with shot noise (as well as the usual Gaussian white noise) provide a highly satisfactory model of EOD variability. A shot noise consists of impulses acting at discrete points in time. It is an idealization of physical processes producing very localized effects in time. For instance, such perturbations of an AR model were used for the detection of discrete secretory pulses of luteinizing hormone, involved in the control and disorders of reproductive functions (Thomas et al. 1992). In our study they are probably related to spikes or fast changes in the spiking rate of pre-pacemaker neurons (see below).

AR models are particularly useful when no other physically-based model is available. For example, there exists no biophysical model of the pacemaker and its synaptic inputs in *Gymnotus carapo*. AR modeling can then guide physical modeling, e.g., by helping to choose between different classes of physical models (e.g., nonlinear deterministic or linear stochastic). Also, EOD intervals generated by any candidate biophysical model should behave like the data, with respect to linear and nonlinear time series statistics.

The study of the resting discharge regime is thus a prerequisite of biophysical modeling studies. It is also required to quantitatively analyze changes in EOD variability during specific behaviors, which can be described only in comparison to the resting variability pattern. For example, in *Gymnotus carapo*, EOD variability increases due to shortening of some intervals: (1) in response to sensory stimulation, i.e., the NR (Bullock 1969), (2) after pharmacological treatments with serotonergic agents (Capurro et al. 1994), (3) during the interaction between two fish (Capurro et al. 1999a), (4) during active phases of the circadian cycle (Black-Cleworth 1970), (5) during the reproductive social behavior (for review, see Hagedorn 1986), and (6) during the escape response (Falconi 1997). Statistical models of the resting discharge can also be used in simulation studies of these behaviors (e.g., Capurro et al. 1998, 1999a).

2 Experimental recordings

Five specimens of *Gymnotus carapo*, with lengths in the range 7–10 cm, were gathered from Laguna del Sauce, a lake in Departamento de Maldonado in southeastern Uruguay. The fish were placed in separate tanks. Each tank contained a plastic tube that was open at both ends, with carbon electrodes to record the EOD. The tube was the preferred position of the fish during most of the light phase (which is the resting phase) of the circadian cycle, so that no constraint was necessary to make the animal enter the tube or remain in it. All experimental procedures were performed during the light phase of the circadian cycle, when the variability in the duration of the intervals is lowest. The voltage of the EOD was digitally recorded at equal time intervals of 0.05 milliseconds (i.e., sample frequency of 20 kHz). From these data, a file was created that contained two columns: one for the discharge time, and the other for the EOD interval ending in that discharge time.

The starting point of our analysis is the time series of the EOD interval sequence from which the mean EOD interval was subtracted. The resulting sequence is denoted throughout this paper as $I_1, I_2, \ldots I_N$. Thus, when we refer to the EOD interval sequence or time series, we refer to the sequence of variations of the intervals around their mean value. Since novel stimuli, such as weak mechanical vibrations or electric fields, as well as small movements of the fish induce transient frequency increases in the EOD (Capurro et al. 1994), we checked

visually that the fish were at rest during recordings, and that even small movements of the fins were absent. Furthermore, the recordings were done in a silent room with no changes in the lighting level, and used an antivibration table. Digital recordings that were contaminated by electric fields due to external sources were discarded.

3 Analysis methods

The time series and its autocorrelation function are first inspected visually. Then we assess whether the fluctuations have a deterministic component or whether they are predominantly stochastic (see Grassberger et al. for a review). This is achieved using nonlinear prediction (Sect. 3.1) and false nearest neighbor analysis (Sect. 3.2), as well as an approach based upon the comparison of best fitted linear and nonlinear ARs (Sect. 3.3). Following these analyses, we develop a statistical model for the sequence of inter-EOD intervals.

3.1 Nonlinear prediction

The first step of the nonlinear prediction algorithm (Farmer and Sidorowich 1987; Sauer 1994) is to construct an m-dimensional delayed embedding space containing vectors of the form $\vec{x}_n = (I_n, I_{n-\tau}, I_{n-2\tau}, \ldots, I_{n-(m-1)\tau})$, where I_k is the kth interval in the sequence $I_1, \ldots, I_N$, and τ is the embedding delay, i.e., the separation in time of the intervals used for each coordinate. τ is chosen to keep these coordinates as "independent" as possible. We present results for $\tau = 40$, for which the autocorrelation of intervals has decayed significantly. We find qualitatively similar results for $\tau = 1$, which assesses neighbors and predictability on a finer time scale. For each $\vec{x}_n$, we identify β nearest neighbors $\vec{x}_{p_j} \equiv (I_{p_j}, I_{p_{j-\tau}}, \ldots, I_{p_{j-(m-1)\tau}})$, $j = 1, \ldots, \beta$; we choose $\beta = 10$. Nearest neighbors close to $\vec{x}_n$ in time are omitted. Finally, the predicted interval f_n is computed as the average of the first components of the $\vec{x}_{p_j+1}$, i.e.,

$$f_n = \frac{1}{\beta}\sum_{j=1}^{\beta} I_{p_j+1} \quad . \tag{1}$$

The prediction error corresponding to f_n is thus the difference $f_n - I_{n+1}$, where I_{n+1} is the actual value of the next interval in the sequence of intervals. For the detection of nonlinearities, a normalized prediction error (NPE) is computed by first averaging the squares of all the prediction errors $(f - I)^2$ over the whole data set, and then dividing this result by the average error incurred by predicting the mean of the interval sequence, $\bar{I}$:

$$\mathrm{NPE} = \left[\frac{\langle (f_n - I_{n+1})^2 \rangle}{\langle (\bar{I} - I_{n+1})^2 \rangle}\right]^{1/2} \quad . \tag{2}$$

An NPE value close to zero signifies that there is predictability in the data; a value near one signifies that there is little predictability, and one can do just as well by forecasting the mean $\bar{I}$. This NPE is calculated as a function of the embedding dimension m, i.e., of the number of intervals used to identify neighbors and make predictions. This is because nonlinear structure in the data is not necessarily visible in a low-dimensional space where trajectories may intersect (i.e., the time evolution is not unique).

Since linear correlations in the data can also yield values of NPE less than one, it is crucial to compare nonlinear prediction results for the raw data with those for surrogate data sets (Theiler et al. 1992), i.e., stochastic processes having similar linear correlations as the raw data. If results for the raw data overlap those for various surrogate sets, one cannot make a case for nonlinearity in the dynamics.

Here we have generated surrogates using AR models for the sequence $I_1, I_2 \ldots, I_n, \ldots$ – see also Farmer and Sidorowich (1987) and Casdagli (1992), in the context of ordinary time series. These models are of the form:

$$I_{n+1} = \rho_1 I_n + \rho_2 I_{n-1} + \rho_3 I_{n-2} + \cdots + \rho_p I_{n-(p-1)} + \sigma W_{n+1} \tag{3}$$

where the W_i are independent, identically distributed, truncated, and centered Gaussian random variables, σ is the noise amplitude, and the coefficients ρ_i control the level of correlation. To determine the AR model for the data, we first estimated the first twenty autocorrelation coefficients of the time series using the Yule-Walker algorithm (Kay 1988). Subsequently, we estimated the coefficients of AR models of order one to ten by applying the Levinson-Durbin algorithm (Kay 1988). These algorithms are described in Appendix A.

Residuals – the difference between the one-step prediction of the AR model and the data – were computed for each of these models. The mean square error (i.e., mean residual) decreased as the order increased, until it reached a minimum. We selected the value of p at the first minimum of the mean square error as the order of the AR model for the data. We checked in all cases that the Akaike information criterion (Kay 1988) also reached its first minimum at the same value of p. The AR models were then used to generate surrogate data sets that captured the linear behavior of the data. Our method for generating surrogates contrasts with the common approach for time series based on Fourier transform phase-randomization procedures (Theiler et al. 1992).

