Journal of Clinical Neurophysiology
18(3):269–274, Lippincott Williams & Wilkins, Inc., Philadelphia

Observations on the Application of the Correlation Dimension and Correlation Integral to the Prediction of Seizures

*†Ivan Osorio, *Mary Ann F. Harrison, ‡Ying-Cheng Lai, and *Mark G. Frei

**Flint Hills Scientific L.L.C., Lawrence, Kansas; †Department of Neurology, University of Kansas Medical Center, Kansas City, Kansas; and ‡Departments of Mathematics, Electrical Engineering, and Physics, and Center for Systems Science and Engineering Research, Arizona State University, Tempe, Arizona, U.S.A.*

Summary: The authors reexamine the correlation integral and the related correlation dimension in the context of EEG analysis with application to seizure prediction. They identify dependencies of the correlation integral and the correlation dimension on frequency and amplitude of the signal, which may result in a reinterpretation of the dynamic importance of these measures and may cast doubts on their predictive abilities for certain classes of seizures. The relevance, for clinical and research purposes, of the distinction between retrospective and prospective inference (prediction) is addressed briefly. The authors point to the need for further research, consisting of long time series, containing multiple seizures, and for the development of objective prediction criteria. **Key Words:** Seizure—Dependencies—Correlation dimension—Amplitude—Frequency.

The popularization and, with it, the inevitable oversimplification of chaos theory (Ruelle, 1991), a branch of nonlinear dynamics, has resulted in the indiscriminate application of its tools of analysis to the study of any system with behavior that appears aperiodic and unstable. Its immense appeal to those devoted to the study of complex systems, and especially to those interested in biologic phenomena, lies in its predictive application. Specifically, chaos holds great allure to epileptologists, because the aperiodic and unstable behavior of the epileptic human brain seems ideally suited to investigation by tools that would allow precise tracking of its temporal evolution. For the first time in the history of epilepsy, a systematic and readily accessible way to do away with seizure unpredictability apparently became available. It is therefore not only understandable, but also highly desirable, that nonlinear dynamics techniques have been sanctioned by epileptologists and applied to the human electrocorticogram (ECoG). Materialization of the predictive potential of these techniques would open the possibility for both understanding and controlling the dynamics of this disease (Schiff et al., 1994b). Analyses of short time series (such as ECoG) suggest that estimation of the correlation integral (Lerner, 1996; Martinerie et al., 1998) and the correlation dimension (Elger and Lehnertz, 1998; Lehnertz and Elger, 1998) allow prediction of the electrographic onset of seizures. The intent of this communication is (1) to enhance our understanding of the type of information extracted by these methods and to help put into perspective its dynamic importance, (2) to review briefly the criteria that any method must fulfill to qualify as predictive, and (3) to stimulate much needed discussion in this important subfield of epilepsy and to assess the contribution of these methods of analysis to attaining the grand objective of enabling seizure prediction at the clinical level.

Address correspondence to Dr. I. Osorio, Flint Hills Scientific L.L.C., 5020 W 15th Street, Suite A, Lawrence, KS 66049, U.S.A.

COMMENTS ON CORRELATION DIMENSION ESTIMATION

Correlation dimension (D_2) is the most widely applied technique to the analysis of ECoG because of the avail-

ability and ease with which algorithms for its estimation can be implemented on the computer. It is also the most common basis on which the claim of chaotic dynamics has been made in biological systems (Guevara, 1997). The procedure (for a more mathematical treatment of D_2, see the appendix) requires several steps that we describe briefly here: First, the ECoG signal is broken into time windows. In each window, an m-dimensional trajectory is built from a discretely sampled ECoG signal by creating m-tuplets of points with members that are separated by a delay time τ. Once the m-dimensional trajectory is constructed, the probability that a randomly selected pair of points is separated by a distance less than or equal to a radius ϵ is computed. This probability can be estimated by the correlation integral ($C(\epsilon,m)$), which counts the fraction of point pairs within a distance ϵ. The quantity D_2 is usually estimated by looking at the linear scaling region on a logarithmic plot of the correlation integral versus ϵ (Fig. 1). The slope of this line generally increases with the embedding dimension m until the slope reaches a maximum value, which is then taken as an estimate of D_2.

