Predicting epileptic seizures using nonlinear dynamics

by

William Marshall

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Mathematics
in
Statistics

Waterloo, Ontario, Canada, 2008

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Epilepsy is a nervous system disorder which affects approximately 1% of the world's population. Nearly 25% of people who have epilepsy are resistant to traditional treatments such as medication and are not candidates for surgery [32]. A new form of treatment has emerged that attempts to disrupt epileptic activity in the brain by electrically stimulating neural tissue. However, the nature of this treatment requires that it is able to accurately predict the onset of a seizure in order to time the intervention correctly. Recent studies suggest that EEG recordings may be generated by a low dimensional nonlinear process [35] [36] [6]. This paper will investigate nonlinearity tests, as well as the use of methods from the theory of nonlinear dynamical systems in the prediction of seizures or seizure like events (SLEs) from complex time series. To do this data is generated from a nonlinear dynamical system with a stochastic time dependent parameter, which attempts to emulate the different states of an epileptic brain. Two kinds of nonlinearity tests were used in simulations, one which specifies a model in the alternative hypothesis (Keenans test) and one which simply states that the process is 'not linear' (Surrogate data test). The tests were applied to the generated data, as well as a short EEG recording from a person with epilepsy and a simple nonstationary example. Both tests were able to correctly identify the model as nonlinear, neither test identified the EEG data as nonlinear and there were conflicting results when the tests were applied to nonstationary data. Estimates of the correlation dimension and Lyapunov exponent were then used to classify the preictal state of the model data. Correlation dimensions showed the best ability to classify states, so they were used in the prediction algorithm. The results of the simulation was that the correlation dimension was able to successfully predict half of the SLEs, however there was an alarmingly high false prediction rate. These results suggest that even though a complicated model may fit the data better, when dealing with prediction it is usually best to use a simple model. A simpler approach with better understood statistical properties may be able to improve on the prediction of SLEs as well as reduce the computational cost of performing them.

Acknowledgements

Dedication

This is dedicated to my friends and family.

Contents

List of Tables

List of Figures

List of Symbols

P - Phase space
S_t - A continuous mapping corresponding to t units of time
T - A set which describes the time range
I - Identity matrix
$x(t)$ - A trajectory of a dynamical system (may be a vector)
x_t - Time series of measurements from observable of a dynamical system (may be a vector)
A - Constant matrix
U, V - Subset of the phase space
u, v - Elements of U or V respectively
$d(\cdot)$ - Distance function
$\mu(\cdot)$ - Measure
D - Box-count dimension
λ - Lyapunov exponent
N - Length of time series
$C(r)$ - Correlation integral
ν - Correlation Dimension
R - Discriminating quantity in surrogate data method
Z - Test statistic
e_t - i.i.d noise
y_t - Vectors in the reconstructed phase space
d_0 - Minimum embedding dimension parameter for method of delays
τ - Time lag parameter for method of delays
$H(\cdot)$ - Entropy
$Pr(\cdot)$ - Probability mass function
$MI(\cdot)$ - Mutual information
$E, E1, a$ - Quantities used in Cao's method
ψ - Prediction quantity
μ_ψ, σ_ψ - Mean and SD of prediction quantity
γ_j - j^{th} autocorrelation

Chapter 1

Introduction

1.1   Introduction to Dynamical Systems

Dynamical systems theory has its roots in classical mechanics where it is used to study problems such as the motion of a pendulum or the flow of water through a pipe. Abstractly speaking, a dynamical system is a mathematical representation of a deterministic rule which describes the time evolution of a state vector in its ambient space. I.D. Chueshov [10] describes dynamical systems theory, 'The mathematical theory of dynamical systems is based on the qualitative theory of ordinary differential equations'. It is sometimes impossible and often difficult to find actual solutions of differential equations, so instead the long term behaviour of a general class of solutions is studied. If a population p_t eventually dies out for all values of $p_0 < 100$ and survives if $p_0 \geq 100$ where p_0 is the initial population, this is an example of a qualitative feature. Qualitative features can also be based on the parameter values of the system.