3.2 False nearest-neighbor analysis

The analysis of the previous section was complemented by false nearest-neighbor analysis, which assesses how neighbors in the reconstructed phase space behave as the dimension of that space is increased (Kennel et al. 1992). For mainly deterministic nonlinear systems (i.e., those with low noise), neighbors are true neighbors if they remain neighbors as the embedding dimension m increases. Therefore, in such systems the percentage of

false nearest neighbors decreases near zero and remains there as m increases. Data generated by stochastic processes usually fill in the phase space differently, and the dependence of the percentage of false nearest neighbors on m can then be different from that above.

3.3 Autoregressive analysis

We further search for nonlinearity by using a technique that is particularly effective for time series heavily contaminated with noise and which, furthermore, directly compares the performance of linear AR models with nonlinear ones (Barahona and Poon 1996). This approach uses nonlinear autoregressive (NAR) models of the form

$$I_n^{\text{calc}} = \rho_0 + \rho_1 I_{n-1} + \cdots + \rho_p I_{n-p} + \rho_{p+1} I_{n-1}^2 + \rho_{p+2} I_{n-1} I_{n-2} + \cdots + \rho_M I_{n-p}^d + \sigma W_n \quad (4)$$

$$= \sum_{m=0}^{M} \rho_m z_m(n) + \sigma W_n \quad (5)$$

where $\{z_m(n)\}$ is composed of all the distinct product combinations of the reconstructed p-dimensional state-space coordinates $(I_{n-1}, I_{n-2}, \ldots, I_{n-p})$ up to degree d, the W_n's are independent and identically distributed Gaussian random variables, and σ is the noise standard deviation.

The basic idea is to compare the best linear model ($d = 1$) and nonlinear model ($d > 1$) for each data set. The best linear model is obtained by searching for the value of p that gives the first minimum of the Akaike information criterion with $d = 1$. Repeating this procedure for values of $d > 1$ yields the best nonlinear model. For each data set and each value of d and p, we estimated the AR and NAR coefficients ρ_i using Korenberg's fast orthogonal algorithm (Korenberg 1988). This algorithm is described in Appendix A.

For each of the resulting models, the residuals were computed and the standard deviations of each of the series of residuals were estimated. The presence of nonlinear determinism is established when the best nonlinear model is more predictive than the best linear model, i.e., σ_{NAR} is significantly smaller than σ_{AR}, where σ_{NAR} and σ_{AR} are the standard deviations of the residuals of the best NAR and best AR, respectively. Practically, we compared the variances σ^2_{NAR} and σ^2_{AR} using the non-parametric squared ranks test for variances (Conover 1980), which is described in Appendix A. We want to test the null hypothesis that $\{W_n\}_{\text{AR}}$ and $\{W_n\}_{\text{NAR}}$ – the residual series of, respectively, the best AR and best NAR models – are identically distributed except for possibly different means, and the alternative hypothesis being $\sigma^2_{\text{NAR}} < \sigma^2_{\text{AR}}$. The null hypothesis is rejected at the level of significance α if the test statistic U_{calc} is greater than its $(1-\alpha)$th quantile, denoted here as U_α.

3.3.1 Linear autoregressive plus innovational outlier model. The observation of a given time series may sometimes be affected by unusual events or disturbances, such as environmental fluctuations or sudden changes in a physical system. Such unusual observations may be referred to as *outliers*. To account for the presence of outliers in the data, we consider that instead of (5), the data can be modeled by a linear AR subject to a combination of both Gaussian and shot noise. The model is of the form:

$$I_n = \rho_0 + \sum_{r=1}^{p} \rho_r I_{n-r} + \sum_{r=1}^{k} a_r P_n^{T_r} + \sigma W_n \quad (6)$$

where $P_n^{T_r}$ represents an innovational outlier (IO) equal to 1 if $n = T_r$, and equal to 0 if $n \neq T_r$; k is the total number of outliers in the series, and a_r is the amplitude of the rth outlier. To fit this (AR + IO) model to the data, i.e., to estimate the parameters, we follow the scheme of Box et al. (1994), which is described in Appendix A. To detect the timing of the outliers, the model is first estimated assuming that no outliers are present. The residuals are computed and the maximum of their absolute values is obtained. If this maximum is larger than three standard deviations, then it is an outlier and the time of occurrence is recorded. A new series of residuals is then obtained by setting the maximum value to zero, and the standard deviation is re-estimated. This procedure is repeated until the maximum value of the residual series is less than three times the standard deviation. After all the outliers are detected, the coefficients in (6) are re-estimated using fast Korenberg algorithm. A revised set of residuals is obtained using this new set of coefficients. The same iterative procedure described above is applied to revised residual series until no new outliers are detected.

The resulting AR + IO models were compared to the best AR models as well as to the best NAR models. Comparisons based on the one-step prediction performance of the models were performed. The nonparametric squared ranks test for variances is used to test whether the best NAR models are performing significantly better than the AR + IO models. Furthermore, we tested whether the residuals for the best fit AR and AR + IO form a sequence of independent Gaussian random variables. To this end, we estimated the first one hundred autocorrelation coefficients of the residuals, and the Ljung and Box statistics (see Appendix A) was used to test for the lack-of-fit of the model (Ljung and Box 1978). The standard deviations of the residuals of the best fit ARs were compared to those of the AR + IO models. An extensive description of the methods used in this section can be found in Bagarinao (2000).

4 Results

4.1 Time series and autocorrelation function

The voltage versus time of the EOD of a *Gymnotus carapo* specimen recorded in water is shown in Fig. 1A and B. The high regularity of the train can be observed in Fig. 1A, while the typical EOD waveform of the species is depicted in Fig. 1B. Figure 1C shows an example of the sequence of inter-EOD intervals. Despite

