Looking for chaos in brain slices

Peter G. Aitken a*, Tim Sauer b, Steven J. Schiff c

a Department of Cell Biology, Duke University Medical Center, Durham, NC 27710, USA
b Department of Mathematics, George Mason University, Fairfax, VA, USA
c Department of Neurosurgery, Children's National Medical Center and the George Washington University School of Medicine, Washington, DC, USA

Received and accepted 4 November 1994

Abstract

Many signals measured from the nervous system exhibit apparently random variability that is usually considered to be noise. The development of chaos theory has revealed that such random appearing variability may not, in fact, be random, but rather may be deterministic behavior that can reveal important information about the system's underlying mechanisms. We present some new methods for distinguishing determinism from randomness in experimental data, and we apply these methods to population neural responses recorded from hippocampal tissue slices.

Keywords: Chaos; Determinism; Hippocampus; Stochastic; Variability

1. Introduction

All neural circuits exhibit a greater or lesser degree of variability. In some cases predictability can be detected in the varying signal using traditional methods of linear signal analysis, such as power spectrum and autocorrelation. The remaining variability has typically been considered to be noise resulting from stochastic processes operating within the system or impinging upon it from outside. We now know, however, that such apparently random noise may, in fact, be neither random nor noise, but rather deterministic variability that can provide important information about the system's dynamics and mechanisms. The term chaos is used to describe such highly erratic yet deterministic behavior. A non-mathematical introduction to this field can be found in Gleick (1987), and more rigorous treatments in Glass and Mackey (1988), Peitgen et al. (1992) and Moon (1992).

The question of whether neural elements operate deterministically or stochastically is clearly important in advancing our understanding of how the brain works. Several methods have been available for the detection of chaotic processes in experimental data. These methods permit treatment of experimental data and also assist in developing a conceptual understanding of the underlying processes. One can calculate the Lyapunov exponent (Wolf et al., 1985), the correlation dimension (Grassberger and Procaccia, 1983a), and the Kolmogorov entropy (Grassberger and Procaccia, 1983b) from time series data. Relative dispersion analysis (Glenny et al., 1991) and Hurst exponent analysis (Mandelbrot, 1983) can reveal the fractal component of time series. These analysis techniques have been applied to data from a variety of neural systems spanning the spectrum of complexity from single invertebrate neurons (e.g., Mepisto et al. 1988) to the human EEG (e.g., Babloyantz and Destexhe, 1986) (for a review, see King, 1991).

Despite the value of these methods, it has become apparent that they can sometimes lead to erroneous results, suggesting the presence of determinism in a time series where none exists or vice versa. There have been recent advances in analysis techniques to distinguish chaotic from stochastic behavior (Ott, 1993). We present here analysis techniques we have developed, as
well as the results of applying these techniques to data from hippocampal tissue slices. While our interest lies in analysis of signals from nervous system preparations, these techniques are applicable to any physical system that produces time series data.

2. Methods

2.1. General considerations for data collection

Data to be used for the type of analysis presented here will ideally possess two characteristics. (1) The system under study, while emitting complex and possibly chaotic signals, should depend on a relatively small number of dynamical degrees of freedom. Excessive degrees of freedom can result in a signal which, even though highly deterministic, has such a high level of complexity that it appears stochastic to even the most sophisticated analysis techniques. (2) Laboratory techniques should be available that permit accurate and low-noise acquisition of relatively long time-series (at least 1000 points).

Brain slice preparations might appear to be excellent in these criteria. Being isolated from the remainder of the nervous system, a brain slice might well have fewer degrees of freedom than if the same neural circuits were left in situ. Furthermore, experimental manipulations can be used to further simplify the slice preparation. For example, the use of micro-knife cuts can isolate a smaller and therefore less complex portion of the slice. Likewise, chemical dissection with specific receptor antagonists can block the influence of certain neural elements on the response. Slice preparations also permit tight control of experimental variables and easy access for high-quality, stable recordings. It is important to note that measurements for this type of analysis should not be simply a superimposed electrical potential, such as scalp EEG, but rather should directly reflect a meaningful physical parameter such as the number of responding neurons in the network. In the hippocampal slice, the amplitude of the population spike recorded in stratum pyramidale reflects the number of discharging neurons (Andersen et al., 1971).