The following definition of a dynamical system comes from Chueshov's book Introduction to the Theory of Infinite-Dimensional Dissipative Dynamical Systems [10]. A dynamical system consists of two components which are represented by the pair (P,S_t). The object P is a set which defines the phase or state space of the dynamical system, each vector in the phase space corresponds to a unique state of the system. The time evolution of the dynamical system is represented by a family of continuous mappings S_t from P $\rightarrow$ P $\forall$ t $\in$ T, where T is a set which is closed under addition and has an identity element. It defines the time domain of the dynamical system and it is usually of the form $[t_0,\infty)$ for continuous systems or $\{t \in \mathbb{N} \mid t \geq t_0\}$ for discrete systems where t_0 is the identity element. The mappings S_t must have two properties:

$$S_{t_1+t_2} = S_{t_1} \circ S_{t_2} \quad \forall t_1, t_2 \in T$$

and

$$S_{t_0} = I.$$

The first property means that it does not matter whether you propagate the whole time in one step or break it down into two or more smaller steps, the final result will be the same. The second property is that the map corresponding to the initial time is the identity map, it does not change anything. In practice the mappings S_t are rarely explicitly known, instead a system of differential equations (difference maps) or a time series of observed measurements is all the information that is known about the system. Additionally, there often exist unknown parameters within the system of differential equations which need to be estimated. There is small class of systems for which the mappings S_t can be solved for analytically using the differential equations, however this means that alternative methods must be used to study the vast majority of systems.

Suppose the mappings S_t define a continuous function x(t)=$S_t x_0$ for $x_0 \in$ P. A dynamical system is called linear if

$$\frac{dx}{dt}(t) = Ax(t)$$

in the continuous case or

$$x_{t+1} = Ax_t$$

in the discrete case, for some constant matrix A. This is an important class of dynamical systems because if the initial condition x_0 is known, the solution x(t) of the differential equation (called a trajectory of the dynamical system) can be solved for analytically. On the other hand, nonlinear dynamical systems, in general cannot be solved analytically. This has led to the development of a number of qualitative methods which can be used to help understand the system. One of the ways to analyze a nonlinear system is to look at the long term behaviour of trajectories in general, instead of the behaviour of a single trajectory with a specific initial condition. This is often the case for ergodic systems where the long term behaviour is the same for all initial conditions (see Section 1.1.2 for more on ergodicity).

A fixed point of a dynamical system is a point $x^* \in P$ such that $S_t x^* = x^*$, $\forall t \in T$. Fixed points are also called stationary or equilibrium points. Fixed points are useful for describing certain systems, however there are many simple situations where they are inadequate for describing the behaviour of the trajectory. As an example of this, a trajectory is called periodic if there exists $\bar{t}$ such that $x(t+\bar{t}) = x(t)$, $\forall t \in T$, $\bar{t}$ is called the period of the trajectory. Another generalization of the concept of fixed points is an invariant set, where a subset $U \subseteq P$ is invariant if $S_t U \subseteq U$, $\forall t \in T$. Fixed points, periodic points and invariant sets are examples of qualitative behaviour and are excellent tools for describing the long term behaviour of certain

2
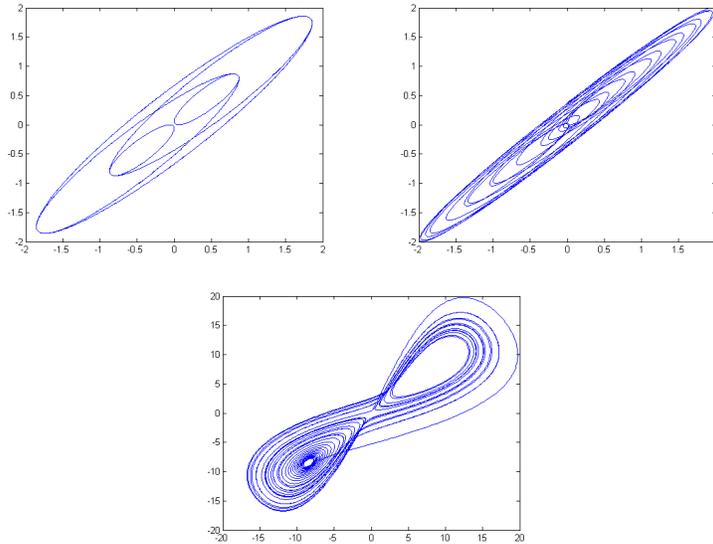
Figure 1.1: Example of phase spaces showing periodic (top left), quasi-periodic (top right) and chaotic (bottom) behaviour.

dynamical systems, however their definition is still not quite broad enough to describe the chaotic dynamics found in some systems.