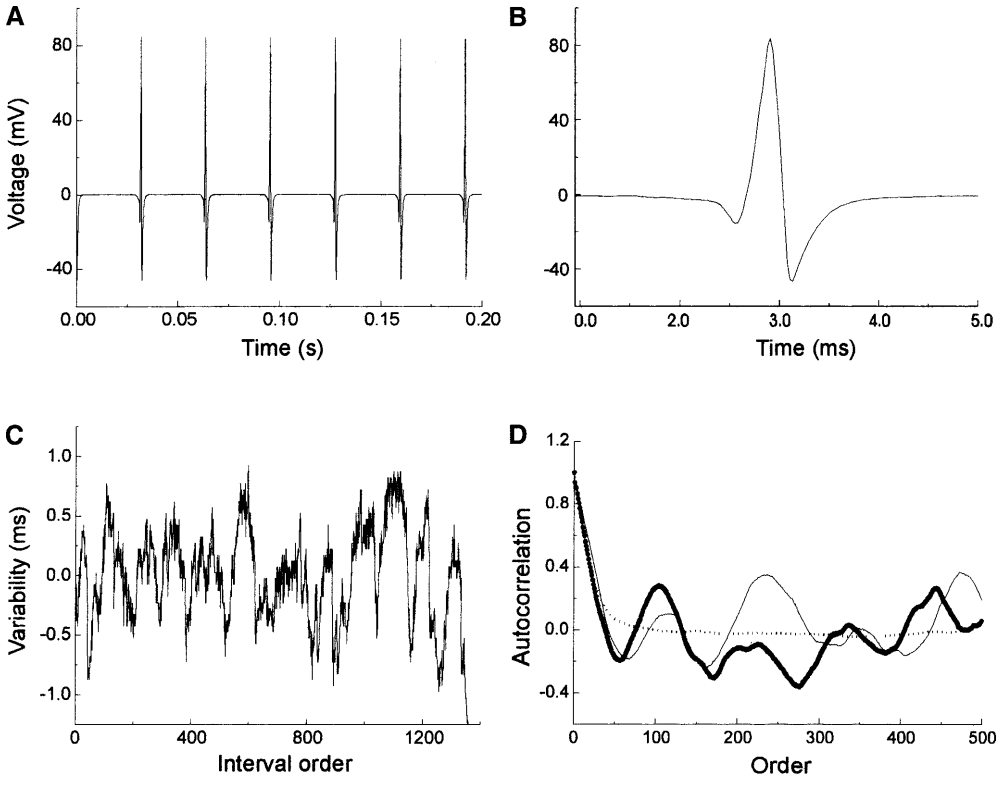

Fig. 1. **A** Electric organ discharge (EOD) voltage vs time (recorded in water with the positive electrode near the head of the fish and the negative electrode near its tail). **B** Same as **A**, but with a more expanded time scale to appreciate the multiphasic EOD waveform, which is a taxonomic feature of *Gymnotus carapo*. **C** Time series of the inter-EOD intervals of a resting *Gymnotus carapo*, 40 s of recording of fish 1. **D** Autocorrelation function of the time series shown in **C** (*circles* with *lines*), autocorrelation function of one realization of the linear autoregressive AR model fitted to the series in **C** (*solid line*) and averaged autocorrelation function of 1000 realizations of the same AR model (*dotted line*). The AR coefficients in this case were $\rho_1 = 0.667790$, $\rho_2 = 0.103294$, $\rho_3 = 0.305663$, $\rho_4 = -0.029653$, $\rho_5 = -0.039228$, $\rho_6 = -0.014113$, and $\rho_7 = -0.043400$

the regular aspect of the train shown in Fig. 1A, the variability of the intervals is noteworthy. One example of the autocorrelation function of the variability is shown in Fig. 1D. Its decay is slow, usually reaching 0.5 in 15–100 EOD cycles, depending on the fish.

4.2 Results of the nonlinear prediction analysis

The results of our NPE analysis for one-step-ahead prediction are shown in Fig. 2. The error increases naturally with the prediction time (not shown), and with embedding dimension m. The fact that the NPE values are significantly less than one suggests some predictability in the data, since it is possible to do better than forecasting the mean.

The results for five surrogate data sets generated from the AR model fitted to the respective data sets (Eq. 3) are also shown in Fig. 2 for the corresponding values of m. The order p was between four and seven in all cases. The AR model for each data set is given in Table 1. It is important to perform the analysis for more than one surrogate data set for each raw data set, because these sets are stochastic by nature (each one uses a different sequence of pseudorandom numbers), and any statistic calculated from them will have some variation about a mean. Comparison of the NPEs for the surrogates with those for the corresponding raw data sets shows that – in all cases – there is some significant overlap between the raw and surrogate NPE values. We thus conclude (see, e.g., Theiler et al. 1992) that we cannot reject the null hypothesis that the nonlinear predictability is in fact due to the linear correlations in the data. In other words, EOD fluctuations or "jitter" about their mean periodic motion are likely to be due to linearly correlated noise, rather than to some deterministic property such as chaotic dynamics. Our results cannot exclude the possibility of noise-perturbed chaotic patterns or periodic phase-locked patterns with noise. The key point is that when there is a minor indication of nonlinearity (as in Fig. 2C and D), this property does not endow the data with much more predictability than the linear data properties do alone; thus, if present, nonlinear dynamics are likely to be only a minor component of the dynamics. Our results are qualitatively similar for an embedding delay $\tau = 1$; NPE values are then relatively constant across embedding dimensions (data not shown).

The percentage of false nearest neighbors was found to remain near zero for $m > 4$ for all data sets (data not shown), for both embedding delays of $\tau = 40$ and $\tau = 1$. This suggests that some nonlinear structure may be present for these data. However, our surrogate data sets also have similar behavior. This is presumably due to the very long-lived linear correlations in these data sets. Thus, the false nearest-neighbor analysis, corroborates the results of the NPE analysis, in that there are no significant differences between the linear surrogates and the data.

4.3 Results of the autoregressive analysis

As the NPE and false nearest-neighbor analyses did not detect significant differences between data and surrogates, we proceed further by comparing the performance of

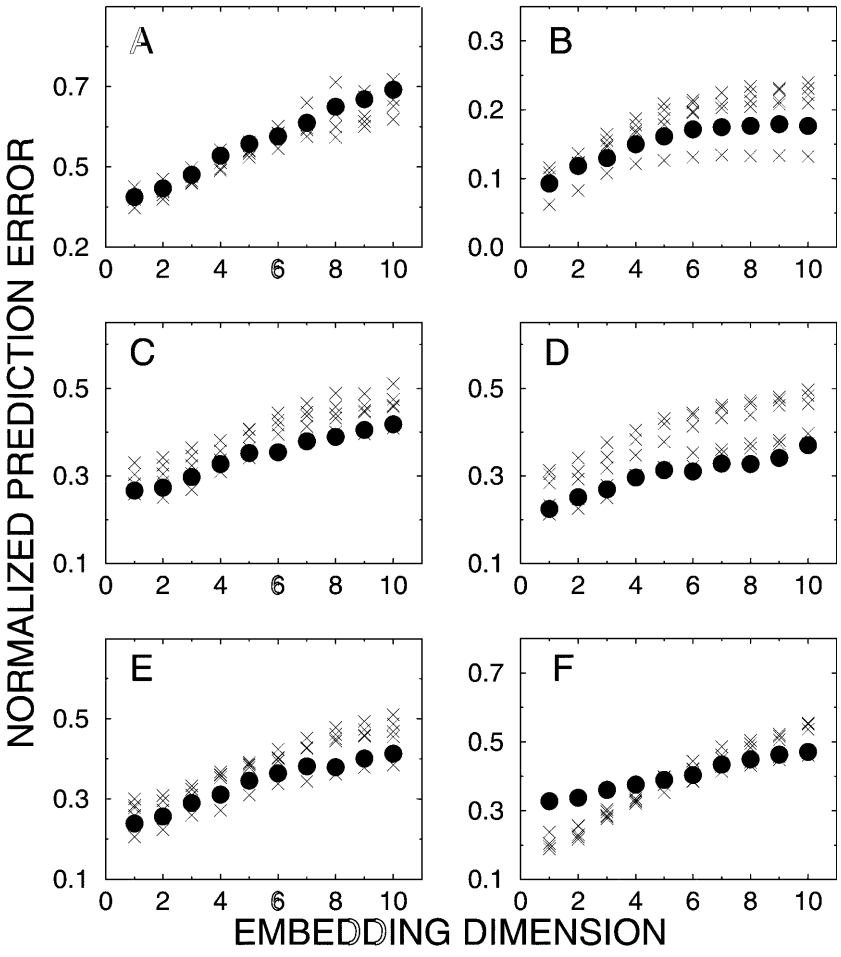

Fig. 2A–F. Normalized prediction error (NPE) as a function of embedding dimension for **A** fish1 (selected segment of 40 s); **B** fish 1 (complete file of 60 s); **C** fish 2; **D** fish 3; **E** fish 4; and **F** fish 5. Results are for one-step-ahead prediction NPE versus m; results for the five surrogate data sets are plotted as *crosses*