One caveat is in order. Simplification of the physical system to reduce degrees of freedom makes sense from a theoretical point of view, but the real nervous system may not operate in this manner. We have shown (Chang et al., 1994) that variability in spinal cord reflexes is stochastic in an isolated spinal segment but may exhibit deterministic structure if the segment is not isolated. One could argue that determinism in brain function results from the integrative activity of relatively large numbers of neurons and that isolating a few hundred or thousand neurons from the remainder of the brain destroys rather than enhances determinism.

2.2. Slice preparation and recording

We cut 400 μm transverse slices from the hippocampi of ether-anesthetized 125–150 g Sprague-Dawley rats. Slices are transferred immediately to an ‘Oslo’-style recording chamber where they are placed on a mesh at the interface between oxygenated artificial cerebrospinal fluid (ACSF) and warmed, humidified 95% O2–5% CO2. The ACSF composition is (in mmol/l) 155 Na+, 136 Cl–, 3.5 K+, 1.2 Mg2+, 1.2 Ca2+, 1.25 PO43–, 24 HCO3–, 1.2 SO42–, 10 dextrose. The ACSF flows at 2 ml/min, and the temperature is 35.0–35.5°C. The population spike (PS) in stratum (st.) pyramidale and the presynaptic volley (PV) in st. radiatum are recorded with glass micropipette electrodes filled with 150 mM NaCl and broken back to a resistance of 2–8 MΩ. Afferent fibers are stimulated with constant current pulses (10–150 μA, 0.05–0.20 ms) delivered via a monopolar tungsten electrode; stimulation frequency is 1.0–0.1 Hz, constant for any given experiment. For each experiment the stimulation intensity is adjusted to give a population spike of approximately 50% maximum amplitude and then kept constant.

Individual evoked potentials are digitized and stored to disk for later semi-automated analysis. For each trial, the amplitude of the PS and the PV are measured and the input/output (IO) ratio calculated as PS/PV. Each experiment provides two time series: one of the raw PS amplitudes and one of the IO ratios.

2.3. Detection of determinism

2.3.1. Time-delay embedding

A physical system that changes with time can be represented by a graph in a state space where each dimension, or coordinate, represents one of the system’s independent, or state, variables. Typically, the individual points in a state space plot are connected in temporal order. If the state space plot has only 2 or 3 dimensions, a visual representation can often reveal structure in the dynamics of a system that is not evident from the more traditional time series plots, where the value of each variable is plotted as a function of time. If the state space plot has more than 3 dimensions, mathematical techniques can be used to detect structure.

With real-world dynamical systems one rarely has access to all of the system’s state variables. Most often we are limited to measuring one variable repeatedly over time to create an experimental time series. With only one variable it is not possible to create a state space plot. The technique of time-delay embedding provides a solution by creating one or more additional variables (coordinates) from previous, or delayed, measurements from the same time series. In the simplest
form, a 2-dimensional time-delay embedding, for a time series of \( n \) measurements \( x_1, x_2, \ldots, x_n \), each measurement \( x_a \) is plotted against the immediately preceding measurement \( x_{(a-1)} \). Additional coordinates at larger time delays can be specified if needed to unravel the data. The term *embedding dimension* is used to specify the number of time-delay coordinates being plotted.

The technique of time-delay embedding is based on the method of non-linear system reconstruction described in (Packard et al., 1980) and given mathematical foundations in (Takens, 1981). This methodology, called delay coordinate reconstruction, uses the fact that the dynamical attractor underlying the system, in a geometric sense, is in one-to-one correspondence with the set of vectors:

\[
(x_{t'-1}, x_{t'-2}, \ldots, x_{t'-d}),
\]

where \( x_t \) denotes the output signal of the system at time \( t \), where \( d \) is sufficiently large and where \( t' \) is a unit of time called the time delay, which can for most purposes be the sampling period. There is an implicit mathematical assumption that the measured time series \( \{x_t\} \) is coupled to all important degrees of freedom in the system. This set of vectors, called delay-coordinate vectors, can be analyzed geometrically for structure. The graph of such vectors is called an embedding. The mathematical foundation extends back to Whitney (1936). An accessible discussion of the experimental applications of this technique can be found in Moon (1992) and a detailed theoretical analysis is given in Sauer et al. (1991).