When the correlation integral is applied to a finite segment of signal resulting from a low-dimensional, stationary dynamical system, it shows a dependence on embedding dimension m, delay time τ, and the number of sampled points. When this measure is applied to a moving window of nonstationary data, however, it varies from window to window as a result of both linear and nonlinear changes in the signal properties, although parameters of the correlation integral computation are held fixed. Here we caution that even simple changes in the signal amplitude or differences in the signal autocorrelation can cause large changes in the correlation integral and correlation dimension. For instance, consider two windows of ECoG data recorded from the same patient, one during the interictal state with amplitude A_1, and the other during ictus with amplitude A_2, where $A_2 > A_1$. Fig. 2 shows the signal trajectory (the presumed attractor) for two windows of interictal and ictal ECoG, where an embedding dimension of $m = 3$ was chosen to facilitate visualization. The correlation integral depends on the number of point pairs separated by at most a distance ϵ. Selecting one arbitrary point x_i, we draw a sphere of radius ϵ around the point and count the number of point pairs enclosed by the sphere, as shown in Fig. 2. We see that in the interictal window (Fig. 2A), the points are clustered into a far tighter group than in the ictal window (Fig. 2B), and that the number of points within the *fixed* sphere of radius ϵ is lower than that generated by the ictal data. This leads to a decrease in the correlation integral during the ictal state relative to the interictal state. Because the probability of a point pair being inside a ball of radius ϵ is proportional to $(\epsilon/A_1)^m$ interictally, and to $(\epsilon/A_2)^m$ ictally, the relative decrease in the correlation integral is thus on the order of $(A_1/A_2)^m$. We note

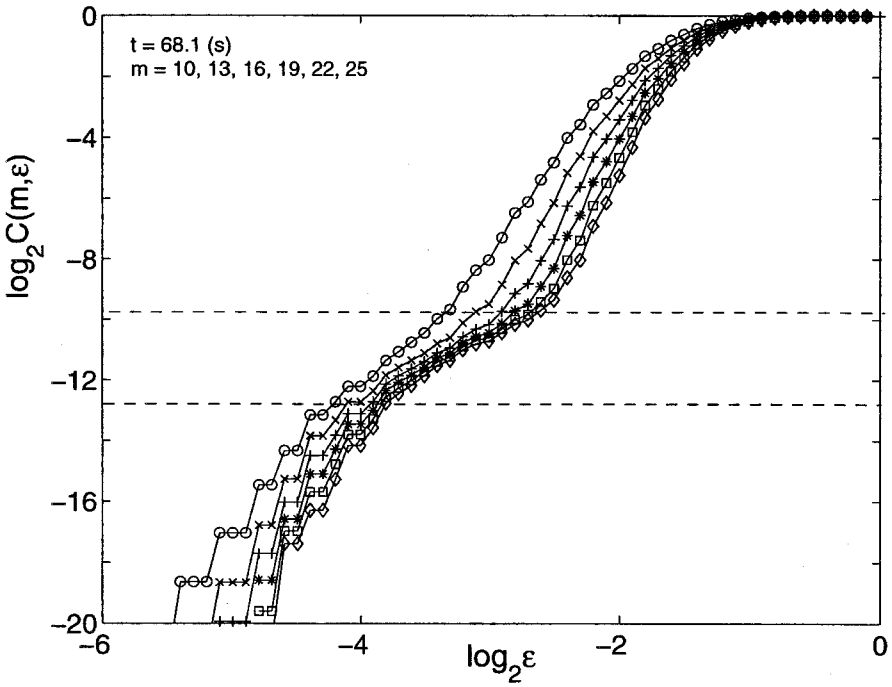

FIG. 1. For $t = 68$ seconds during the preictal phase and six values of the embedding dimension, the figure shows $\log_2 C(m, \epsilon)$ versus $\log_2 \epsilon$ on a logarithmic scale. The circles, *x*'s, pluses, stars, squares, and diamonds correspond to $m = 10$, 13, 16, 19, 22, and 25 respectively. The region between the dashed lines is the region of approximately linear scaling (the region of "anomalous slope") resulting from the autocorrelation of the data.