Table 1. Coefficients of the surrogate ARs

Fish	ρ_1	ρ_2	ρ_3	ρ_4	ρ_5	ρ_6	ρ_7
Fish 1	0.806564	0.127666	0.164781	−0.066841	−0.034220		
Fish 2	0.539676	0.315605	0.252485	−0.056097	−0.022186	−0.046117	−0.003731
Fish 3	0.544062	0.275485	0.250865	−0.006365	0.013795	−0.067874	−0.024061
Fish 4	0.800585	0.010637	0.166458	0.003430			
Fish 5	1.001973	0.051346	−0.009777	−0.063952	0.032197	0.005926	−0.045951

Table 2. Comparison of the best AR and the best NAR model. σ_{AR} is the SD of the residuals of the AR model, σ_{NAR} is the SD of the residuals of the best NAR model, U_{calc} represents the estimated squared ranks statistics, and $U_{0.05}$ is the value of U at the 0.05 significance level

Fish	σ_{AR}	σ_{NAR}	U_{calc}	$U_{0.05}$
Fish 1	0.133	0.125	1.766	1.645
Fish 2	0.115	0.113	0.665	1.645
Fish 3	0.192	0.185	0.876	1.645
Fish 4	0.289	0.283	0.428	1.645
Fish 5	0.171	0.144	3.438	1.645

linear AR models with nonlinear ones. These results are summarized in Table 2. The U values indicate that the nonlinear models are not significantly better than the linear ones for fish 2, 3, and 4. This agrees with the NPE and the false nearest-neighbor analyses. Surprisingly, for fish 1 and 5, the U test indicates that the NAR is a better one-step predictor than the AR. This seems to contradict the results of the NPE and false nearest-neighbor analyses. However, in the following, we show that the reason for this apparent discrepancy is not the presence of nonlinearities that went undetected in the previous methods, but rather it is due to the fact that the assumption about the noise in the AR models is not appropriate.

To establish this latter point, we consider – instead of (5) – that the data can be best modeled by a linear AR subject to a combination of Gaussian and shot noise (AR + IO; see Eq. 6). Tables 3 and 4 summarize the results of the comparison between AR + IO and NAR models and AR + IO and AR models, respectively. The U values indicate that the NARs are not significantly better than the AR + IOs. In fact, for fish 5, which has the largest number of outliers, the performance of the AR + IO model as a one-step predictor is significantly better than the best NAR model. Furthermore, the

Table 3. Comparison of the AR + IO and the best NAR model. σ_{AR+IO} is the SD of the residuals of the AR + IO model

Fish	σ_{AR+IO}	σ_{NAR}	U_{calc}	$U_{0.05}$
Fish 1	0.124	0.125	0.780	1.645
Fish 2	0.106	0.113	−0.711	1.645
Fish 3	0.180	0.185	−0.092	1.645
Fish 4	0.262	0.283	−1.024	1.645
Fish 5	0.132	0.144	−0.242	1.645

residual time series of the ARs with IOs are completely free of outliers, as compared to that of the AR alone. These result in a significant decrease of the standard deviations of the residuals. Furthermore, for the AR + IO model, all lack-of-fit tests of the autocorrelation of the residuals were satisfactory. For the standard ARs, this was not the case for fish 3. These results show that the AR + IOs perform better or as well as the best NARs and the best ARs. Furthermore, the residuals of the AR + IOs, in contrast with some of those of the best ARs and best NARs, successfully pass all the tests pertaining to their independence and Gaussian distribution. This is illustrated in Fig. 3, which shows the residuals of the best fit AR, NAR, and AR + IO models for fish 5. The AR coefficients of the final models are given in Table 5.

We would like to remark that the oscillations that have variable amplitude and regularity present in the autocorrelation function of the experimental data (circles with lines in Fig. 1D), are not incompatible with linear AR models. As illustrated by the solid line in Fig. 1D, similar oscillations were also obtained when the autocorrelation coefficients were estimated from the same length of data obtained from the simulations of the AR model (fitted to the original data). Given that such marked oscillations are absent from the averaged autocorrelation function of 1000 realizations of AR simulations (dotted line), their presence can be attributed to the sample path variability resulting from the finite length of the data used for the estimation of the autocorrelation coefficients (the bigger oscillations appeared in the autocorrelation function of fish 1, while in the other fish they were present to a lesser extent). This is also supported by the Ljung and Box test performed on the residuals of the fish data, which indicate that the sequence of residuals is compatible with independent realizations of a Gaussian random variable, so that henceforth the data can be accurately modeled by AR models.

In summary, our analysis indicates that the EOD interval fluctuations can be essentially modeled by a linear AR, with sudden jumps. The present data do not allow the development of a statistical model for occurrence times and amplitudes of these jumps; this will be dealt with in a future study.

5 Discussion

We studied the fluctuations of the EOD intervals of the electric fish *Gymnotus carapo* in resting conditions, and we presented a statistical model for the sequence of intervals observed. The analysis of the correlations in the inter-EOD intervals constitutes, to our knowledge, the first study that qualitatively and quantitatively describes the higher-order statistics of the EOD intervals in electric fish. Indeed previous studies have mainly concentrated on the evaluation and comparisons of the means, standard deviations, and coefficient of variation of the inter-EOD intervals.

Our analysis shows that a linear autoregressive model perturbed by both white Gaussian noise and shot noise is the most satisfactory model for the data. This model is the only one that fits all the data, in the sense that the resulting sequence of residuals are independent and have a Gaussian distribution in all cases. A remarkable property of this model is that most of the time it corresponds to the standard AR, with occasional interval shortening caused by the shot noise. More precisely, segments of the data devoid of outliers (of 20–40 duration) were systematically analyzed and were found to be satisfactorily modeled by the standard AR with Gaussian white noise (data not shown).

Our recording method allows effective *non-invasive* monitoring of the fluctuations of the pacemaker nucleus, and indirectly of the pre-pacemaker nucleus. In light of our current knowledge of mechanisms of pacemaker variability and of our results, here we discuss the significance of linear versus nonlinear correlations, comment on the functional role of noise in neural systems, and provide an outlook into future directions of research.

Table 4. Comparison of the AR and the AR + IO models. N is the number of residuals used in the analysis, Q is the estimated Ljung and Box statistics, Q_{crit} gives the value of Q at the 0.01 significance level, DSD is the decrease in the standard deviation computed as $DSD = (\sigma_{AR} - \sigma_{AR+IO})/\sigma_{AR}$, and NO is the number of outliers present in the residual series

Fish	Model	N	p	σ	Q	Q_{crit}	DSD (%)	NO
Fish 1	AR	2055	5	0.133	87.847	129.97	–	16
	AR + IO	2055	5	0.124	93.571	129.97	7.1	0
Fish 2	AR	1940	6	0.115	98.759	128.80	–	21
	AR + IO	1940	6	0.106	97.021	128.80	7.8	0
Fish 3	AR	1875	6	0.192	133.123	128.80	–	13
	AR + IO	1875	7	0.180	126.647	127.63	6.2	0
Fish 4	AR	1775	3	0.289	100.931	132.31	–	21
	AR + IO	1775	3	0.262	98.139	132.31	9.3	0
fish 5	AR	1587	4	0.171	86.409	131.14	–	27
	AR + IO	1587	7	0.132	100.678	127.63	22.8	0