1.1.1 Chaos

Before we describe the phenomenon known as chaos, we define the concept of an attractor. This is important because chaotic dynamics are often exhibited in the presence of an attractor. Collet and Eckman [34] informally describe an attractor as, 'the set of points to which most points evolve'. A set $U \subseteq P$ is defined to be an attractor if: U is an invariant set and U has the property that for every bounded set $V \subseteq P$,

$$\lim_{t \to \infty} \sup_{u \in U} \inf_{v \in V} d(S_t u, v) = 0$$

where $d(\cdot)$ is a distance function in P. In plain language this means that any element $u \in U$ will eventually be arbitrarily close to an element of V. Some definitions of an attractor also require that there are no subsets of U which also have these properties (U is minimal) [27].

There are many different definitions of chaos, for a more extensive look at the definitions and their implications the reader is referred to [1]. The definition that

is used here is, a subset $U \subseteq P$ of a dynamical system is called chaotic if it satisfies these two properties

- sensitivity to initial conditions,

- transitivity.

Sensitivity to initial conditions means that there exists $r > 0$ such that for each initial condition $u_0 \in U$ and $\epsilon > 0$ there exists another point $u_1 \in U$ and $t \in T$ such that

$$d(u_0, u_1) < \epsilon$$

and

$$\|S_t u_0 - S_t y_0\| \geq r.$$

The implication of this is that two trajectories which are initially arbitrarily close will eventually be separated by at least r units, making prediction difficult in the presence of any uncertainty or measurement error. Sensitivity to initial conditions is closely related to the condition of a positive largest Lyapunov exponent (see Section 1.1.3) in other definitions of chaos.

Transitivity means that for any two open sets $V_1, V_2 \subseteq U$ there exists $t \in T$ such that

$$S_t V_1 \cap V_2 \neq \emptyset,$$

The important of transitivity is to exclude situations where the sensitivity to initial conditions is caused by trajectories that are simply diverging to infinity. Chaotic dynamics can exist on the entire phase space or only in a subset of the phase space. The most interesting situation occurs when the chaotic region is an attractor because this guarantees that the system will eventually be chaotic for a large number of initial conditions. Such attractors are sometimes called strange or fractal attractors, however this term is also used for describing the geometry of the attractor [27]. For an illustrative example of periodic and chaotic dynamics, see Figure 1.1.

1.1.2   Ergodicity

Another concept which is central to the study of dynamical systems is ergodicity. Ergodic theory is developed in a measure theoretic environment and it is based on the idea of invariant measures. In our context, a measure μ is said to be invariant if

$$\mu(S_t^{-1}U) = \mu(U)$$

for all measurable sets $U \subseteq P$ and all $t \in T$. It is not always easy, or even possible, to find an invariant measure for a given dynamical system. A dynamical system is said to be ergodic with respect to the invariant measure μ, if every invariant set $U \subseteq P$ has $\mu(U) = 0$ or $\mu(U) = 1$. One useful result for ergodic systems is that the long term behaviour of the system will be the same for almost all initial conditions. This is especially important for time series data where the initial condition may not be known, it allows results of analyses to be generalized across the system.

From the definition of ergodicity, it is not obvious how important a role it plays in statistical analysis of dynamical systems. This connection is made by what is possibly the most important result for ergodic dynamical systems, the ergodic theorem.

Theorem 1.1.1 *Let (P,S_t) be a dynamical system which is ergodic with respect to the invariant measure μ and f an integrable function on P. The limit*

$$f^*(x) = \lim_{N \to \infty} \frac{1}{N} \sum_{t=0}^{N-1} f \circ S_t(x)$$

exists μ-almost everywhere, moreover it is equal to the spatial average

$$f^* = \frac{1}{\mu(P)} \int f d\mu$$

The ergodic theorem is such an important result that it is often viewed as an alternative and equivalent definition for ergodicity. The relationship between the ergodic theorem and time series data is analogous to the one between the strong law of large numbers and independent data. It tells us that even though the data from a single trajectory is dependent, the sample mean will converge to the true mean of the system. This allows us to compute inferences on the mean using the well understood methods for independent data.

