Biophysical Journal Volume 69 November 1995 1748–1757

Predictability of EEG Interictal Spikes

David A. Scott and Steven J. Schiff
Department of Neurosurgery, Children's National Medical Center, and The George Washington University School of Medicine, Washington, DC 20010 USA

ABSTRACT To determine whether EEG spikes are predictable, time series of EEG spike intervals were generated from subdural and depth electrode recordings from four patients. The intervals between EEG spikes were hand edited to ensure high accuracy and eliminate false positive and negative spikes. Spike rates (per minute) were generated from longer time series, but for these data hand editing was usually not feasible. Linear and nonlinear models were fit to both types of data. One patient had no linear or nonlinear predictability, two had predictability that could be well accounted for with a linear stochastic model, and one had a degree of nonlinear predictability for both interval and rate data that no linear model could adequately account for.

INTRODUCTION

The relationship between interictal spikes and seizures remains unclear. Several recent studies have inquired whether the rate of EEG spike generation changes in a consistent manner before a seizure (Gotman, 1991; Katz et al., 1991), but no such patterns were identified. These studies employed automated spike detection systems, although one study (Katz et al., 1991) employed hand editing of detected spikes to eliminate false positive detections. The issue is significant, because if spikes predict seizures, then the prospect of an automated system to warn of impending seizure onset, or perhaps to suppress seizure generation, might be feasible.

There is a growing interest in whether neuronal ensembles generate behavior that is predictable (Chang et al., 1994; Schiff et al., 1994a). In addition, new techniques now exist that are capable of exerting control over complex and chaotic systems (Ditto et al., 1990; Garfinkel et al., 1992), and such strategies have recently been extended to a model epileptic focus (Schiff et al., 1994b).

Because in principle what is predictable is controllable, we sought to examine interictal spike patterns. Although our long-term goal is to establish whether interictal spikes predict seizures, we sought to begin with a detailed study of the predictability of the interictal spikes themselves. Although linear time series methods are well established for such data (Box and Jenkins, 1976), we felt it useful to ask whether there were patterns among these spikes that would require nonlinear methods to model. Evidence for nonlinear predictability would be necessary but insufficient evidence that these patterns might be chaotic (Chang et al., 1994). To be sure that even subtle inaccuracies in the time series would not mask predictability, especially with the nonlinear methods, we prepared sets of data where all traces were hand edited.

METHODS

Data were collected from patients undergoing routine evaluation for epilepsy surgery that required implanted subdural or depth electrodes for medical purposes unrelated to this study. Institutional Review Board approval from the Children's National Medical Center was obtained for this research.

Data were recorded by interposing a bioelectric isolation unit (model IMEB2, Grass Corporation) between the patient and preamplifiers (model 12A5, Grass Corporation), using a 100 Hz anti-aliasing filter, and digitizing across 12 bits at 200 Hz. These data were recorded directly on a personal computer utilizing Axotape 2.0 (Axon Instruments). Interictal spikes were identified with Datapac 2.0 (Run Technologies), using routine signal processing methods to preliminarily identify the majority of these spikes (high-pass filtering, first derivatives, rectification, etc.). These time series were then hand edited to remove false positive and false negative spikes and to be sure that all clear spikes were accurately resolved for this study. We attempted to collect at least 1 h of data just before the onset of a seizure in these patients. Because of the length of the data sets sought, controlling for state of consciousness was not practical; the segments controlled for state tended to be quite short, often only 5–10 min, with correspondingly fewer spikes. Because spurious results are anticipated with the use of the nonlinear methods with small data sets, the analysis to follow will utilize long time series without control for state.

Linear time series modeling was performed using a generalized Auto Regressive Integrated Moving Average (ARIMA) model (Box and Jenkins, 1976), implemented using StatGraphics Plus (Manugistics, Inc.) or Matlab (The Mathworks, Inc.). Details of the ARIMA modeling process are presented in the Appendix. ARIMA modeling is a method of performing linear prediction that assumes that a given time series is a function of its own past behavior (autoregressive, AR), the convolution of a given filter with a series of Gaussian distributed shocks (moving average, MA), and if nonstationary, a difference term (Integrated, I). These models take a linear model of a time series and incorporate a random term so that the end result is a linear stochastic model. We attempt to achieve a "best" linear fit of our time series, although we limit the number of moving average and autoregressive coefficients that we will use to characterize the systems (the limit we will typically incorporate in this study is 9).

Time series may not be well fit by a linear model, and we have done extensive testing to determine which of currently available methods are best to identify nonlinear determinism for biological or theoretical time series (Chang et al., 1995). We choose for this study a nonlinear predictor that is "zero order," which means that it fits local constant linear maps to a time series that has been "embedded" into a phase space with the method

Received for publication 18 April 1995 and in final form 26 July 1995.

Address reprint requests to Steven J. Schiff, M.D., Ph.D., Department of Neurosurgery, Children's National Medical Center, 111 Michigan Avenue NW, Washington, DC 20010. Tel: 202-884-3020; Fax: 202-884-3091; E-mail: sjs@gwuvm.gwu.edu.

0006-3495/95/11/1748/00 $2.00

of delay coordinate embedding. The embedding method derives from a theory that states that the dynamics of a system of several variables may be reconstructed using coordinates derived from measurements of a single variable at multiple points in time (Takens, 1981). That delay coordinate embedding of inter-event intervals (such as EEG spikes) may also be used to reconstruct the dynamics of a system was recently demonstrated by Sauer (1994). The zero-order nonlinear predictor that we use is a rather simple implementation of a method that can be extended to higher-order predictors (Farmer and Sidorowich, 1988). We discuss the details of the mathematics of this nonlinear predictor in the Appendix. These algorithms were implemented with Matlab (The Mathworks, Inc.).

The delay coordinate embedding method and nonlinear predictor just described can identify both linear and nonlinear predictability in a time series. To discriminate between linear and nonlinear predictability, data are compared to randomized surrogate data that possess linear properties similar to those of the original data (i.e., autocorrelation, power spectrum). A common feature of the surrogate methods we employ involves taking the Fourier transform of data and adding a random phase, and then performing an inverse transform to regenerate a random time series (Theiler et al., 1992). Used alone, this method alone is called phase randomization. This method assumes the data are produced by a Gaussian random process, akin to the ARIMA model, and that the data reflect a Gaussian distribution. One alternative, called the Gaussian scaled method, assumes that any such Gaussian random process is modified by a static nonlinear transformation (Theiler et al., 1992). In this method the data are scaled to a normal distribution before phase randomization, and the rank order of these phase-randomized data is used to rank order the original data set to generate the surrogate data. Another surrogate method that seeks to preserve the original distribution of signal amplitudes is one where the data produced by phase randomization are scaled using the rank ordered values of the original amplitude distribution. We call this latter method Fourier shuffling. A complete description of these methods can be found in Chang et al. (1994), and source code listings in Matlab to generate these surrogates can be found in Schiff et al. (1994c).