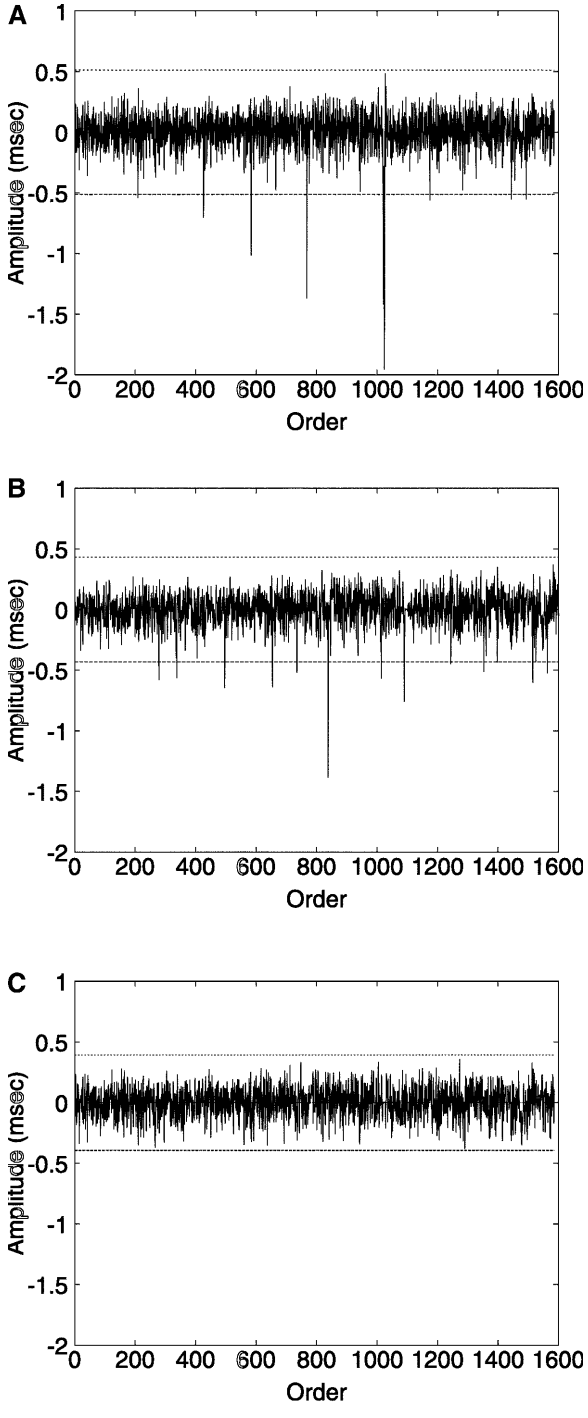

Fig. 3A–C. Residuals of fish 5 models: **A** AR; **B** nonlinear autoregressive (NAR); **C** linear autoregressive plus innovational outlier (AR + IO)

5.1 Linear and nonlinear correlations

Nonlinearity is the rule rather than the exception in biological systems, especially for pacemakers (e.g., Glass and Mackey 1988; Hayashi and Ishizuka 1992; Longtin and Racicot 1997a; Segundo et al. 1998). An important point to realize is that the pacemaker nucleus oscillates without external periodic input. This system is then likely to be in a autonomous limit cycle state, and this is a nonlinear property. This limit cycle has a mean period equal to the mean EOD interval, which was subtracted from our data prior to our analyses. The results of our analysis do not question this nonlinear behavior. Rather, they concern the fluctuations around the mean inter-EOD period, arising from intrinsic and extrinsic influences on this limit cycle.

We have assessed whether these fluctuations can be modeled by linear AR processes, or whether further nonlinear processes are involved. Our analysis reveals that they are best modeled by linear AR processes with long-decaying correlation and outliers, i.e., by noise. Our analysis thus excludes the possibility that the fluctuations arise from strong dynamic nonlinearities, as would be the case if, for example the oscillation were chaotic rather than a limit cycle. It is of course possible that weaker dynamic nonlinearities are at play, and that the "noise" is in fact an unresolvable (using our methods) combination of noise and determinism.

The variability of firing in simple neuron models, measured for example by the coefficient of variation of firing intervals, is known to depend on a number of parameters such as the distance between resting potential and threshold and, for pacemakers, on the precise dynamical mechanisms underlying periodic spike generation (see, e.g., Lánský and Sato 1999). There is little knowledge however on the correlation properties of interval fluctuations for different models. And to our knowledge, no models yet account for the long-lived correlations seen in our pacemaker data. These correlations will depend on the stability of the pacemaker limit cycle to perturbations. Such issues are currently under investigation.

5.2 The origin and functional significance of the variability

To a certain extent the origin of variability may be separable from that of its potential functional significance. Functional implications may however depend on

Table 5. AR coefficients of the final models

Fish	ρ_0	ρ_1	ρ_2	ρ_3	ρ_4	ρ_5	ρ_6	ρ_7
Fish 1	0.000942	0.671543	0.203362	0.237763	−0.036813	−0.076328		
Fish 2	0.005311	0.491533	0.314534	0.283100	−0.052195	−0.008962	−0.048943	
Fish 3	0.007252	0.503950	0.287591	0.249927	0.015180	0.024365	−0.071875	−0.031984
Fish 4	0.012477	0.777074	0.025463	0.176697				
Fish 5	0.013912	0.784023	0.089700	0.047697	0.009994	0.090963	0.012050	−0.070122

how sources of variability are modified by behaviorally relevant stimuli. A strong interaction between experiment, data analysis and theoretical modeling, guided by the results of the present study, will be key to understanding these issues.

We know from our previous study (Capurro et al. 1999b) that what is observed in water is already present in the pacemaker. Our analyses of the EOD train suggest significant variability and correlations in these pacemaker discharges. The variability depends on several factors, which can be divided into two groups: intrinsic and extrinsic to the pacemaker nucleus. The internal sources, common to all excitable cells, include conductance fluctuations, channel noise, thermal fluctuations, and ionic-pump current fluctuations. The coupling and resulting synchrony of the pacemaker cells will also affect the expression of these noises. The main extrinsic source is likely to be synaptic noise arising from the synaptic release mechanism itself as well as from the random arrival times of action potentials. One can also add external electrical fields, and other biological rhythms such as respiration (Westby 1975).

The extrinsic influence that is likely to have the most important influence on the correlations of the pacemaker is its synaptic input from a pre-pacemaker nucleus. It is possible that post-synaptic potentials fit the time scale corresponding to 3–7 EOD intervals, i.e., the order of our best AR + OI models (90–210 ms considering a mean EOD interval of 30 ms). Correlations may be related to the resting activity of the pre-pacemaker neurons. Further, shots of noise may be caused by fast changes in the discharge pattern of these neurons, or by the activation of one or many pre-pacemaker neurons that are usually silent under resting conditions (for a review on pre-pacemaker neurons, see Heiligenberg 1991). Whether these last changes are of the same type as those induced by novelties in the sensory system remains an open question that deserves further investigation. Indeed, the neuronal systems that produce a transient shortening of intervals in response to environmental stimuli include the pre-pacemaker neurons. Thus, the shots may be due to spontaneous variations in their discharging pattern, or correspond to small NRs due to very weak environmental stimulus that could have escaped our experimental controls. This last possibility, which could reflect the exquisite sensitivity of the fish to environmental fluctuations (Westby 1975), is more likely when the IOs occur in trains, because the NR involves shortening of many intervals. Although most outliers are isolated (i.e., not in trains) we detected a train of outliers around interval 1100 of fish 5 (upper panel of Fig. 3), and the shortening of intervals that caused this train has the typical time course of a NR (data not shown). The possible relation between NR and the shots of noise deserves further investigation.