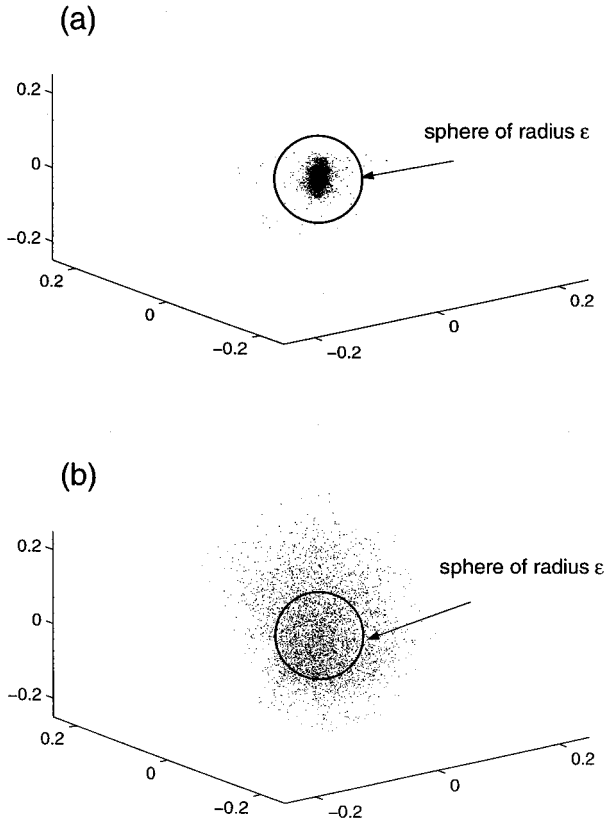

FIG. 2. (**A, B**) With embedding dimension $m = 3$ and delay time $\tau = 1/12$ second, the trajectory of the system interictally (**A**), and ictally (**B**). To compute the correlation integral, the number of points inside an m-dimensional sphere of radius ϵ is counted. A high-amplitude signal fills out a larger region of m-space, thus fewer points are contained within a sphere of radius ϵ than for a lower amplitude signal. These density differences account in part for the sensitivity of D_2 to signal normalization.