The technique of time-delay embedding is based upon the assumptions that the state of a deterministic system at a given time depends on its state in the recent past and that the physical variable being measured provides valid information about the system’s ‘true’ state variables. While visual inspection of a 2- or 3-dimensional time delay plot can be useful for detecting obvious chaotic behavior, more rigorous mathematical methods are clearly required. Our approach is based on the fact that determinism is a necessary (albeit insufficient) condition for a chaotic process. Compared to a stochastic process, a deterministic process is one that should, in the short term, be predictable to some degree. In addition, there should be some local smoothness in state space to the structure of the system’s dynamics, reflecting the underlying ‘rules’ (equations) that describe the behavior of the system. A stochastic system lacks such local smoothness. We are primarily interested in those situations when the data appear irregular and stochastic, even when subjected to conventional signal processing techniques, yet possess exploitable predictability. Because determinism, or predictability, is a necessary feature distinguishing chaotic from stochastic systems, detecting this determinism is therefore the first step to both defining a model to explain the dynamics of the system and planning a strategy to consider controlling such systems. We describe three independent tests for determinism.

### 2.3.2. Local Flow

The Local Flow method begins with the time-delay embedded representation of the data and uses a discrete adaptation (Kaplan, 1993) of a technique for continuous systems (Kaplan and Glass, 1992). This discrete adaptation first selects points in the embedding space that are close to each other. For each point, a vector is calculated to the point in the embedding space at a selected time in the future. The selected time translation is called the translation horizon. These vectors are then normalized and averaged; larger average vectors are produced by deterministic rather than stochastic systems. In other words, if one translates nearby points in embedding space a short time into the future, they will tend to move in the same direction if the system is deterministic but in random different directions if the system is stochastic.

From Kaplan (1993), for a discrete time series \( s_i \) \((i = 1, 2, \ldots, N)\), we construct a time-delay coordinate embedding based on an embedding dimension \( E \), and a time delay \( L \):

\[
\mathbf{x}_q = (x_{q+1}, x_{q+2}, \ldots, x_{q+(E-1)\lambda})
\]

for \( q = 1, 2, \ldots, N - L \). The embedding space is covered with a grid of cubes. The number of points in each cube \( j \) is \( n_j \), and their time indices are \( t_{1j}, t_{2j}, \ldots, t_{nj} \). For a given translation horizon, \( H \), the change in state from time \( t_{ijk} \) to \( t_{ijk} + H \) for each of the \( n_j \) points in the cube \( j \) is

\[
\Delta X_{ijk} = x(t_{ijk} + H) - x(t_{ijk})
\]

for \( k = 1, 2, \ldots, n_j \). Since points near the edge of the cloud of points will have a directional bias towards the middle of the cloud, \( \Delta X_{ijk} \) is mapped onto a sine function as

\[
\Delta X_{ijk} = \left( \sin \left( \frac{2\pi}{\lambda} \left[ x(t_{ijk} + H) - x(t_{ijk}) \right] \right), \ldots, \sin \left( \frac{2\pi}{\lambda} \left[ x(t_{ijk} + (E-1)L) - x(t_{ijk} + (E-1)L) \right] \right) \right)
\]

where \( \lambda \) is a characteristic length of the embedded attractor. An averaged vector, \( \mathbf{r}_j \), arises from cube \( j \), where the number of vectors passing through the cube is \( n_j \) for \( n > 2 \).

For different values of \( n \) and \( E \) these averaged
vectors form a family of values \( L_i^E \). These values are averaged over all values of \( n \) as

\[
N = \left( \frac{(L_i^E)^2 - c_E^2/n}{1 - c_E^2/n} \right)
\]

where \( \langle \rangle \) denotes average over \( n \), \( c_E^2/n \) is the expectation value of \( L_i^E \) if the process was random. A more complete discussion of the derivation of \( c \) can be found in Kaplan and Glass (1992). For the analysis of data presented above we used an embedding dimension of 4 and mesh grain of \( 16^E \) cubes.