Synaptic noise in the form of spontaneous release of synaptic vesicles is not likely to cause the shots, because the membrane potential of the pacemaker rises in a smooth way, and reaches the threshold with a high slope thus making the pacemaker quite robust to such small perturbations. After a peripheral stimuli the slope of the pacemaker increases, producing a shortening of the intervals (D. Lorenzo, personal communication), in a similar way to that which occurs with an intracellular current pulse (Bennett et al. 1967). Thus, in order to produce a shot of noise, at least one spike in the pre-pacemaker neurons (spontaneous or induced by peripheral stimuli) is probably needed. The origin and statistics of these shots deserve further study, because they may be related to the changes that occur during behavior. A final issue is whether correlations are due mainly to inputs to the pacemaker, or whether they arise from the response of the pacemaker to noise that is largely uncorrelated.

5.3 The AR + OI model as a basis for further behavioral studies

The statistics of the times of occurrence of the shots in the AR + OI model may vary during different behavioral displays. For example, shots of noise are quite rare in resting conditions, but not in the active stage of the circadian cycle. Such models can also provide a basis for understanding the changes that occur during other behaviors. For example, the resting discharge pattern, simulated by an AR model, can be used to simulate the jamming avoidance response, e.g., by adding adequate frequency transients to the predicted intervals (Capurro et al. 1999a). The efficiency of the transients in preventing jamming (i.e., avoiding simultaneous discharges between two EOD trains) may improve significantly when the actual variability of the EOD is simulated with our fitted AR models, rather than with less correlated or uncorrelated noise (Capurro 1999).

5.4 Concluding remarks

The inter-EOD intervals are modulated by pre-pacemaker neurons that receive input from very wide areas of the brain. Thus, it seems natural that every behavioral display, including resting conditions (which is the goal of the present effort), will be reflected in the correlation of the inter-EOD intervals. This time series contains information that can then be used to characterize the state of the system that produces this signal, since changes in electroencephalograms or electrocardiograms are related to the behavioral state of mammals.

Acknowledgements. This work was partially supported by JSPS and Osaka University (Japan), NSERC (Canada), and PEDECIBA (Uruguay). AC gratefully acknowledges personal contributions from Mrs. DSU and Mrs. ASC (Uruguay).

Appendix A: Statistical methods

A.1 Yule–Walker equations

Consider an autoregressive process of order p given by

$$x_n = a_1 x_{n-1} + a_2 x_{n-2} + \cdots + a_p x_{n-p} + \varepsilon_n \ . \tag{A1}$$

An important relation of the autocorrelation function of this process is found by multiplying x_{n-k} throughout, to obtain

$$x_{n-k}x_n = a_1 x_{n-k}x_{n-1} + \cdots + a_p x_{n-k}x_{n-p} + x_{n-k}\varepsilon_n \ . \tag{A2}$$

Taking the expectation gives

$$\gamma_k = a_1\gamma_{k-1} + a_2\gamma_{k-2} + \cdots + a_p\gamma_{k-p} \ , \tag{A3}$$

where $\gamma_k = E[x_{n-k}x_n]$ and $E[\varepsilon_n x_{n-k}] = 0$ when $k > 0$. Dividing this equation by γ_0 gives the autocorrelation function ρ_k satisfying the relation

$$\rho_k = a_1\rho_{k-1} + a_2\rho_{k-2} + \cdots + a_p\rho_{k-p} \ . \tag{A4}$$

To get the coefficients a_k values from ρ_k, k is substituted for values $1, \ldots, p$ in (A4) to form the set of equations

$$\begin{aligned} \rho_1 &= a_1 + a_2\rho_1 + \cdots + a_p\rho_{p-1} \\ \rho_2 &= a_1\rho_1 + a_2 + \cdots + a_p\rho_{p-2} \\ &\vdots \quad\quad \vdots \\ \rho_p &= a_1\rho_{p-1} + a_2\rho_{p-2} + \cdots + a_p \end{aligned} \tag{A5}$$

where $\rho_0 = 1$ and $\rho_{-k} = \rho_k$. The above equations are usually called the *Yule–Walker* equations, and the *Yule–Walker estimates* of the parameters are obtained by replacing the theoretical autocorrelations ρ_k by the estimated autocorrelations. The solution of (A5) for the parameters in terms of the autocorrelations may be written as

$$\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{p-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{p-2} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \rho_{p-1} & \rho_{p-2} & \rho_{p-3} & \cdots & 1 \end{bmatrix}^{-1} \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_p \end{bmatrix} \ . \tag{A6}$$

A.2 Estimation of AR coefficients by Levinson–Durbin algorithm

The *Yule–Walker* estimates of the parameters of an AR($p+1$) model may be obtained when the estimates for an AR(p) model, fitted to the same time series, are known. This is given by the following recursive formulas (Levinson–Durbin algorithm):

$$a_{p+1}(p+1) = \frac{\rho_{p+1} - \sum_{j=1}^{p} a_j(p)\rho_{p+1-j}}{1 - \sum_{j=1}^{p} a_j(p)\rho_j} \tag{A7}$$

$$a_j(p+1) = a_j(p) - a_{p+1}(p+1)a_{p+1-j}(p) \tag{A8}$$

where the $a_j(p)$ values are the known parameters of the AR(p) model, the $a_j(p+1)$ values are the parameters of the AR($p+1$) model that needs to be computed, and $j = 1, 2, \ldots, p$.

A.3 Korenberg's fast orthogonal algorithm

This section briefly explains Korenberg's algorithm that is used to estimate the coefficients of a nonlinear autoregressive model given by

$$\begin{aligned} y_n^{\text{calc}} &= g(Y_{n-1}; \mathbf{a}) + \epsilon_n \\ &= a_0 + a_1 y_{n-1} + \cdots + a_d y_{n-d} + a_{d+1} y_{n-1}^2 \\ &\quad + a_{d+2} y_{n-1} y_{n-2} + \cdots + a_M y_{n-d}^k + \epsilon_n \qquad \text{(A9)} \\ &= \sum_{m=0}^{M} a_m z_m(Y_{n-1}) + \epsilon_n \qquad \text{(A10)} \end{aligned}$$

where $Y_{n-1} = (y_{n-1}, y_{n-2}, \ldots, y_{n-d})$ represents a vector in the reconstructed d-dimensional state space (embedding space), the functional basis $\{z_m(X)\}$ is composed of all the distinct combinations of coordinates up to degree k, d is the delay which also corresponds to the dimension of the embedding space, ϵ_n represents the random forcing of the system, and $M + 1 = (k+d)!/(d!k!)$ determines the number of coefficients to be computed.

The basic idea is to use auxiliary orthogonal functions to estimate the model parameters. To do this, the right-hand side of (A10) is rearranged into a sum of terms that are mutually orthogonal over the given time series:

$$g(n; \mathbf{a}) = \sum_{m=0}^{M} b_m w_m(n) \tag{A11}$$

where $w_m(n)$ are orthogonal functions constructed from the $z_m(n)$ using the Gram-Schmidt orthogonalization procedure

$$w_m(n) = z_m(n) - \sum_{r=0}^{m-1} \alpha_{mr} w_r(n) \tag{A12}$$

where

$$\alpha_{mr} = \frac{\sum_{n=0}^{N} z_m(n) w_r(n)}{\sum_{n=0}^{N} [w_r(n)]^2} \tag{A13}$$

The expansion coefficients, b_m, are computed using the orthogonality of $w_m(n)$ and is given by

$$b_m = \frac{\sum_{n=0}^{N} y_n w_m(n)}{\sum_{n=0}^{N} [w_r(n)]^2} \ . \tag{A14}$$