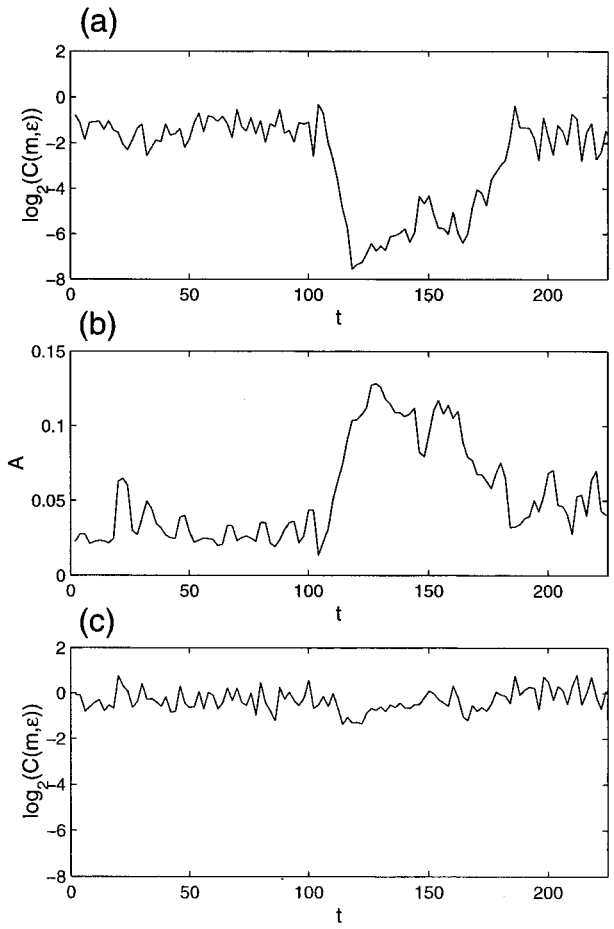

FIG. 3. For a segment of electrocorticogram containing a seizure, the following are seen: (**A**) The correlation integral with parameters $m = 7$, $\tau = 1/20$ second, and $\epsilon = 0.02$ mV. (**B**) The amplitude of the signal in each window. (**C**) The correlation integral of the signal normalized in each window ($x_{norm} = x/A$) with parameters $m = 7$, $\tau = 1/20$ second, and $\epsilon = 0.75$. The drop in D_2 parallels the increase in signal amplitude.

that this is on the same order of magnitude as the decrease reported by Lerner (1996) for the same seizure recording. These observations suggest that the correlation integral and D_2 are highly sensitive to amplitude changes—a characteristic that could result in nonspecificity, given the nonstationarity of the signal (that is, amplitude increases are not seen *only* with seizures). Therefore, it seems relevant to test the performance of these methods on an ECoG signal normalized for amplitudes. For the seizure in Fig. 2, we compute the correlation integral with embedding dimension $m = 7$, delay time $\tau = 1/20$ second, and $\epsilon = 0.020$ mV, which is approximately 75% of the median of the amplitudes in interictal windows. Here, the amplitude A is approximated for a noisy signal by dividing the original signal into windows, sorting in increasing order, and computing this amplitude as the 95th percentile minus the fifth percentile and dividing by two. The correlation integral for the signal is shown in Fig. 3A). To normalize for amplitude changes, we divide each window by its respective amplitude (computed as mentioned) and recompute the correlation integral. Fig. 3B shows the amplitude of the signal in each window, and Fig. 3 C shows the correlation integral of the signal *normalized* in each window, $x_{norm}(t) = x(t)/A$. We see that for this seizure, this normalization scheme destroys most of the discriminating ability of the correlation integral. The previously reported decreases in the correlation integral (Lerner, 1996) and the correlation dimension (Martinerie et al., 1998), in the absence of normalization schemes, may reflect (at least in part) simple amplitude variations that can be detected more efficiently through other means. In short, changes in correlation integral and D_2 of a non-

normalized signal may not be the result of changes in the nonlinear dynamic properties of the signal.

We now examine the application of D_2, the slope of the correlation integral versus ϵ on a logarithmic scale to a signal normalized for amplitude. For a stochastic system, this plot generates a line with a slope dependent on the embedding dimension *m* that never reaches a maximum value as *m* is increased because of the infinite dimensionality of the system. However, Theiler (1986) points out that for finite, autocorrelated, stochastic datasets, the plot (see Fig. 1) of the correlation integral on a logarithmic scale can exhibit approximately linear regions with slopes distinctly smaller than in other regions (we term this the *anomalous* scaling region) that are not good indicators of the dimensionality of the system. By analyzing ECoG data in a moving window of finite width, we find that the autocorrelation of the data typically varies widely among windows, and can undergo particularly abrupt changes on entering the ictal state. Changes such as these in the time–frequency distributions are used typically to score seizures visually. It is well-known that the energy spectral density of a signal is the Fourier transform of its autocorrelation sequence by the Wiener–Khintchine Theorem (Proakis and Manolakis, 1996). When a D_2 computation is made on the basis of the slopes in this linear region, changes in the autocorrelation are linked to changes in D_2 through the value of the slopes, as shown in Fig. 4. Thus changes in D_2 values *when computed in this region of anomalous slope* may be largely be the result of changes in the frequency spectra of the "windowed" signal.