Because the length of time that data can be fully hand edited is limited, we also attempted to use spike rates in very long time series (several hours), for which hand editing was generally impractical. For these data, we defined the sensitivity of detection, and the false positive and false negative detection rates, by comparing the portions of the automatically detected spikes that overlapped with the corresponding hand-edited data. The above linear and nonlinear prediction techniques were applied to these rate data.

Although no general method exists to directly compare the ability of the linear and nonlinear methods to predict these time series, we attempted the following comparison. The linear model was an AR model determined by autoregression on the first 400 data points. All possible AR models AR(1) to AR(199) were considered. The model with the lowest prediction error sums of squares (PRESS) was the model used to predict the next 100 points. The PRESS criterion is a method based on deleted residuals, or the ability of the model to predict any one of the points within the data, after that data point has been removed from consideration in developing the model (Neter et al., 1990), and was implemented with Matlab. Points were predicted one step into the future using the immediately preceding data (p points before the predicted point for an AR model of order p). We compared the mean squared error for these 100 linear predictions with the nonlinear predictions of the same points using again only data preceding each predicted point. These findings were contrasted with the results from the above methods.

RESULTS

Fig. 1 illustrates segments of raw data from four patients. Although there is a great deal of variation in the morphology of these spikes, all of them tend to be characterized by a spike followed by a slow wave, and some of the spikes are polymorphic. In each of the panels in this figure, a heavy black bar is placed above the clear spikes. The black bar defines the onset and offset of the spikes, and we will use this to define a region where the spike occurs to compare with automated methods later in the paper. Note that for RO, there is an open box above a small spike and wave. This particular patient had a tracing that was difficult to interpret because there were large clear spikes in addition to smaller, less well defined spikes. To deal with this, we have analyzed it in two ways, both including and excluding the small spikes with a cut-off of 0.15 mV peak to peak. We will designate RO1 as the file with only the large spikes, and RO2 as the file with both large and small spikes included.

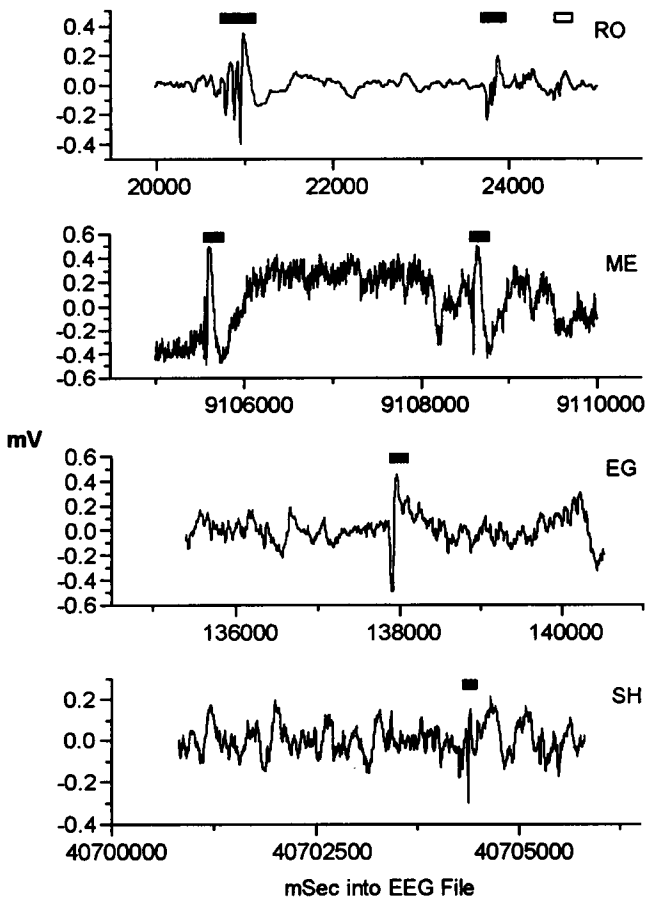

FIGURE 1 Representative raw EEG samples from the four subjects: RO, ME, EG, SH. The bars indicate the detected spikes. For subject RO a second, more sensitive analysis was done that included smaller spikes (<0.15 mV), as shown by the open bar.

Fig. 2 illustrates the interspike intervals for these data. Fig. 3 shows the autocorrelations for the interspike intervals from these spike series. Autocorrelation is a measure of the amount of linear correlation in a series of events. Here the events are interictal spikes, and the autocorrelation measures the linear correlation between the time intervals between these spikes. The dotted lines in the figures indicate confidence limits within which the linear correlation is not significantly different from zero (i.e., uncorrelated or white noise, $\alpha = 0.05$). A good review of autocorrelation can be

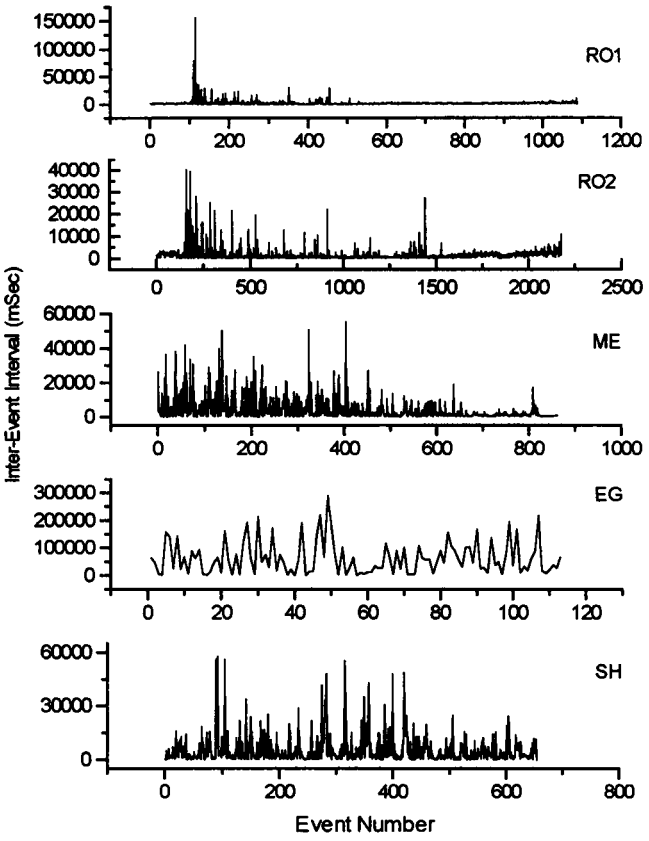

FIGURE 2 Intervals between interictal spikes, plotted as a function of the event number.