Another concept which is often studied along side ergodicity is mixing. A dynamical system is said to be strongly mixing with respect to an invariant measure μ if for every pair of sets $U, V \subseteq P$ it has the property

$$\lim_{t \to \infty} \mu(U \cap S_t V) = \mu(U)\mu(V).$$

A mixing dynamical system has the property that after enough time, the initial conditions can no longer be recovered from the current data. This statement is stronger than corresponding statements about ergodic systems, so it is not surprising that mixing implies ergodicity.

1.1.3 Lyapunov Exponents

The Lyapunov exponents of a dynamical system are numbers which quantify the divergence of two trajectories of the phase space which are initially very close. For a dynamical system with a D_P dimensional phase space, the exponents are defined by the long term evolution of the principal axes of an infinitesimal D_P-sphere centered at x_0. In general the Lyapunov exponents are dependent on the initial condition, however for an ergodic system they will be the same for almost all initial conditions. If we denote the length of the ith principal axis at time t by $d_i(t)$ then the ith Lyapunov exponents of the dynamical system are defined as

$$\lambda_i = \lim_{t \to \infty} \lim_{d_i(0) \to 0} \frac{1}{t} \log \left(\frac{d_i(t)}{d_i(0)} \right) \quad i = 1..D_P.$$

Existence of this limit has been shown under certain conditions (Oseledec theorem [31]) however they do not exist for all systems. Traditionally the exponents are ordered from largest to smallest, $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_{D_P}$.

The largest exponent λ_1 corresponds to the vector of maximum divergence while the sum of the two largest exponents $\lambda_1 + \lambda_2$ corresponds to the plane of maximum divergence and in general the sum of the first k exponents corresponds to the maximal k-dimensional divergence [44]. Many properties of a dynamical system can be deduced from the exponents, for example a positive largest exponent indicates sensitivity to initial conditions (one of the requirements of chaos). If the sum of all the exponents is negative then the overall volume of objects in the phase space goes to zero as $t \to \infty$ i.e., the system is dissipative. It is extremely difficult to solve for the exponents analytically and quite often it is necessary to solve them numerically. For an in depth discussion on Lyapunov exponents and a numerical method to calculate them from equations of motion, see Benettin et al [2].

1.1.4 Dimension

The box-counting dimension D is a geometric quantity which is used to partially characterize an attractor. It measures the extent to which the attractor fills space. If we denote N_ϵ to be the number of boxes of side length ϵ that are required to cover a set, then D is defined as

$$D = \lim_{\epsilon \to \infty} \frac{\log(N_\epsilon)}{\log(\frac{1}{epsilon})}.$$

Methods to estimate the fractal dimension from the definition are based on counting algorithms which can be very costly to compute. These algorithms require large
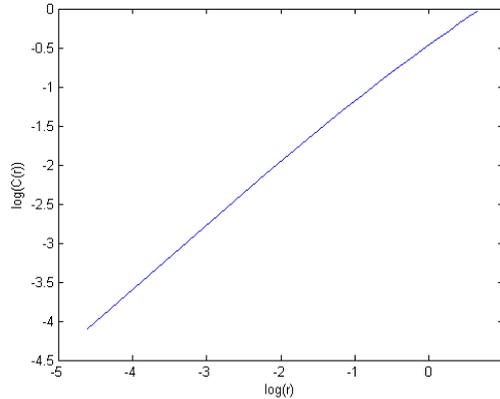
Figure 1.2: The correlation integral C(r) scales with r, $C(r) \propto r^\nu$. By plotting $\log(C(r))$ against $\log(r)$ and estimating the slope of the resulting line we can estimate the correlation dimension ν.