This interplay between autocorrelation and D_2 is best illustrated through an example: Let us compute D_2 in a 20-second window "slid" by 2 seconds at each step, using a delay time $\tau = 1/12$ second, embedding dimensions from 10 through 25, and an amplitude normalization scheme to make the signal be between 0 and 1 in each window. Fig. 1 shows a plot of the correlation integral versus ϵ on a logarithmic scale for an interictal window. The segment between the dashed lines indicates the region of approximately linear scaling, where the slope is influenced by the autocorrelation. Above and below these regions, we see that the lines exhibit slopes that increase with embedding dimension *m*, consistent with a stochastic (or very high-dimensionally chaotic) dynamic origin. During ictus, slopes in the linear scaling region can vary widely, indicating variation in the autocorrelation and, consequently, the frequency spectrum. As shown in Fig. 4, there is correspondence between decreases in the autocorrelation and increases in the D_2 computation. This indicates that D_2 may be dependent on frequency changes—a dependency that does not bode

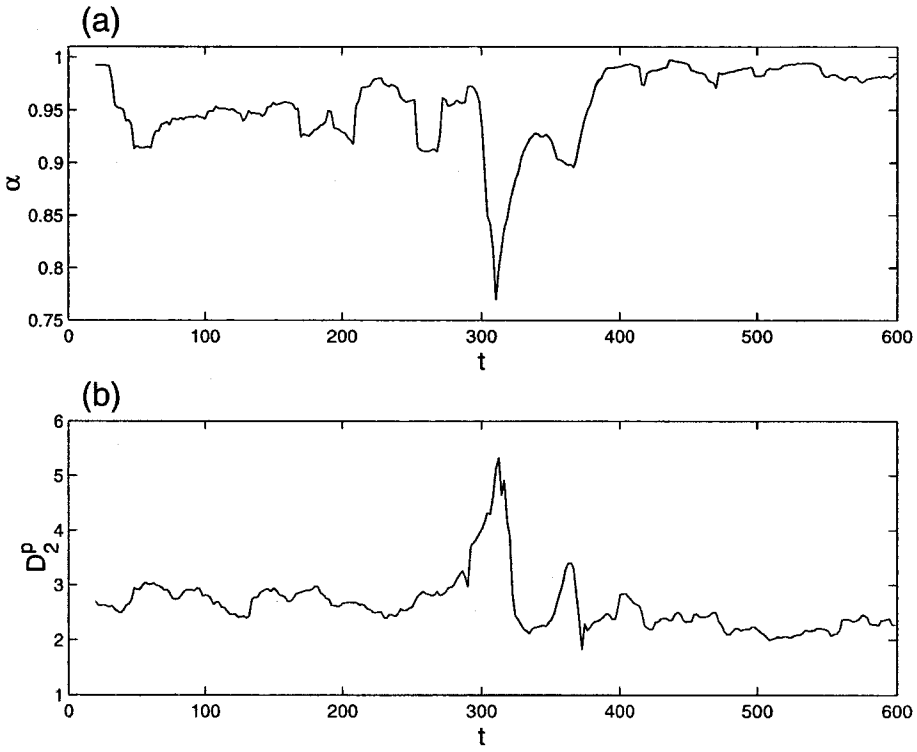

FIG. 4. (**A, B**) For a segment of electrocorticographic data containing a seizure, the time evolution of the autocorrelation α (**A**) and the correlation dimension D_2 (**B**). Notice that the decrease in autocorrelation occurs simultaneously with the increase in D_2.

well for the specificity of this method with regard to ECoG data. The extent of the interplay between D_2 and frequency spectral changes needs to be further explored.

Other measures in seizure prediction, such as the Lyapunov exponent (Iasemidis and Sackellares, 1991; Iasemidis et al., 1990, 1998), a measure of sensitive dependence on initial conditions, may also be sensitive to changes in signal amplitude and power spectrum. Although we have not applied this method to ECoG series, its limitations are similar to those of D_2: It requires large amounts of stationary data, and is sensitive to noise, m, and τ, and perhaps to fluctuations in amplitude and autocorrelation because the exponents are estimated using an attractor reconstructed in an identical manner to that used for the computation of D_2.

RETROSPECTIVE INFERENCE?