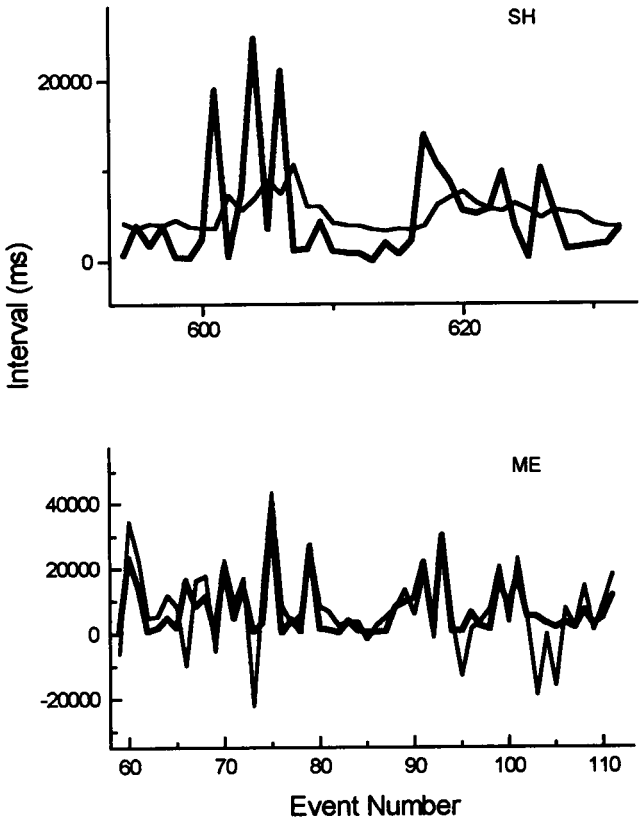

FIGURE 4 Detail of short segment of linear model (*light lines*) of interspike intervals (*heavy lines*). The model fit for SH is an AR(3) on undifferenced data (3, 0, 0). For ME the fit is a MA(1) on first differenced data (0, 1, 1). See Appendix for further details on methodology and terms.

found in Box and Jenkins (1976). Note in this figure that both RO1 and RO2 have substantial linear correlation, as does SH. Less linear correlation is seen in ME, and no correlation is identified in EG.

Fig. 4 illustrates segments of the ARIMA model fits to time series from SH and ME. For both of these time series,

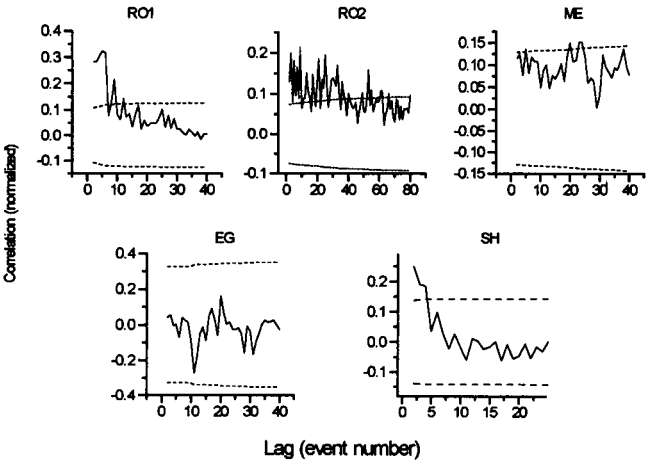

FIGURE 3 Autocorrelation of interspike intervals as shown in Fig. 2. Note significant linear correlation in subjects RO and SH, and a lesser amount in ME.

with a relatively small number of parameters, we were able to fit an ARIMA model (light lines) that was a good linear fit to the original data (heavy lines) in that a χ^2 statistic on the first 20 residual autocorrelations was not significantly different from white noise ($\alpha = 0.05$). We were not able to obtain a parsimonious (<10 MA or AR coefficients) linear fit to any of the data sets from RO (RO1 or RO2) that could remove all significant residual autocorrelation. Because patient EG had no significant linear correlation, no attempt at linear time series prediction could be made for this individual.

Table 1 illustrates the statistics involved in providing an adequate linear fit. Shown are the statistics for the models for both SH and ME. In the table, parameters refer to the designation of AR or MA coefficients, the estimate of the parameters reflects the values of the AR or MA parameters that yield the best fit of the data, and the T-value is a statistic that reflects the likelihood that the given estimate is significantly different from zero. The probability of making an error when assessing statistical significance to the estimate is given as p. The next column gives the difference term (Diff.), if any (the integrated component of the ARIMA model). Also estimated by this method is the variance of the driving shocks or white noise variance (σ) and degrees of freedom (d.f.). To show the adequacy of the linear stochastic model we examined the residual autocorrelations, com-

TABLE 1 Statistics for ARIMA models—individual spikes

							White noise*		First 20 residuals‡	
Subject	Parameter	Estimate	SE	T-value	p	Diff.	σ^2	d.f	χ^2	P(wn)
ME (a)	MA (1)	0.90013	0.01663	54.128	0.00000	1	5.7423E7	729	16.1852	0.579626
	Mean	−22.170	30.507	−0.72671	0.46764					
	Constant	−22.170								
ME (b)	MA (1)	1.5145	0.02987	50.708	0.00000	2	8.2768E7	727	13.6738	0.690041
	MA (2)	−0.60921	0.03027	−20.126	0.00000					
	Mean	16.841	31.496	0.53470	0.59302					
	Constant	16.841								
SH (a)	AR (1)	0.20036	0.03894	5.1458	0.00000	0	6.0551E7	651	14.82	0.537857
	AR (2)	0.11033	0.03951	2.7923	0.00539					
	AR (3)	0.11904	0.03895	3.0558	0.00234					
	Mean	5462.1	531.70	10.272	0.00000					
	Constant	3114.9								
SH (b)	MA (1)	−0.21865	0.03859	−5.6654	0.00000	0	6.0664E7	651	15.6479	0.477793
	MA (2)	−0.14317	0.03917	−3.6551	0.00028					
	MA (3)	−0.16349	0.03870	−4.2248	0.00003					
	Mean	5464.4	473.88	11.5313	0.00000					
	Constant	5464.4								

*Estimate of the white noise variance for the linear model, with the given number of degrees of freedom.
‡χ^2 test on the first 20 residual autocorrelation terms. P(wn) represents the probability of having a larger χ^2 value for the residual autocorrelations given white noise. We ask, what is the probability of making an error in saying that the ARIMA fit does not adequately explain the data? The relatively large P(wn) values indicate that the residuals are consistent with white noise, and the model therefore fits the data.

paring the first 20 autocorrelations estimated to white noise with a χ^2 statistic. The statistic P(wn) reflects the probability of making a mistake in saying that the residual autocorrelation is larger than would be expected for white noise; a large value of P(wn) (P(wn) > 0.05) implies an adequate fit of the data. As can be seen in Table 1, for two patients we were able to identify models (two models for each patient) that provided good fits with few coefficients. Such low-order fits are termed "parsimonious" fits by Box and Jenkins (1976).