2.3.3. Local Dispersion

The Local Dispersion method is closely related to the Local Flow method. It examines the dispersion of nearby points in the embedding space as they are translated in time (Wayland et al., 1993). If nearby points move similar distances in similar directions, this is evidence for determinism. If the distances and directions wildly vary, the process is more likely stochastic.

From Wayland et al. (1993), \( x_q \) is defined as above, and \( x_1, \ldots, x_k \) are the \( k \) nearest neighbors of \( x_0 \), an arbitrary value of \( x_q \). If these nearest neighbors are translated into the future by a certain horizon, \( H \), then \( y_0, \ldots, y_k \) are the \( k \) nearest ‘images’ of the index point \( x_0 \) and its \( k \) nearest neighbors. The error, \( v \), produced between the images and original points:

\[
v_j = y_j - x_j
\]

should be small in deterministic systems. The average error is

\[
\langle v \rangle = \frac{1}{k + 1} \sum_{j=0}^{k} v_j
\]

and the translation error is defined as

\[
\epsilon_{\text{trans}} = \frac{1}{k + 1} \sum_{j=0}^{k} \| v_j - \langle v \rangle \|^2
\]

where \( \| \| \) indicates the length of the enclosed vector. \( N \) points are randomly chosen from the embedded attractor, and the median value of \( \epsilon_{\text{median}} \) is calculated. For the calculations reported in this paper we chose a translation horizon \( H \) of 1, and a value of \( k \) equal to 2% of the number of values in the time series.

2.3.4. Non-linear Prediction

The Non-linear Prediction method selects index points in the embedding space and for each index point determines how the nearest neighbors of that point move with time. The average of the movement of the neighboring points is treated as a prediction of how the index point will move; this prediction is compared with the actual movement of each index point. The difference between the actual and predicted movement is the error of prediction. This non-linear prediction technique was explored in detail by Farmer and Sidorowich (1988). We apply a fairly simple implementation of this technique to our data, called zero order prediction, using local constant maps. There are more elaborate implementations of non-linear prediction that use powerful signal processing techniques in conjunction with embedding methodology to handle higher noise levels as well as higher order local maps. Our implementation is published in a recent volume on the Santa Fe Institute Time Series Prediction Competition (Sauer, 1993).

Techniques with varying levels of sophistication and sensitivity are available for non-linear prediction. The specific method used depends on the quality and length of the time series data. The simplest approach, which we used, is as follows. For an embedded time series, \( x \), as above, again choose \( k \) nearest neighbors of a given index point \( x_0 \). The \( k \) nearest neighbors are translated by the horizon, \( H \), and their average translation is now

\[
\langle v \rangle = \frac{1}{k} \sum_{j=1}^{k} x_{j+H}
\]

which can be viewed as a zero-order prediction of the translation of \( x_0 \) by \( H \) time units. The prediction error for the index point is

\[
\epsilon_{\text{trans}} = | x_{0+H} - \langle v \rangle |
\]

and the error of predicting the mean of the time series is

\[
\epsilon_{\text{trans}} = | x_{0+H} - \text{mean}(x) |
\]

where \( | \| \) indicates absolute value. The normalized prediction error, \( NPE \), is

\[
NPE = \text{RMS}(\epsilon_{\text{trans}})/\text{RMS}(\epsilon_{\text{mean}})
\]

where RMS denotes root mean square. For the calculations above, we chose an embedding dimension of 8 and a number of nearest neighbors equal to 2% of the number of values in the time series.