For any method to qualify as predictive (Osorio et al., 1998; Stone, 1992), the system's output at a given time t must be solely a function of the information available to that system at or before time t. In other words, the prediction that a seizure will occur at time $t + n$, should have been made earlier (at t), without using any new data between t and $t + n$. This allows for objective assessment of the predictive sensitivity and specificity (accuracy) of the method, through calculation of the prediction error. Methods in which the time series of a predictive measure is tracked backward from the electrographic onset to the first point where it crosses a "threshold" (or where "thresholds" are set retrospectively for each seizure) are examples of retrospective inference, not of true prediction.

CONCLUDING COMMENTS

Our findings indicate that correlation integral and D_2 estimates of ECoG are sensitive to changes in amplitude and, through the autocorrelation, to variation in the frequency spectrum of the signal. In particular, we contrast our approach with that proposed by Lerner (1998) and adopted by Martinerie et al. (1998). The decrease of the correlation density measure observed by Lerner (1996) has a very simple dynamic explanation and it reveals no more information about seizures than expert visual inspection of the ECoG. The existing results that the correlation dimension (Martinerie et al., 1998) or the correlation density (Lerner, 1996) decrease to a lower value during the seizure may not reflect correctly the underlying dynamics of the seizure. Taken one step further, our observations may be used in support of the claim that the information these methods provide about nonstationary systems' dynamics, in this case the brain's, can be at least in part a reflection of changes in the amplitude or autocorrelation of the signal. This observation, although possibly trivial or even obvious to many, has not been emphasized in the epilepsy literature. The quantity computed by application of the correlation dimension to the ECoG has been believed in the past to provide information about the dimension of the system, giving an estimate of the number of degrees of freedom necessary to specify its trajectory. Recently, as computational problems with estimating dimension on finite, nonstationary data have been recognized, most researchers have chosen to use D_2 as a discriminating measure of the changes in the measure between windows.

The nature or properties of the signal that result in increases or decreases in this relative D_2 remain obscure. For certain classes of seizures, we suggest that the information provided by digital signal processing methods, which are well-suited for time–frequency analysis (D'Attellis et al., 1997; Gotman and Gloor, 1976; Osorio et al., 1998; Principe et al., 1991; Schiff et al., 1994a, c), will be highly similar qualitatively to that obtained by applying D_2, but much faster. This statement does not imply that amplitude and frequency are the only or even the main signal features detectable by the correlation integral and the correlation dimension.

In our view, the top research priorities for the near future are (1) large-scale studies consisting of long time series containing multiple seizures; (2) the selection and implementation of criteria to determine, in a reproducible and objective manner, the time of seizure onset prediction; and (3) the development of algorithms that operate in real time. The diligence and scientific rigor with which these objectives are pursued, will determine the probability of success.

Acknowledgment: The work is supported in part by a grant from the Alliance for Epilepsy Research. Dr. Lai is grateful for the hospitality and stimulating discussions at Flint Hills Scientific, where this work was initiated.

APPENDIX

To compute the correlation integral ($C(m, \epsilon)$) and the correlation dimension (D_2), we first delay–coordinate embed the signal:

$$x(t) = \{x(t), x(t + \tau), \ldots, x(t + (m - 1)\tau)\}, \qquad (1)$$

where τ is the delay time, and m is the embedding dimension. Grassberger and Procaccia (1983) show that D_2 can be evaluated using $C(m, \epsilon)$, which is defined to be

the probability that a randomly selected pair of delay–coordinate points is separated by a distance less than ϵ. We let N be the number of points in the reconstructed vector time series x(t). The correlation integral can be approximated by the following sum for large N:

$$C(m,\epsilon) \approx \frac{2}{N(N-1)} \sum_{j=1}^{N} \sum_{i=j+1}^{N} \theta(\epsilon - |x_i - x_j|), \quad (2)$$

where $\theta(\cdot)$ is the Heaviside function given by $\theta(x) = 1$ for $x \geq 0$ and 0 otherwise, and $|x_i - x_j|$ is the distance between points x_i and x_j. The correlation dimension is then given by the following (Grassberger and Procaccia, 1983):

$$D_2 \approx \lim_{\epsilon \to 0} \lim_{N \to \infty} \frac{\log C(m,\epsilon)}{\log \epsilon}. \quad (3)$$