Next, we examine whether a nonlinear model provides any advantages over linear modeling. First we perform a delay coordinate embedding of these spike series. Parameters used for this embedding were dimension of 4, and a lag was determined by the decorrelation lag of the series (the point where the autocorrelation is no longer significantly different from zero). For RO1 this lag was 6; for RO2, 34; for SH, 4; and for ME, 9. Nineteen surrogates for each of these data sets were chosen so that we would make few unnecessary assumptions about the distribution of the surrogates; these 19 surrogates permit the use of a simplified one-tailed Monte Carlo significance test, where if prediction error of the data is less than all surrogate results, then the null hypothesis that the data are linearly stochastic can be rejected at the 0.05 significance level (Hope, 1968). Rejecting this null hypothesis implies that a nonlinear model is required for at least a portion of these data. We examined three different types of surrogates, and the surrogates shown in Fig. 5 are the best fitting of the surrogates: for RO1 it was the Fourier shuffled surrogate, for RO_2 Gaussian scaled, for SH Fourier shuffled, for ME Gaussian scaled, and for EG Gaussian scaled. If any of the prediction errors from the surrogate data are smaller than the experimental data, then we can assume that a linearly stochastic model can account well for these data. Because chaotic time series display an exponential loss in predictability as time evolves, we focus on whether we can predict just a few events ahead; indeed, such short-term prediction can be seen for ME.

As an alternative to looking at the interval between individual spikes, we attempted to look at the rate of spike generation in terms of the number of spikes per minute, similar to previous EEG spike data sets used for rate anal-

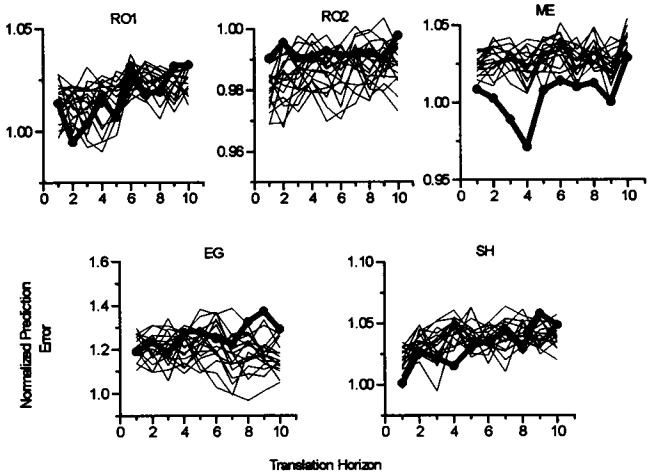

FIGURE 5 Results of nonlinear prediction for spike intervals shown in Fig. 2. Heavy black lines indicate experimental data, and 19 surrogate data set results are plotted with thin black lines. The embedding dimension is 4 and the lag is determined by the decorrelation time. Some nonlinear predictability appears to be present for ME. Further details in text

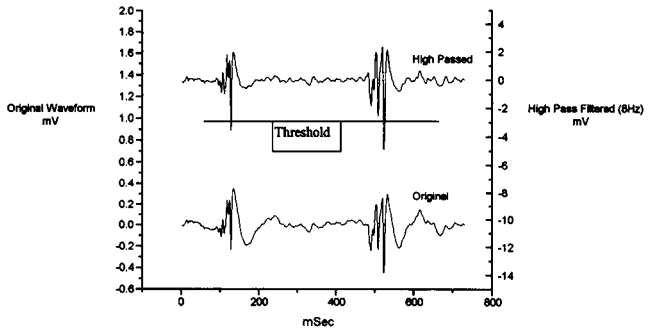

FIGURE 6 Illustration of preparation of data for automatically detecting EEG spikes from RO. After high-pass filtering, the spikes in this data set could be reasonably discriminated with a negative going threshold detector. Statistics for this and the other data sets are in Table 2.

ysis (Katz et al., 1991; Gotman, 1991). We used automated spike discrimination for these data, customized to each data set. The attempt here was not to definitively detect all of the spikes within the signal but to produce a measure highly correlated to the rate of spiking. For the particular signals we obtained this was possible to do using simple signal processing techniques such as high-pass filtering and threshold detection. To illustrate, Fig. 6 shows the effect on data from patient RO, obtained by using a high-pass filter, which generates a signal from which the spikes from this patient could be discriminated with high accuracy with a negative going threshold detector (each patient required different signal processing). For the region of overlap, we compared the accuracy of the automatically discriminated spikes with the hand-edited data (summary given in Table 2). False positives or false negatives were detected when there was no coincidence between a manually detected and an automatically detected spike. Because RO1 contains fewer and larger spikes than RO2, the sensitivity of the analysis applied to the larger data set, RO2, was greater than that for RO1; for the less sensitive data, RO1, there are fewer false negatives, but for the more sensitive data, RO2, there are fewer false positives. For the data set EG, spikes were relatively infrequent, and only for this data set could spikes per minute be calculated from hand-edited data.

Fig. 7 illustrates the measured rate of spiking for these data. The autocorrelations for these data are shown in Fig. 8, which displays the high degree of linear correlation in these rate data, with the exception of patient EG. It is important to realize that the spike rate was not determined by a sliding-

TABLE 2 Automatically discriminated versus hand-edited spikes

Subject	No. points, manual	No. points, automatic	False positive	False negative
RO1	1088	1473	379	7
RO2	2175	1473	5	633
ME	861	790	61	162
EG	113	113	All Manual	
SH	655	645	16	32

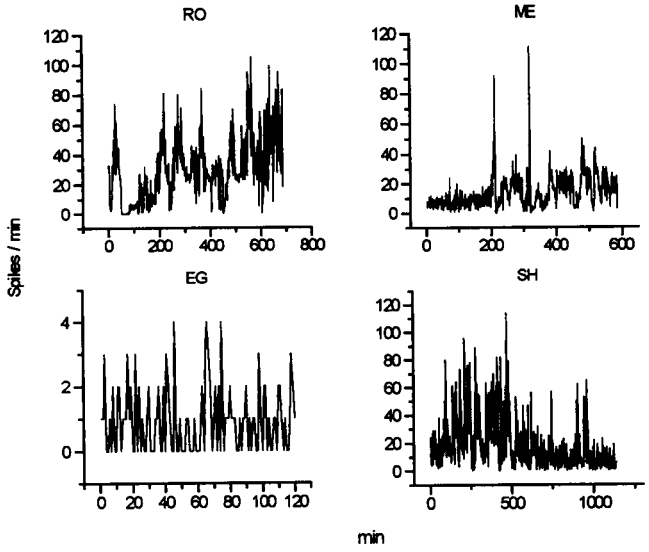

FIGURE 7 Rates of interictal spiking, as measured using automated algorithms.

window averaging method, which would introduce a moving average type autocorrelation to the data, but instead was created by partitioning the data into discrete 1-min bins.

Segments of the linear fits for the spikes per minute data are shown in Fig. 9. The three data sets for which there was substantial linear correlation in Fig. 8 can each be modeled with parsimonious linear models, and the statistics showing the adequacy of these models are shown in Table 3.