amounts of data in order to get good estimates and they take quite a long time to compute. Additionally estimates of fractal dimension from time series have been shown to be impractical due to the amount of data required to get precise estimates [16]. Instead a related statistical quantity, $\nu \leq D$, known as the correlation dimension is often used because it is more easily estimated. The correlation integral of a set of points $\{X_i\}_{i=1}^N$ is defined as

$$C(r) = \lim_{N \to \infty} \frac{1}{N^2} \sum_{i,j=1}^N \rho(r - ||X_i - X_j||)$$

where ρ is the Heaviside function defined as

$$\rho(x) = \begin{cases} 0 & x < 0 \\ 1 & x \geq 0 \end{cases}.$$

The correlation dimension scales with the correlation integral

$$C(r) \propto r^\nu$$

for small values of r. The motivation behind the correlation dimension is that in higher dimension spaces there are more ways for points to be close together, thus the rate at which C(r) increases depends on the dimension of the space. The standard way to estimate ν is to look at plots of $\log(C(r))$ against $\log(r)$ and estimate the slope of the line. For an example of these plots, see Figure 1.2
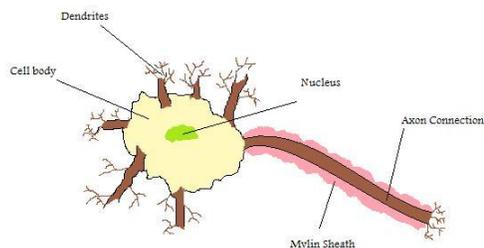
7

Figure 1.3: A diagram depicting the components of a typical neuron

1.2 Introduction to brain biology

The following information on the nervous system can be found in the book Psychology by David Myers, module 3 [29]. The nervous system is the human body's primary mode of information transmission. With the brain as its control centre the nervous system controls movement, regulates internal organs and performs more complex tasks such as learning. Information is transferred through the nervous system by billions of cells called neurons or nerve cells. While there are many different kinds of neurons, they all share the same basic structure. A typical neuron has a main cell body surrounded by tiny branch-like fibers called dendrites which are used to receive information and a longer axon fiber which branches out at the end to communicate information to other neurons. Figure 1.3 is a picture depicting the typical neuron. The axon fiber may be surrounded by an insulating tissue called the myelin sheath which increases the speed of information transmission. When a neuron is stimulated by pressure, heat, light or other neurons it sends a brief electrical charge down the axon called an action potential. Once the action potential reaches the ends of the axon branches it triggers the release of a chemical stimulant to be received by the dendrites of other neurons.

The following information about brain functions can be found in module 4 [29]. The nervous system is very compartmentalized, with many different groups of neurons each performing very specific tasks. It should not be surprising that different regions of the brain have their own functions. The deepest parts of the brain perform the more basic functions required for survival. Included in these deep brain structures is the brain stem which regulates heartbeat, breathing and sleep. Directly on top of the brain stem is the thalamus which acts as a hub for sensory information, receiving the information and then relaying it to the appropriate area. The cerebellum is located right behind the brain stem, it is used in voluntary motion and in maintaining balance. The final part of these lower brain structures is the limbic system, often thought of as the border between low and high levels of

brain function. The limbic system consists of the amygdala, hypothalamus, hippocampus and pituitary gland. Using the pituitary gland as a link to the hormone system, the limbic system plays an important role in emotions such as fear and anger as well moderating appetite and sex drive. Myers also writes that, 'One limbic system component, the hippocampus, is essential to memory processing' [29]. Surrounding the interior regions is the largest part of the brain, the left and right cerebral hemispheres. The cerebral hemispheres consist mostly of axon fibers that connect the surface of the brain to the interior. Finally there is the cerebral cortex, a thin layer of tightly packed cells about an eighth of an inch thick which covers the outside of the brain. It is the cortex and its role in high level brain function which distinguishes us from other animals. While most complicated actions require interaction between the different lobes, it has been possible to localize the primary region for many of them. The cortex has no natural divisions, however there is human imposed structure which is used to help map out brain functions. The frontal lobes are located at the front of the brain, they contain the part of the brain that controls motor functions. The frontal lobe is also mainly responsible for many advanced functions which are unique to humans such as language and problem solving. Behind the frontal lobes at the top rear of the head is the parietal lobe, it contains the part of the brain which interprets body sensations. At the back of the brain are the occipital lobes, this part of the brain receives visual information. Finally at the sides of the brain just above your ears is the temporal lobes, where auditory information is received. An map of different regions of the brain can be found in Figure 1.4.