The autocorrelation is computed through the following average of each window of the signal:

$$\alpha = \frac{1}{M} \sum_{j=1}^{M} (\alpha_j)^{1/j}, \quad (4)$$

where we use $M = 6$ and

$$\alpha_k = \frac{\langle x_i x_{i+j} \rangle}{\langle x_i^2 \rangle} \quad (5)$$

REFERENCES

D'Attellis CED, Isaacson SI, Sirne RO. Detection of epileptic events in ECoG using wavelet analysis. *Ann Biomed Eng* 1997;25:308–15.

Elger CE, Lehnertz K. Seizure prediction by nonlinear time series analysis of brain electrical activity. *Eur J Neurosci* 1998;10: 786–9.

Grassberger P, Procaccia I. Measuring the strangeness of strange attractors. *Physica D* 1983;9:189–208.

Gotman J, Gloor P. Automatic recognition and quantification of interictal epileptic activity in human scalp EEG. *Electroencephalogr Clin Neurophysiol* 1976;41:513–29.

Guevara MR. Chaos in electrophysiology. In: Billette J, LeBlanc A-R, eds. *Concepts and techniques in bioelectric measurements: is the medium carrying the message*? Montréal: Éditions de l'École Polytechnique, 1997:67–87.

Iasemidis LD, Sackellares JC. The evolution with time of the spatial distribution of the largest Lyapunov exponent on the human epileptic cortex. In: Duke DW, Pritchard WS, eds. *Measuring chaos in the human brain*. Singapore: World Scientific, 1991:49–82.

Iasemidis LD, Sackellares JC, Gilmore RL, Roper SN. Automated seizure prediction paradigm [abstract]. *Epilepsia* 1998;39(S6):207.

Iasemidis LD, Sackellares JC, Zaveri HP, Williams WJ. Phase space topography and the Lyapunov exponent of electrocorticograms in partial seizures. *Brain Topography* 1990;2:187–201.

Lehnertz K, Elger CE. Can epileptic seizures be predicted? Evidence from nonlinear time series analysis of brain electrical activity. *Phys Rev Lett* 1998;80:5019–22.

Lerner DE. Monitoring changing dynamics with correlation integrals: case study of an epileptic seizure. *Physica D* 1996;97:563–76.

Martinerie J, Adam C, Le Van Quyen M, et al. Epileptic seizures can be anticipated by non-linear analysis. *Nat Med* 1998;4:1173–6.

Osorio I, Frei MG, Wilkinson SB. Real-time automated detection and quantitative analysis of seizures and short-term prediction of clinical onset. *Epilepsia* 1998;39:615–27.

Principe J, Hsu H, Reid SA. Construction and visualization of ECoG state space portraits. In: Duke DW, Pritchard WS, eds. *Measuring chaos in the human brain*. Singapore: World Scientific, 1991:167–80.

Proakis JG, Manolakis DG. *Digital signal processing: Principles, algorithms, and applications*. Upper Saddle River, NJ: Prentice Hall, 1996:299.

Ruelle D. *Chance and chaos*. Princeton: Princeton University Press, 1991:66–72.

Schiff SJ, Aldroubi A, Unser M, Sato S. Fast wavelet transform of the ECoG. *Electroencephalogr Clin Neurophysiol* 1994*a*;91:442–55.

Schiff SJ, Jerger K, Duong DH, Chang T, Spano ML, Ditto WL. Controlling chaos in the brain. *Nature* 1994*b*;370:615–20.

Schiff SJ, Milton J, Heller H, Weinstein S. Wavelet transform and surrogate data for electroencephalographic spike and seizure localization. *Opt Eng* 1994*c*;33:2162–9.

Stone M. Chaos, prediction, and Laplacian determinism. *Am Philos Q* 1992;26:123–31.

Theiler J. Spurious dimension from correlation algorithms applied to limited time-series data. *Phys Rev A* 1986;34:2427–32.