We again tried to identify evidence of nonlinear predictability that could not be accounted for with a linear model. Fig. 10 shows that such a model was unnecessary except for ME, where for 4 min ahead there was a component of predictability that was better explained with a nonlinear model than for any linear model that we could find. The embedding dimension for these tests was 4. The lag was

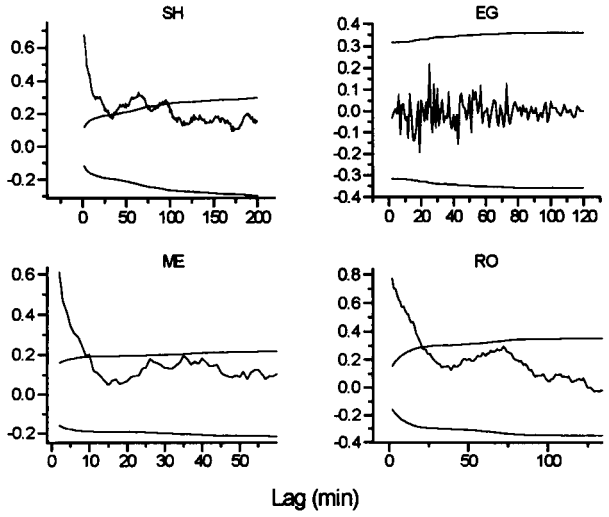

FIGURE 8 Autocorrelation for the spike rates shown in Fig. 7. Note significant linear correlation for all subjects except EG.

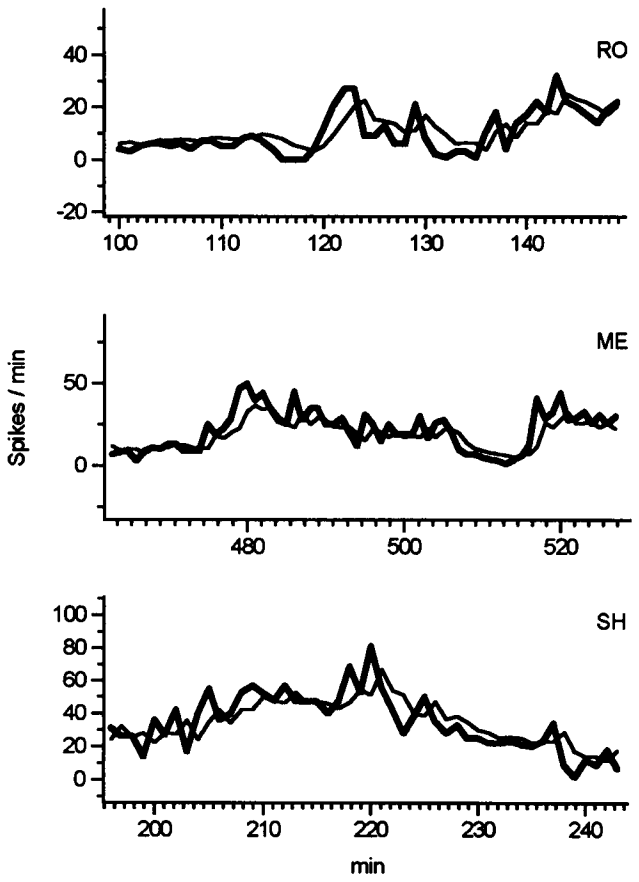

FIGURE 9 Detail of short segments of linear fits of spike rate time series. All data sets were undifferenced. RO is fit by an AR(9) model (9,0,0), ME by an AR(3) (3,0,0) model, and SH by an AR(7) (7,0,0) model.

determined by the decorrelation time; for RO the lag was 23, for SH it was 33, and for ME it was 10. For EG's data, which were uncorrelated, a lag of 1 was used. Nonlinear predictability persisted even after the data were passed through a Hanning window (Press et al., 1992), indicating that this nonlinearity was not an artifact of end effects or due to the distortion of the Fourier spectrum from a finite data series. For these data, three types of surrogate data, as discussed in Methods, were used to see if any could account for the nonlinear predictability observed. The surrogates displayed in Fig. 10 (as above in Fig. 5) reflect the surrogates that best accounted for the experimental data: for RO the surrogates shown were phase randomized, for ME Gaussian scaled, for SH phase randomized, and for EG Gaussian scaled.

We next ask whether linear or nonlinear prediction actually does better at predicting the data from patient ME. The ability to explicitly predict 100 successive points was tested for ME's interval and rate data, using the best AR model and the nonlinear predictor, blinded to points occurring after the prediction point. The results of the 100 predictions are shown in Fig. 11. One can see a clear improvement in the prediction with spikes per minute rather than interval data. The mean squared error for the prediction of the intervals was 73 s^2 for the linear predictor and 51 s^2 for the nonlinear predictor. Although the nonlinear predictor had smaller errors, neither was particularly convincing when visually inspected in the top panel of Fig. 11. For the spike rate data the mean squared error was 58.3 (spikes squared per minute squared) for the linear predictor and 55.8 for the nonlinear predictor. For comparison, the variance of the data about the mean was 105.1, and the mean squared error obtained by assuming the predicted point had the same value as the previous point was 58.0. Note that this "sample and hold" prediction method is equivalent to an AR(1) model and, interestingly, has prediction errors similar to those of the "best linear predictor." Again, the nonlinear predictor did marginally better than either linear predictor. Therefore, for both spike intervals and spikes per minute for patient ME, the explicit prediction error (Fig. 11) and the global fit of the data with the nonlinear predictor (Figs. 5 and 10) suggest that there may be a nonlinear component to the signal.

As noted above, the "best linear predictor" was not necessarily better at prediction than simply using the previous point as a prediction in a simplified AR(1) fashion. We are most concerned that the apparent linear properties of the signal changed with time, either because the system is inherently nonlinear or because the dynamics of the underlying system changed (nonstationarity). In addition, we note that the errors here are prediction errors and not regression errors, which may further explain why the more complex model does not necessarily have smaller errors. It is not likely that the complex model is overfit, as the use of deleted residuals is specifically designed to minimize overfitting (see Appendix).

Although it seems clear by visual inspection of Figs. 2 and 7 that the data set ME is nonstationary, we sought to display just how much the linear characteristics of these data changed with time. After these data were windowed, the autocorrelation function was calculated for the individual partitions of data. The autocorrelation was normalized by the standard error of the estimated linear correlation. Shown in Fig. 12 are the ± 2, ± 4, and ± 8 standard error confidence bands for the windowed autocorrelogram of ME's data. As can be seen, there are statistically significant changes in the linear properties of the signal over time that are highly statistically significant and would clearly have an impact on the ability of any linear predictor to accurately estimate future spike occurrences.