1.2.1   Recording brain activity

Scientists have been able to map out these areas of the brain by recording the electrical, chemical and magnetic signals that are constantly being created by our brains. Magnetic signals are recorded from the brain using fMRI (functional magnetic resonance imaging) scans. fMRI scans are a special case of MRI scans, which work by putting the brain into a magnetic field to cause its atoms to align and then disrupting them by sending a radio wave through. As the atoms work to realign themselves they disrupt the magnetic field and allow the researcher to generate 2D images of the concentration of atoms in the brain. When a neuron is active it requires more oxygen than an inactive neuron, this increase in oxygen alters the magnetic effects in the surrounding area and allows us to determine which regions of the brain are most active. Chemical signals in the brain are recorded using a PET (positron emission tomography) scan. The PET scan uses the fact that neurons use glucose as 'food' to determine which regions of the brain are most active. To do this they give the brain a temporarily radioactive form of glucose and monitor where the food is going to determine which brain regions are active.
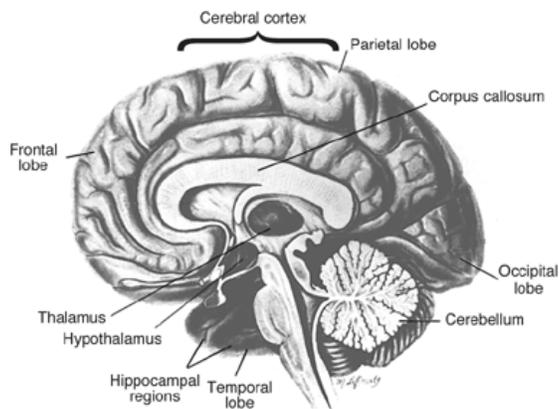
Figure 1.4: This image is from [30]. This image shows the different parts of the brain, as well as the different regions of the cerebral cortex.

EEG (electroencephalogram) recordings are amplified recordings of the electric potentials generated by the neurons in the brain. They are recorded using tiny electrodes placed on the scalp or directly against brain tissue. Unlike MRI or PET scans which use time averages to monitor which regions of the brain are active, EEG recordings give an overall picture of what is happening in the brain on much smaller time scales. The spatial scale of EEG recordings depends on both the number and placement of electrodes. The number of electrodes can be as low as two for intracranial EEG recordings where the target area is very localized. For more exploratory kinds of analyses, caps are worn against the scalp which can have up to 256 electrodes distributed around the head. As the number of electrodes on a net increase so to does the price of the net, which means many researchers must compromise spatial resolution in order to lower cost.

One way which EEG recordings can be analyzed is by computing their Fourier transform and looking at the power spectrum. Researchers have shown that large amounts of power in different frequency bands correspond to the different stages of the sleep cycle. For example a large portion of the power in alpha band (8-13Hz) is a characteristic of a wakeful person where as power in the theta band (4-7Hz) indicates the first stage of sleeping. Another common way to analyze EEG recordings is with the idea of an event related potential (ERP). An ERP is a very specific waveform that is generated when a stimulus is shown to a participant, however that waveform is usually hidden amongst other brain activity. ERP analysis is based on repeating the same stimulus a large number of times and then averaging all the signals together in hopes to cancel out unrelated brain activity and get a clearer picture of the ERP. Although the general shape of an ERP is the same for all people and all stimuli, they can differ in their timing and amplitude. This allows

researchers to look for differences in the ERP across different groups of people, for example comparing the ERP of a university student population to that of an elderly population. Another way that researchers are able to use ERPs is to change the instructions given to a participant, by instructing a person whether or not to remember the stimuli which are presented they are able to affect the ERP.

There are a number of issues which need to be dealt with before an EEG recording can be analyzed in a meaningful way. The recording picks up a 60Hz electrical signal which is given off by the electrical equipment in the room. This is generally dealt with by running the signal through a notch filter[1] to get rid of all 60Hz activity, including anything that was generated by the brain. Additional noise comes from electric potentials generated by muscle activity, such as blinking or jaw clenching. It is sometimes possible to regress out the contribution from blinking, however most data which contains muscle activity is simply thrown out of the analysis. Another issue is that correlations between electrodes may be spurious due to volume conduction in the head.