More elaborate versions of non-linear prediction can be applied when high-quality data series of 2000 or more points are collected. In this case 1% of the points on the reconstructed attractor amounts to at least 20 points. Sufficient information is then available in each small neighborhood within the attractor to form a high-dimensional embedding and project down to the local principal directions of the attractor using the singular value decomposition (SVD). A local linear model is parametrically fit along the (relatively few) principal directions. This is a significant improvement in the local modeling process. The benefits of a high-dimensional embedding to untangle the attractor can be used, while relatively few parameters need to be fit. Only those in the principal directions are fit, and the other directions are presumed to be noise. The reduct-
2.4. Surrogate data controls

If the system behaves chaotically, and the time series measurement has been chosen wisely, then the time-delay embedding will encode the deterministic structure of the system's behavior. The prediction techniques discussed above can then be used to detect this structure. If the reconstructed map is imperfect (and it almost surely will be), the predictions will have errors but short-term accuracy may still be good. A 'good' prediction is one that is better than a random choice; more precisely, one that is statistically superior to the results of non-linear prediction on a stochastic process that has autocorrelation similar to the experimental data. It is essential to establish this distinction, which is the first indication that the process under study contains deterministic elements that warrant further experimental investigation.

We test the goodness of prediction by comparing the predictions from the experimental data with the predictions from a control system that is intentionally stochastic (Kaplan and Glass, 1993). The control system is created by generating surrogate data sets from the original data. The process of surrogate generation retains certain characteristics of the original data (such as number of data points, mean, standard deviation or autocorrelation) but randomizes the data so that any deterministic structure that may be present is destroyed. The same non-linear prediction techniques are then applied to the original data and to the surrogate data. If the results are similar for the experimental and surrogate - that is, if they provide equally good (or bad) predictions - then we cannot reject the null hypothesis that the experimental data can be explained by a linear stochastic model.

We use three types of surrogate data. The first type, called randomized phase, preserves the mean, standard deviation and autocorrelation of the observed time series while randomizing the phases. A Fourier transform is performed on the original data, generating a sequence of complex numbers with amplitudes corresponding to the power spectrum. The phases of these complex numbers are essential to reconstruct the original data through an inverse Fourier transform. A random phase is added to each coefficient of the Fourier transform before performing an inverse Fourier transform. The result is a random time series that has the same mean, standard deviation, and autocorrelation as the original (Schiff and Chang, 1992; Theiler et al., 1992). Preserving autocorrelation implies that the phase randomized surrogate time series has the same 'smoothness' as the original time series and that mathematically it has the same frequency content (power spectrum).

The gaussian scaled surrogate assumes that the data came from a normally distributed (gaussian) random process that was filtered through a non-linear filter (e.g., synapses, bad amplifiers, etc.). In this surrogate, autocorrelation is not fully preserved, but the original data values of our experimental time series are used and shuffled, preserving the amplitude distribution of our data. This surrogate was originally proposed by Theiler et al. (1992).

The Fourier shuffled surrogate was devised by us to overcome some of the difficulties encountered with the gaussian scaled surrogates. This surrogate again uses the original data values from the experimental time series, but it tries to approximate as closely as possible the linear-correlation or smoothness of the original data. First a phase randomized surrogate is created as described above, and then the original data are substituted by rank ordering.

Surrogate data sets are used as follows. For each experimental time series three separate realizations of each of the three types of surrogate data sets are generated. The nine resulting surrogate data sets are analyzed in the same manner as the experimental data. A technique suggested by Theler et al. (1992) is used to determine if the experimental data yield significantly

![Fig. 1. Results of local flow, local dispersion, and non-linear prediction algorithms on the Hénon data, compared with 3 sets of each of the 3 types of surrogate data. Diamonds (••) represent 'experimental' (Hénon) results and dashed lines (−−−) represent surrogate data results. As can be seen, each method is readily capable of detecting the determinism in these data, showing clear separation from the results obtained with the surrogate data. The abscissa represent translation horizon for the Local Flow and Non-linear Prediction plots, and embedding dimension for the Local Dispersion plots. Embedding dimensions are 4 and 8 for Local Flow and Non-linear Prediction, respectively, and Local Dispersion translation horizons are 1 in this and subsequent plots.](image)
Fig. 2. Input/output (IO) measurements from 1000 constant intensity stimulations of the CA1 region of a hippocampal slice. Each measurement is the IO ratio calculated as the population spike amplitude divided by the presynaptic volley amplitude. Data have been detrended and normalized to zero.

better prediction than the surrogate data. For each point plotted along the abscissa, the standard deviation is calculated for the surrogate data, and the number of standard deviations ('sigmas') separating the surrogate mean from the experimental value is determined. It is important to focus on the surrogate data set giving the least separation between surrogate and experimental results, because, if any surrogate data can account for the experimental findings, the null hypotheses that the data can be modeled by a linear stochastic process cannot be rejected.