DISCUSSION

This study asked whether EEG spikes had predictability, and if so, whether linear or nonlinear predictors would be best suited to these data. In an attempt to produce very accurate time series, we examined hand-edited data, reducing false positive and negative events as much as possible. Three of four patients so examined showed evidence of linear correlation, and for two of these we could find low-order linear models that could serve as efficient

TABLE 3 Statistics for ARIMA models—spike rates

							White noise*		First 20 residuals‡	
Subject	Parameter	Estimate	SE	T-value	p	Diff.	σ^2	d.f	χ^2	P(wn)
RO	AR (1)	0.49012	0.03864	12.68278	0.00000	0	140.576	678	18.64	0.0678701
	AR (2)	0.07682	0.04338	1.77084	0.07704					
	AR (3)	0.19322	0.04339	4.45323	0.00001					
	AR (4)	0.03355	0.04407	0.76133	0.44673					
	AR (5)	0.00391	0.04410	0.08857	0.92945					
	AR (6)	0.07102	0.04354	1.63124	0.10330					
	AR (7)	−0.07048	0.04355	−1.61822	0.10608					
	AR (8)	0.11293	0.03910	2.88851	0.00399					
	Mean	29.98275	4.39279	6.82545	0.00000					
	Constant	2.66538								
ME	AR (1)	0.49801	0.04120	12.08670	0.00000	0	71.4476	582	14.8511	0.535574
	AR (2)	0.08765	0.04594	1.90790	0.05690					
	AR (3)	0.12607	0.04126	3.05540	0.00235					
	Mean	13.87207	1.19637	11.59510	0.00000					
	Constant	3.99885								
SH	AR (1)	0.41575	0.02968	14.00703	0.00000	0	136.079	1130	14.2049	0.287814
	AR (2)	0.20535	0.03214	6.38877	0.00000					
	AR (3)	0.11853	0.03270	3.62495	0.00029					
	AR (4)	−0.02429	0.03288	−0.73860	0.46015					
	AR (5)	0.03827	0.03270	1.17023	0.24191					
	AR (6)	0.02589	0.03215	0.80539	0.42059					
	AR (7)	0.06662	0.02970	2.24336	0.02487					
	Mean	18.36119	2.16888	8.46575	0.00000					
	Constant	2.82526								

*Estimate of the white noise variance for the linear model, with the given number of degrees of freedom.
‡χ^2 test on the first 20 residual autocorrelation terms. P(wn) represents the probability of having a larger χ^2 value for the residual autocorrelations given white noise. We ask, what is the probability of making an error in saying that the ARIMA fit does not adequately explain the data? The relatively large P(wn) values indicate that the residuals are consistent with white noise, and the model therefore fits the data.

linear predictors. For one of these spike series, an element of nonlinear predictability could be identified as well, which the linear models we tested could not adequately account for.

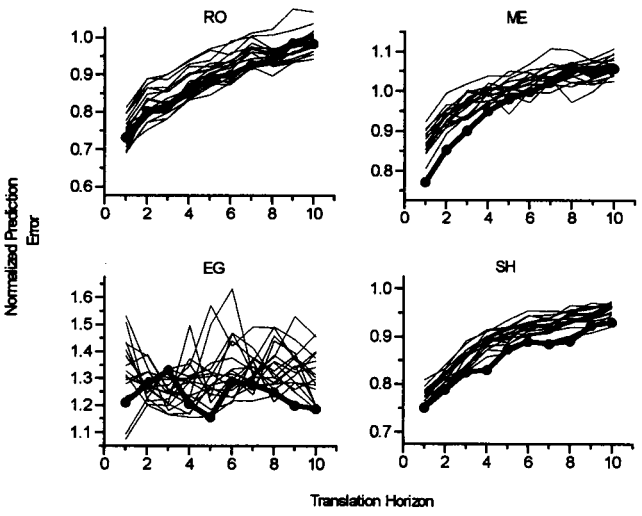

FIGURE 10 Results of nonlinear prediction for spike rates shown in Fig. 7. Heavy black lines indicate experimental data, and 19 surrogate data set results are plotted with thin black lines. Nonlinear predictability appears likely for ME. The embedding dimension was 4 and the lag was determined by the decorrelation time as given in the text.

To examine much longer data sets we studied spikes per minute, but with one exception were unable to rely on hand editing. Nevertheless, we validated the accuracy of our spike detection with the hand-edited segments, and again could account well for three of these data sets with a low-order linear model. Again, only one data set showed evidence of nonlinear predictability that could not be well accounted for with a linear stochastic system, the same data set that demonstrated this property for the hand-edited spike interval data. For this series, explicit linear and nonlinear predictions were also consistent with a nonlinear component not fully accounted for by the linear models. If nonlinear, these spike generators may create very irregular time series that derive from relatively few degrees of freedom—one of the hallmarks of deterministic chaos.

We were struck by the differences in prediction requirements of these spike time series. All were recorded from single implanted depth or subdural electrodes. Nevertheless, the findings ranged from complete lack of predictability, to linear predictability only, and in one case to a blend of linear and nonlinear components. The implication is that if one needed a predictor for a given patient's EEG focus, no single method would suffice a priori. After analysis of a focus' activity, a linear, nonlinear, or mixed prediction method may be required to achieve a certain degree of prediction accuracy. It would

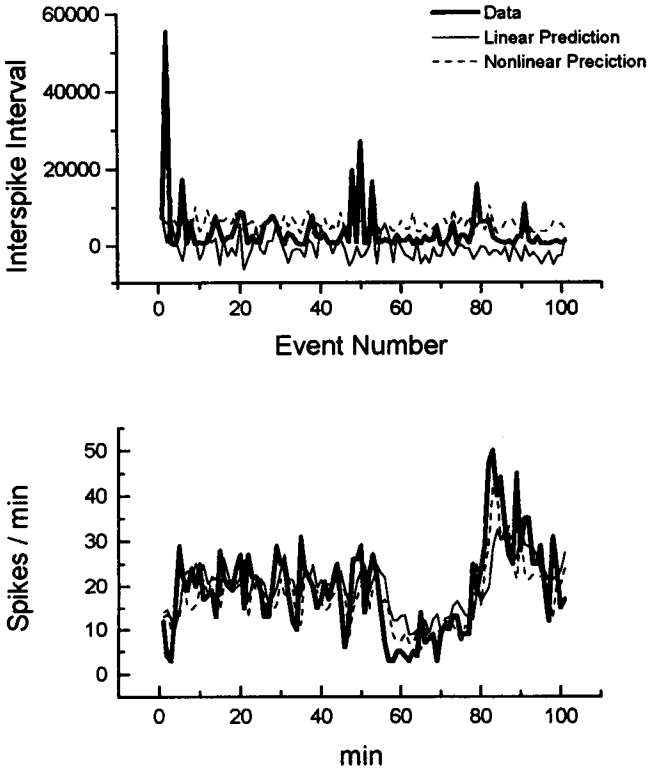

FIGURE 11 Prediction of the interspike intervals and spiking rate for ME. Linear prediction was done using the AR model with the best prediction error sums of squares. Nonlinear prediction was accomplished in the embedding phase space, with the predictor blind to future points.

appear that in certain circumstances, using spike rate data may increase the length of time over which prediction is possible; on the other hand, for other patients the spikes are stochastic events, and no prediction technique will be particularly reliable.