A property of field potentials is that they can only be measured up to a constant, meaning that one must look at the difference between electric potentials in order to generate a physically meaningful quantity. In EEG recordings differences are obtained by comparing the electric potential at each electrode to a single reference point. The ideal reference point is one where there is no electrical activity present however there is no place on the scalp which fits this criteria. Reference electrodes are often placed on the ears or mastoid bones because they are thought to be the area with the least amount of electrical activity. Sometimes electrodes are referenced to two points on opposite sides, this is called a bipolar reference. The advantage of a bipolar reference is that it does a good job capturing localized activity, however global activity that is common to both references is lost during this process. Another option is to rereference against the average voltage of all the electrodes, this method is independent of the initial reference point.

---

[1]A notch filter is used to filter out the contributions from a specific frequency

Chapter 2

Epilepsy and Seizure Prediction

Epilepsy is a nervous system disorder where abnormally coherent firing of a region of neurons causes a temporary loss of brain function [9]. This loss of brain function is called a seizure and their intensity can range from no physical signs to severe convulsions. From an electrical perspective the transition into the seizure corresponds to a change from high complexity possibly chaotic activity to low complexity possibly rhythmic behaviour [9]. The region where a seizure originates is called the epileptogenic focus, from there the electrical patterns spread to other areas of the brain. Iasemidis et al [17] write that, 'For some types of epilepsy (e.g focal or partial) these events are caused by structural changes in neuronal circuitry within localized regions of the cerebral cortex'. In focal epilepsy, the epileptogenic focus is the region of the brain where the seizure activity originates. When the epileptogenic focus is in the hippocampus some of the structural abnormalities that can occur include loss of neurons and dendritic simplification[1]. Characteristic activities of the neurons inside the epileptogenic focus are periodic firing of action potentials, followed by prolonged periods without firing. Hypothesized reasons for the spread of seizure activity include lack of inhibitory ability of surrounding cells, propagation of action potentials and/or sustained synchrony recruiting distant neurons into the process [23].

Epilepsy affects about 1% of the world's population and of those people about 25% are immune to traditional therapies [32]. Additionally, treatment through surgery requires the removal of a portion of the brain, a process which often leads to impaired cognition, language and motor control [9]. Recently a new method of treating epilepsy has been the focus of much research. The method entails the suppression and interruption of epileptic EEG activity using low voltage electrical stimulation of neuronal tissue at the appropriate time before seizure onset [9]. However sufficient knowledge about the mechanics of epilepsy are not known to effectively control the seizures. In order to successfully suppress imminent seizures,

---

[1]Dendritic simplification is characterized by retracting branches in dendrite fibers

we must be able to accurately predict their onset time. When the control method is applied at the appropriate time it can interrupt the synchronization of neurons, returning the brain to a normal state [13].

2.1 Linear versus nonlinear models

Due to its complex and unpredictable behaviour, EEG signals can be thought to be generated by a high dimensional linear stochastic process [20]. Methods based on linear stochastic models have been shown to only be able to detect seizures at most a few seconds before their onset [14]. During the last two decades there has been a shift in thinking that perhaps the complex EEG signals are generated by a low dimensional nonlinear process. This is largely due to a greater understanding of chaos and the ability of low dimensional nonlinear systems to exhibit complex dynamics. It should be noted though, that just because low dimensional systems may exhibit complex dynamics does not imply that all complex systems are low dimensional.

2.1.1 Testing for nonlinearity

There are many different methods available to test the linearity hypothesis in time series data. Note that by linear stochastic process it is meant that the time series it can be written as

$$X_t = e_t + \sum_{i=1}^{\infty} a_i X_{t-i} + b_i e_{t-i}$$

where $\sum_{i=1}^{\infty} \|a_i\| < \infty$, $\sum_{i=1}^{\infty} \|b_i\| < \infty$ and e_t is i.i.d with mean zero and finite variance. In the case where e_t is normally distributed the process is completely described by it's covariance structure [7]. These methods can vary in the statement of their assumptions and alternate hypotheses, so it is necessary that these factors be carefully considered when choosing which test is suitable for a specific situation.