The main advantage of this formulation is that the explicit creation of the orthogonal functions $w_m(n)$ in (A12) is unnecessary. Define

$$D(m, r) = \frac{1}{N+1} \sum_{n=0}^{N} z_m(n) w_r(n) \tag{A15}$$

$$E(m) = \frac{1}{N+1} \sum_{n=0}^{N} [w_m(n)]^2 \tag{A16}$$

$$C(m) = \frac{1}{N+1} \sum_{n=0}^{N} y_n w_m(n) \ . \tag{A17}$$

Using (A12) and the orthogonality of $w_m(n)$, it can be shown that these quantities satisfy the following relations

$$D(m,r) = \frac{1}{N+1}\sum_{n=0}^{N} z_m(n) z_r(n) - \sum_{i=0}^{r-1} \alpha_{ri} D(m,i) \qquad \text{(A18)}$$

$$E(m) = \frac{1}{N+1}\sum_{n=1}^{N} [z_m(n)]^2 - \sum_{r=0}^{m-1} \alpha_{mr}^2 E(r) \qquad \text{(A19)}$$

$$C(m) = \frac{1}{N+1}\sum_{n=0}^{N} y_n z_m(n) - \sum_{r=0}^{m-1} \alpha_{mr} C(r) \ . \qquad \text{(A20)}$$

The above equations can be used to compute $D(m,r)$, $E(m)$, and $C(m)$ iteratively. α_{mr} and b_m can now be obtained using

$$\alpha_{mr} = \frac{D(m,r)}{E(r)} \qquad \text{(A21)}$$

$$b_m = \frac{C(m)}{E(m)} \ . \qquad \text{(A22)}$$

The original model parameters a_m are determined from the computed b_m and α_{mr} using the following relations:

$$a_m = \sum_{i=m}^{M} b_i V_i \ , \qquad \text{(A23)}$$

where $V_m = 1$ and

$$V_i = -\sum_{r=m}^{i-1} \alpha_{ir} V_r \qquad \text{(A24)}$$

for $i = m+1, \ldots, M$.

A.4 Nonparametric squared ranks test for variances

Let $X_1, X_2, \ldots, X_N$ denote the residual series of the best AR model. Also, let $Y_1, Y_2, \ldots, Y_N$ denote the residual series of the best NAR model. To compute U_{calc}, the series $\{X_i\}$ and $\{Y_i\}$ are first converted to their absolute deviation from their respective means, that is, $U_i = |X_i - \mu_1|$ and $V_i = |Y_i - \mu_2|$, where $i = 1, \ldots, N$ and μ_1 and μ_2 are their respective means. The combined samples are then ranked from 1 to $2N$, where N is the length of each residual series. If a tie occurs among several values of the absolute deviations, each is assigned the average of the ranks that would have been assigned to them had there been no ties. Let $R(U_i)$ and $R(V_i)$ denote the ranks and the averaged ranks thus assigned for the absolute deviations of the two residual series $\{X_i\}$ and $\{Y_i\}$. The test statistics is defined as

$$T = \sum_{i=1}^{N} [R(U_i)]^2 \qquad \text{(A25)}$$

if there are no ties. Or,

$$T_1 = \frac{T - N\bar{R}^2}{\sqrt{\frac{N^2}{2N(2N-1)}\sum_{i=1}^{2N} R_i^4 - \frac{N^2}{2N-1}(\bar{R}^2)^2}} \qquad \text{(A26)}$$

where

$$\bar{R}^2 = \frac{1}{2N}\left\{\sum_{i=1}^{N}[R(U_i)]^2 + \sum_{i=1}^{N}[R(V_i)]^2\right\} \qquad \text{(A27)}$$

$$\sum_{i=1}^{2N} R_i^4 = \sum_{i=1}^{N}[R(U_i)]^4 + \sum_{i=1}^{N}[R(V_i)]^4 \qquad \text{(A28)}$$

if there are ties. U_{calc} is defined as being equal to T if there are no ties, or equal to T_1 if there are ties. The null hypothesis H_o is that X and Y are distributed identically except for possibly different means, and the alternative hypothesis H_1 is that the variance of X is greater than the variance of Y. One can reject H_o at the level of significance α if U_{calc} is greater than its $1-\alpha$th quantile.

A.5 Estimation of AR + IO model parameters using the scheme of Box

Let z_n denote the underlying time series process which is free of the impact of outliers, and let Y_n denote the observed time series. Assume that

$$z_n = \sum_{i=1}^{p} a_i z_{n-i} \ . \qquad \text{(A29)}$$

In the presence of observational outliers (additive outliers), the observed time series can be written as

$$Y_n = w_T P_n^T + z_n \qquad \text{(A30)}$$

where $P_n^T = 1$ if $n = T$ or $P_n^T = 0$ if $n \neq T$ is the impulse indicator at time T, and w_T is the corresponding weight.

To detect the timing of the outliers, the following iterative procedure is used:

1. Estimate an AR model using the observed time series $\{Y_n\}$, that is, $Y_n^{\text{calc}} = \sum_{i=1}^{p} b_i Y_{n-i}$.
2. Compute the residuals and the variance using $\hat{e}_n = Y_n - Y_n^{\text{calc}}$ and $\hat{\sigma}_a^2 = N^{-1}\sum_{n=1}^{N} \hat{e}_n^2$, respectively.
3. For each time n, compute $\hat{\lambda}_n = \hat{e}_n/\hat{\sigma}_a$.
4. Compute $\gamma_T = \max_n|\hat{\lambda}_n|$, where T denotes the time of the occurrence of the maximum. An outlier occurs at T if $\gamma_T > 3$. Record the value of T.
5. Set the residual to zero at time T, i.e., $\hat{e}_T = 0$. Then, compute again $\hat{\sigma}_a$ from the modified residuals.
6. Repeat step 3 using the modified residuals until all outliers are identified and their timing is recorded.
7. After all the outliers are detected, compute again the parameters of the model $Y_n^{\text{calc}} = \sum_{j=1}^{K} w_j P_n^{T_j} + \sum_{i=1}^{p} b_i Y_{n-i}$, where T_j denotes the occurrence times of the K outliers.
8. With the new estimates of the model parameters, steps 1–6 can be repeated to check for the presence of additional outliers.
9. The AR + IO model parameters can now be estimated using the known timing of the outliers by Korenberg's algorithm, with P_n^T as an additional term to the model.

A.6 Ljung and Box statistics

Consider the residuals of the best AR model:

$$\hat{e}_n = x_n - \sum_{i=1}^{p} a_i x_{n-i} \quad . \tag{A31}$$

It is useful to study the adequacy of the fit by examining $\{\hat{e}_n\}$ and, in particular, their autocorrelations

$$\hat{r}_k = \frac{\sum_{n=k+1}^{N} \hat{e}_n \hat{e}_{n-k}}{\sum_{n=1}^{N} \hat{e}_n^2} \tag{A32}$$

for $k = 1, 2, \ldots, m$. The test statistic Q is given by

$$Q(\hat{r}) = N(N+2) \sum_{k=1}^{m} (N-k)^{-1} \hat{r}_k^2 \tag{A33}$$

which for large N would be distributed as χ^2_{m-p} if the fitted model is appropriate. The values of Q estimated from the residuals is compared to that of the χ^2_{m-p} to check for the lack of fit of the model used.