A valid criticism of the present work is the lack of control for state of consciousness. As discussed above, we chose not to control for state of consciousness because by breaking our data sets up into smaller segments (generally less than 10 min in these tracings), the number of spikes remaining in these segments was quite small (see the total number of spikes in the hour-long segments from Table 2). The nonlinear methods are especially prone to spurious results faced with short data sets, and the presence of linear correlation reduces the effective size of the data sets even more when it is present.

We were concerned that the subjective nature of the "hand editing" of our data would introduce bias into our analysis. Although we did not find that we were likely to falsely identify spikes in random data (we examined surrogate data derived from our original EEG tracings), we recognize that a certain amount of bias may be introduced by the process of hand editing long time series of such data. Nevertheless, hand editing remains the "gold standard" for spike identification in EEG and we have relied on it in this study.

Another difficulty with this work is the problem of nonstationarity. For ME, such nonstationarity is evident from inspection of Figs. 2 and 7, and quantitated in Fig. 12. Although differencing for ME easily led to a parsimonious linear fit (Fig. 4, Table 1), it is less clear that such differencing should be done with the nonlinear methods. In addition, we speculate whether nonstationarities in the linear properties would be correlated with an upcoming seizure and are actively exploring this question.

For EEG spike rate data, a more local measurement of nonlinear prediction that would track a system's drift over

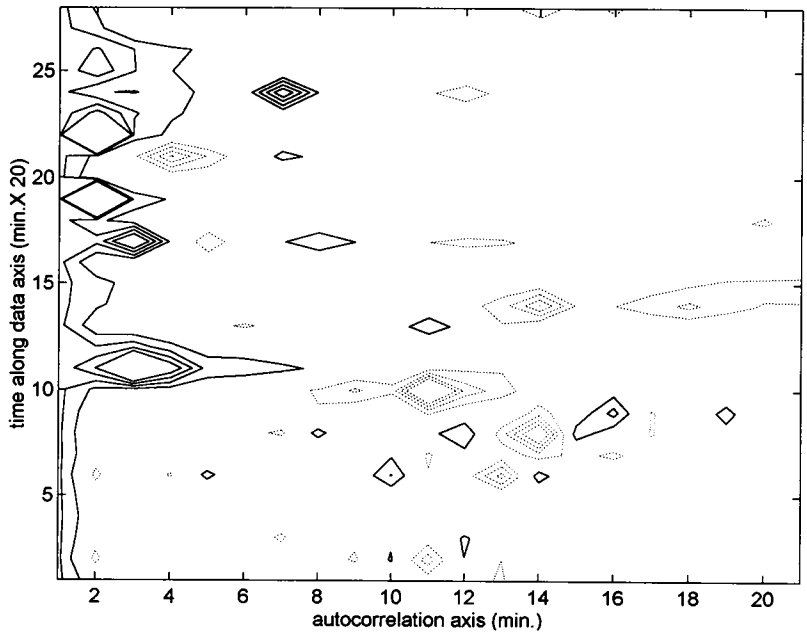

FIGURE 12 Autocorrelogram consisting of windowed autocorrelations of the spike rate data from ME. Contours indicate standard errors of the estimate for the autocorrelations. Solid lines indicate contours of +2, +4, and +8 standard errors. Dashed lines indicate −2, −4, and −8 standard errors. This plot quantifies how the linear properties of the signal change with time.

time would be advantageous. We suspect we might be able to employ an approach for short-term prediction that involves the analysis of *local* properties for predictability (geometrically local using coordinates from a delay coordinate embedding) rather than the more *global* nonlinear statistics we have employed here. Such local methods may demonstrate rather convincing evidence for determinism in neuronal ensembles when burst-firing in a manner similar to interictal spikes (Schiff et al., 1994b). In addition, such local methods have been employed for many of the successful chaos control experiments done in recent years, whether mechanical (Ditto et al., 1990), cardiac (Garfinkel et al., 1992), or neuronal (Schiff et al., 1994b). Whether human interictal spikes possess predictability amenable to such local analysis is entirely unknown.

Although this study establishes predictability (linear and nonlinear) of interictal spikes for all but one of these patients, the more interesting question remains unaddressed: Do spikes predict seizures? Future work will explore this most pertinent issue.

APPENDIX

ARIMA modeling

The ARIMA modeling method we employed is based on a method described by Box and Jenkins (1976). The method allows for a characterization of stochastic systems with linear predictability. ARIMA stands for autoregressive integrated moving average, and applies to three techniques that might be used to characterize a given system.

Moving average (MA) refers to the discrete time convolution of the data with a given filter, a common technique in signal processing. In this case the input signal is assumed to be a series of random shocks with a normal distribution. Such a signal, z_t, with a mean, μ, could be represented as

$$z_t = \mu + a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \cdots \tag{A.1}$$

where θ is a series of weights that make up the MA filter, and a_t is the sequence of random shocks. θ is the MA parameter estimated in the tables given in the article. The variable a_t is the white noise for which the variance is estimated in the same tables.

Data may be "integrated," that is, differenced by taking the first or second difference of the data. This is useful for data that are not stationary (e.g., varying about a fixed mean), or data where the stationary nature of the signal is not apparent at the level of an ARIMA fit.

Autoregression (AR) is a linear stochastic method whereby the present value of signal is regressed on past values of the same signal, as in

$$\tilde{z}_t = \phi_1 \tilde{z}_{t-1} + \phi_2 \tilde{z}_{t-2} + \cdots + \phi_p \tilde{z}_{t-p} + a_t \tag{A.2}$$

Here ϕ is the autoregressive parameter, and this is what is estimated by the AR parameters in the tables. The variable $\tilde{z}$ is the signal, z-μ. Autoregression is a special case of the moving average model, and it can be shown that any model of AR or MA can be represented in terms of the other. The number of terms to represent a simple model of one type in terms of the other type may be large, however, so it is useful to be able to utilize both techniques. In the most general case, one can have mixed models that include both AR and MA terms, as in

$$\begin{aligned}\tilde{z}_t = {} & \phi_1 \tilde{z}_{t-1} + \phi_2 \tilde{z}_{t-2} + \cdots + \phi_p \tilde{z}_{t-p} \\ & + a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \cdots \theta_q a_{t-q}\end{aligned} \tag{A.3}$$

An ARIMA model can thus be described in terms of the number of AR coefficients, the degree of differencing, and the number of moving average coefficients, with the notation generally being a three-unit vector. Thus a third-order AR model with first differencing and five MA coefficients would be referred to as a (3, 1, 5) model.

The overall goal is to develop a parsimonious description of the data's linear and stochastic properties. Care must be taken to avoid overfitting the data, because this may lead to instability in the estimates of the MA and AR parameters.