3. Results

3.1. Validation of determinism detection algorithms

To validate the determinism algorithms described above, we tested them with a time series known to be chaotic, generated from the Hénon (1976) equations:

\[ x_{n+1} = 1 - Ax_n^2 + y_n \]
\[ y_{n+1} = Bx_n \]

We generated a chaotic time series from one variable of these equations (using \( A = 1.4 \), and \( B = 0.3 \)), and then from this series we generated three instances of each of the three types of surrogate data set. Fig. 1

Fig. 3. Results of applying the three determinism algorithms to the IO data time series recorded from a hippocampal slice (○) and to three instances of each type of surrogate data (---) generated from the experimental time series.
shows the results of each of the determinism algorithms applied to the Hénon time series and to three realizations of each type of surrogate data set. The abscissa represents translation horizon for the Local Flow and Non-linear Prediction plots, and embedding dimension for the Local Dispersion plots. Fig. 1 shows clearly that each of the determinism algorithms can distinguish the deterministic Hénon time series (diamonds, Fig. 1) from the stochastic surrogate data sets (dashed lines, Fig. 1). When quantified as described above, these results show a minimum separation between experimental and surrogate data of 6.4, 13.1 and 29 sigmas for the Local Flow, Local Dispersion and Non-linear Prediction algorithms, respectively.

3.2. Analysis of hippocampal data

In each of six separate experiments we generated time series of at least 1000 points. The data were detrended to remove any long-term linear components and then subjected to the analysis methods described above. Fig. 2 shows the IO time series collected from one hippocampal slice, and Fig. 3 shows the results of applying the three determinism algorithms to this time series. It is clear that in almost all cases the predictions from the experimental data (solid lines, circles) overlap, or are separated only slightly from, the predictions from the surrogate random data sets (dashed lines). These results argue that the time series is stochastic. The greatest separation between experimental and surrogate data is seen with non-linear prediction, where the prediction errors are an average of 2.1 standard deviations below the surrogate data mean for the first 5 points. Note, however, that the absolute values of the normalized prediction errors are greater than 1.0, indicating that the algorithm actually provided worse predictions than a guess based on the mean of the time series. This too is a stochastic result.

4. Discussion

In the past dozen years or so, the development of the field of chaos has shown that a signal which, to traditional analysis methods, appears to be random noise may, in fact, contain deterministic components that provide important information about the underlying mechanisms of the system. When developing theories of nervous system function, it is essential to know whether the component neural elements behave in a purely stochastic manner or whether there is a degree of determinism present. The question of determinism in brain function has been addressed over the past 50 years by a variety of scientists, including Schrödinger (1944), Eccles (1953), Rosenbleuth (1970), and Crick (1994). While the question of determinism in large scale brain function remains beyond the reach of quantitative science, recent advances in the analysis of complex systems now make possible the investigation of determinism in small populations of neurons.

Once a complex signal has been shown to contain deterministic elements, then what? A deterministic signal is predictable, at least in theory. There clearly would be great value in being able to predict the future behavior of the nervous system. Even if the functioning of the nervous system is shown to contain significant chaotic elements; however, the fact that chaotic processes are exquisitely sensitive to initial conditions makes accurate predictions unlikely. The best we might realistically hope for is a useful degree of accuracy in short-term prediction.

Our initial results with analysis of evoked potentials in the hippocampal slice preparation suggest that the variability in this response is stochastic. As is usually the case, however, one cannot prove the null hypothesis. That we could detect no determinism in this system may mean that it is, in fact, stochastic. It may also mean that whatever determinism is present in the system is too complex to be detected by our methods of analysis given the size of the data sets available, or that determinism is totally or partially lost when the hippocampal slice is isolated from the remainder of the nervous system.

References