References

Bagarinao E (2000) Reconstruction of the bifurcation structure of a dynamical system from time series data. Doctoral thesis, Osaka University

Barahona M, Poon C (1996) Detection of nonlinear dynamics in short, noisy time series. Nature 381: 215–217

Bastian J (1986) Electrolocation: behavior, anatomy and physiology. In: Bullock TH, Heiligenberg W (eds) Electroreception. John Wiley, New York, pp 577–612

Bennett M, Pappas G, Gimenez M, Nakashima Y (1967) Physiology and ultrastructure of electrotonic junctions. IV. Medullary electromotor nuclei in gymnotid fish. J Neurophysiol 30: 236–300

Bernasconi C, Koenig P (1999) On the directionality of cortical interactions studied by structural analysis of electrophysiological recordings. Biol Cybern 81: 199–210

Black-Cleworth P (1970) The role of electrical discharges in the non-reproductive social behavior of *Gymnotus carapo* (Gymnotidae, Pisces). Anim Behav Monogr 31: 1–77

Box GEP, Jenkins GM, Reinsel GC (1994) Time series analysis forecasting and control, Prentice-Hall, Englewood Cliffs, New Jersey

Bullock TH (1969) Species differences in effect of electroreceptor input on electric organ pacemakers and other aspects of behavior in electric fish. Brain Behav Evol 2: 85–118

Capurro A (1994) Novelty response in *Gymnotus carapo*: the rate of a biological oscillator as a behavioral manifestation. MSc thesis, PEDECIBA, Montevideo, Uruguay

Capurro A, Reyes-Parada M, Ardanaz JL, Silveira R, Macadar O (1994) Serotonergic control of electric organ discharge in *Gymnotus carapo*. Role of 5-HT2A/2c receptor subtypes. Comp Biochem Physiol A 109: 583–591

Capurro A, Macadar O, Perrone R, Pakdaman K (1998) Computational model of the jamming avoidance response in the electric fish *Gymnotus carapo*. Biosystems 48: 21–27

Capurro A, Pakdaman K, Perrone R, Macadar O (1999a) Analysis of the jamming avoidance response in the electric fish *Gymnotus carapo*. Biol Cybern 80: 269–283

Capurro A, Reyes P, Lorenzo D, Pakdaman K, Macadar O (1999b) The intervals between electric organ discharges as a low-distortion window to observe the behavior of the pacemaker. Proceedings of the 3rd Int. Workshop on Neural Coding 99, Osaka, Japan

Capurro A (1999) Experimental characterization and development of statistical models of the spontaneous discharge and the jamming avoidance response in the electric fish *Gymnotus carapo*. PhD thesis, PEDECIBA, Montevideo, Uruguay

Caputi A (1999) The electric organ discharge in pulse type gymnotiforms: the transformation of a simple impulse into a complex spatio-temporal electromotor pattern. J Exp Biol 202: 1229–1241

Casdagli MC (1992) Chaos and deterministic versus stochastic nonlinear modeling. J Roy Stat Soc B 54: 303–328

Conover WJ (1980) Practical nonparametric statistics. John Wiley, New York

Ding M, Bressler S, Yang W, Liang H (2000) Short-window spectral analysis of cortical event-related potentials by adaptive multivariate autoregressive modeling: data preprocessing, model validation, and variablity assessment. Biol Cybern 83: 35–45

Dye J (1991) Ionic and synaptic mechanisms underlying a brainstem oscillator: An in vitro study of the pacemaker nucleus of Apteronotus. J Comp Physiol A 168: 521–532

Falconi A, Lorenzo D, Curti S, Morales F, Borde M (1997) Mauthner cell-evoked synaptic actions on pacemaker medullary neurons of a weakly electric fish. J Comp Physiol A 181: 143–151

Farmer JD, Sidorowich JJ (1987) Predicting chaotic time series. Phys Rev Lett 59: 845–848

Franaszczuk P, Blinowska K, Kowalczyk M (1985) The application of parametric multichannel spectral estimates in the study of electrical brain activity. Biol Cybern 51: 239–247

Franaszczuk P, Bergey G (1999) An autoregressive method for the measurement of synchronization of interictal and ictal EEG signals. Biol Cybern 81: 3–9

Fukunishi K, Murai N (1995) Temporal coding in the guinea-pig auditory cortex as revealed by optical imaging and its pattern-time-series analysis. Biol Cybern 72: 463–473

Gersh W (1970) Spectral analysis of EEG's by autoregressive decomposition of time series. Math Biosci 7: 205–222

Giszter SF, Koreisha SG, Franklin RF (1984) A vector autoregressive moving average time series approach for describing asymmetries of antennal control of two millipede species. J Math Biol 19: 281–302

Glass L, Mackey MC (1988) From Clocks to Chaos. The Rhythms of Life. Princeton University Press, New Jersey

Grassberger P, Schreiber T, Schaffrath C (1991) Non-linear time sequence analysis. Int J Bifurc Chaos 1: 521–547

Hagedorn M (1986) The ecology, courtship, and mating of Gymnotiform Fish. In: Bullock TH, Heiligenberg W (eds) Electroreception. John Wiley, New York, pp 497–526

Hayashi H, Ishizuka S (1992) Chaotic nature of bursting discharge in the *Onchidium* pacemaker neuron. J Theor Biol 156: 269–291

Heiligenberg W (1991) Neural nets in electric fish. MIT Press, Cambridge Mass, pp 151

Kay SM (1988) Modern spectral estimation. Theory and application. Prentice-Hall, Englewood Cliffs, New Jersey

Kennel MB, Brown R, Abarbanel HDI (1992) Determining embedding dimension for phase-space reconstruction using a geometrical construction. Phys Rev A 45: 3403–3411

Korenberg MJ (1988) Identifying nonlinear difference equation and functional expansion representations: the fast orthogonal algorithm. Ann Biomed Eng 16: 123–142

Kramer B (1990) Electrocommunication in teleost fishes. Springer, Berlin, Heidelberg, New York

Lánský P, Sato S (1999) The stochastic diffusion models of nerve membrane depolarization and interspike interval generation. J Peripher Nerv Syst 4: 27–42

Ljung GM, Box GEP (1978) On a measure of lack of fit in time series models. Biometrika 65: 297–303

Longtin A (1993) Nonlinear forecasting of spike trains from sensory neurons. Int J Bifurc Chaos 3: 651–661

Longtin A, Racicot DM (1997a) Spike train patterning and forecastability. Biosystems 40: 111–118

Lorenzo D (1990) Characteristics of the electrogenerator system of the fish *Gymnotus carapo*. PhD thesis, PEDECIBA, Montevideo, Uruguay

Moortgat KT, Keller CH, Bullock TH, Sejnowski TJ (1998) Submicrosecond pacemaker precision is behaviorally modulated: the gymnotiform electromotor pathway. Proc Natl Acad Sci USA 95: 4684–4689

Russell CJ, Myers JP, Bell CC (1974) The echo response in *Gnathonemus petersii* (Mormyridae). J Comp Physiol 92: 181–200

Sauer T (1994) Reconstruction of dynamical systems from interspike intervals. Phys Rev Lett 72: 3811–3814

Schiff SJ, Jerger K, Chang T, Sauer T, Aitken PG (1994) Stochastic versus deterministic variability in simple neuronal circuits: II. Hippocampal slice. Biophys J 67: 684–691

Segundo JP, Sugihara G, Dixon P, Stiber M, Bersier LF (1998) The spike trains of inhibited pacemaker neurons seen through the magnifying glass of nonlinear analyses. Neurosci 87: 741–766

Theiler J, Galdrikian B, Longtin A, Eubank S, Farmer JD (1992) Testing for nonlinearity in time series: the method of surrogate data. Physica D 58: 77–94

Thomas G, Plu G, Thalabard JC (1992) Identification of pulses in hormone time series using outlier detection methods. Stat Med 11: 2133–2145

Westby GWM (1975) Has the latency-dependent response of *Gymnotus carapo* to discharge triggered stimuli a bearing on electric fish communication? J Comp Physiol 96: 307–341