Prediction error sums of squares

Prediction error sums of squares (PRESS) is a technique of evaluating a regression based on its ability to predict the data. It is based on deleted residuals defined as

$$d_i = Y_i - \hat{Y}_{i(i)} \tag{A.4}$$

where $\hat{Y}_{i(i)}$ is the predicted value for the ith data point Y_i when the regression is fitted without the ith case. The PRESS criterion is the sums of squares of these deleted residuals,

$$\text{PRESS} = \sum_{i=1}^{n} d_i^2 = \sum_{i=1}^{n} (Y_i - \hat{Y}_{i(i)}) \tag{A.5}$$

It is expedient to calculate the deleted residuals directly from the residuals instead of running the regression over for each point. The form used is

$$d_i = \frac{e_i}{1 - h_{ii}} \tag{A.6}$$

where e_i is the ith residual and h_{ii} is the leverage value, a constant related to the degree of influence the ith case has on the overall residual. The variables h_{ii} are the diagonal elements of the hat matrix (Neter, 1990),

$$H = X(X'X)^{-1}X' \tag{A.7}$$

The matrix X is the independent variable of the regression as described below.

The overall regression would stand as

$$Y = X\beta + e \tag{A.8}$$

With the depiction of the regression matrices one can illustrate the method of deleted residuals. For a series of data, S_i, Eq. A.8 could be written as

$$\begin{bmatrix} S_n \\ S_{n-1}{}^* \\ \vdots \\ S_{p+1} \end{bmatrix} = \begin{bmatrix} 1 & S_{n-1} & S_{n-2} & \cdots & S_{n-p-1} \\ 1 & S_{n-2} & S_{n-3} & \cdots & S_{n-p-2}{}^* \\ \vdots & \vdots & \vdots & & \\ 1 & S_p & S_{p-1} & & S_0 \end{bmatrix} \cdot \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2{}^* \\ \vdots \\ \varepsilon_{n-p} \end{bmatrix} \tag{A.9}$$

Suppose that we remove the second point from the regression. This would correspond to all elements in the row marked with an asterisk in Eq. A.9. One could then obtain a regression equation and predict the value of S_{n-1}. This corresponds to $\hat{Y}_{i(i)}$, and d_i would be the error between the actual and predicted value. Now if we do this individually for all of the points in the regression one can get a series of d_i values and thus the value of PRESS by Eq. A.5.

Nonlinear prediction

For nonlinear prediction, the spike interval series, x_i, $i = 1 \cdots N$, was embedded in a time delay embedding space of embedding dimension E and a lag time of L using vectors $\vec{x}_q$

$$\vec{x}_q = (x_q, x_{q+L}, \cdots x_{q+L(E-1)}) \quad \text{(A.10)}$$

where $q = 1, 2, \cdots, N - L$. We chose for our lag time, L, the decorrelation time, which is the time required for the autocorrelation function to decay to a value not significantly different from zero.

In the nonlinear prediction method we compare, within the embedding space, the translation of selected index points to the average translation of its nearest neighbors; the difference is the prediction error for the index points. For an index point x_0, k nearest neighbors are selected, where k = 2% of the number of elements in the time series. The average translation of the k neighbors for a given translation horizon, H, is

$$\langle v \rangle = \frac{1}{k} \sum_{j=1}^{k} x_{j+H} \quad \text{(A.11)}$$

and the translation of the index point is x_{0+H}. Therefore the prediction error is

$$\varepsilon_{trans} = |x_{0+H} - \langle v \rangle| \quad \text{(A.12)}$$

and the prediction error for the mean of the time series is

$$\varepsilon m_{trans} = |x_{0+H} - \text{mean}(x)| \quad \text{(A.13)}$$

Therefore, the normalized prediction error (NPE) is

$$\text{NPE} = \frac{\text{rms}(\varepsilon_{trans})}{\text{rms}(\varepsilon m_{trans})} \quad \text{(A.14)}$$

where rms indicates root mean square. For the nonlinear prediction in this study, we used scalar values of x from the original time series to calculate the average prediction of the nearest neighbors, and denote this method as "zero order." An alternative and slightly more complex method is to construct local linear first-order maps of the embedded data, and this method was discussed in detail by Farmer and Sidorowich (1988). Full source code listings of our nonlinear predictor can be found in Schiff et al. (1994c).

We are grateful to T. Sauer for mathematical advice, to A. S. Makoui and J. Kast for starting the task of hand editing data, and to J. Conry and W. Gaillard for teaching us how to read EEG spikes.

S.J.S. receives support from National Institutes of Health grant 1-R29-MH50006-03, ONR grant N00014-95-1-0138, and the Children's Research Institute. D.A.S. received support from a W. T. Gill Fellowship from the George Washington University School of Medicine. We gratefully acknowledge the contribution of software from Manugistics, Inc., and the granting of a perpetual site license from the Mathworks, Inc.

REFERENCES

Box, G. E. P., and G. M. Jenkins. 1976. Time Series Analysis, rev. ed., Holden-Day, Oakland. 575 pp.

Chang, T., T. Sauer, and S. J. Schiff. 1995. Tests for nonlinearity in short stationary time series. *Chaos.* 5:118–126.

Chang, T., S. J. Schiff, T. Sauer, J.-P. Gossard, and R. E. Burke. 1994. Stochastic versus deterministic variability in simple neuronal circuits. I. Monosynaptic spinal cord reflexes. *Biophys. J.* 67:671–683.

Ditto, W. L., S. N. Rauseo, and M. L. Spano. 1990. Experimental control of chaos. *Phys. Rev. Lett.* 65:3211–3214.

Farmer, J. D., and J. J. Sidorowich. 1988. Exploiting chaos to predict the future and reduce noise. *In* Evolution, Learning and Cognition. Y. C. Lee, editor. World Scientific, Singapore. 279–330.

Garfinkel, A., M. Spano, W. L. Ditto, and J. Weiss. 1992. Controlling cardiac chaos. *Science.* 257:1230–1235.

Gotman, J. 1991. Relationships between interictal spiking and seizures: human and experimental evidence. *Can. J. Neurol. Sci.* 18:573–576.

Hope, A. C. A. 1968. A simplified Monte Carlo significance test procedure. *J. R. Statist. Soc. B.* 30:582–598.

Katz, A., D. A. Marks, G. Mccarthy, and S. S. Spencer. 1991. Does interictal spiking change prior to seizures? *Electroencephalog. Clin. Neurophysiol.* 79: 153–156.

Neter, J. 1990. Applied Linear Statistical Models, 3rd ed. Irwin Press, Baldwinsville.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. 1992. Numerical Recipes, 2nd ed. Cambridge University Press, Cambridge. 545–551.

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

Schiff, S. J., K. Jerger, T. Chang, T. Sauer, and P. G. Aitken. 1994a. Stochastic versus deterministic variability in simple neuronal circuits. II. Hippocampal slice. *Biophys. J.* 67:684–691.

Schiff, S. J., K. Jerger, D. H. Duong, T. Chang, M. L. Spano, and W. L. Ditto. 1994b. Determinism, control and anti-control in a neuronal system. *Nature.* 370:615–620.

Schiff, S. J., T. Sauer, and T. Chang. 1994c. Discriminating deterministic versus stochastic dynamics in neuronal activity. *Integr. Physiol. Behav. Sci.* 29:246–261 .

Takens, F. 1981. Detecting strange attractors in turbulence. *Lecture Notes Math.* 898:366–381.

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