There are a number of tests which attempt to reject the linearity hypothesis by looking for quadratic (or higher order, finite) components in the data. [21] [37]. The benefits of these methods are easy implementation, fast computation and knowledge of the distribution under the null hypothesis. The downside of such methods is that they are very restrictive in the kind of nonlinearity they can detect. It is quite easy to construct counter-examples which do not fit the model in the alternate hypothesis, causing the test accept the null hypothesis even though the process is

nonlinear.

Another class of tests look for general dependencies in the data which are not explained by a linear model [42] [3]. These tests are more robust than the previous ones because they do not restrict themselves to only detecting certain types of non-linearity. These methods however, have none of the benefits of the previous tests. Implementation is difficult due to a large number of parameters which must be estimated, quantities which detect general dependencies are usually quite difficult to estimate and their distribution must be estimated using bootstrap techniques.

It should be noted that both types of tests have a few other problems which still need to be overcome. The results of these tests all depend on a stationarity assumption, which is difficult to justify without already knowing the form of the model. Additionally linear models can be arbitrarily complicated, hence the hypothesis that a process is linear is infinite dimensional. Having an infinite dimensional null hypothesis changes the dynamics of hypothesis testing, common properties such as degrees of freedom are no longer applicable and comparisons between the and the null and the data now involve both finite and infinite objects.

The surrogate data method tests the hypothesis that the time series was generated by a linear stochastic process. This method will be discussed in more depth, because it has previously been applied in the context of EEG recordings [42] [35] [36]. For a review of some of the other possible tests, see [37] and [8].

Linear surrogate data

A linear surrogate data set is an artificially generated set which has the same linear properties (autocorrelation function) as the original but does not constrain any other properties. These surrogate datasets can be created using ARMA or Fourier transform (FT) based methods. A discussion on these methods and how they differ can be found in a paper by Theiler et al [42]. The FT method involves computing the FT of the original signal, randomizing the phase of the coefficients while keeping their magnitude constant and then finding the inverse FT to get a new time series. The result of this process is a time series which has the same linear properties of the original series.

Surrogate data method

The surrogate data method as described in [42] is performed as follows:

- select a null hypothesis which specifies the model of the data,

- create surrogate data sets which share certain properties with the original data set,

- calculate values of a quantity R for original and surrogate data sets,

- estimate probability distribution of R under H_0 using surrogate values,

- calculate probability of getting a value R_{orig} under H_0(p-value).

The properties which must be inherited by the surrogate data depend on the model specified in the null hypothesis. The power of this test will be determined by the choice of the statistic R, ideally it would be a quantity which is not completely determined by the model in the null hypothesis. In the case of of looking for non-linearity in data, the hypothesis is that the data is generated by a linear stochastic process $\{X_t\}$. With this hypothesis we can then use the linear surrogate data method which is mentioned above to create additional datasets having the same covariance structure as the original data. The choice of the quantity R should be one that is sensitive to nonlinearities in the data. Some popular options include correlation dimension, maximum Lyapunov exponent or redundancy which is an extension of the concept of mutual information (see Section 3.1.1) defined as

$$R(X_1, ..., X_N) = \sum_{i=1}^{N} H(X_i) - H(X_1, ..., X_N).$$

The Lyapunov exponents may not be defined for the system, however this is not a concern. Wolf [44] states, 'While the existence of this limit has been questioned, the fact is that the orbital divergence of any data set may be quantified'. There are several options for estimating the distribution of R, including normality tests like Kolmogorov-Smirnov or bootstrapping methods. To compute the p-value of the test, we define a statistic

$$Z_{obs} = \frac{\|R_{orig} - \bar{R}\|}{s_R}$$

where $\bar{R}, s_R$ are the sample mean and standard deviation respectively. In the case where R is normally distributed, Z follows a t-distribution with $N_s - 1$ degrees of freedom, where N_s is the number of surrogate data sets used. Using this method nonlinearities have been found in human EEG when R is the redundancy [35] [36] and when R is the correlation dimension [6].

There are some problems with the method of surrogate data which may result in higher false rejection rates than anticipated. The Fourier transform method for generating surrogate data has difficulty reproducing long correlation times and the periodic assumptions of the transform can sometimes introduce spurious high frequency components [42]. In order to ensure that the the test is performing up to theoretical expectations, it is required to have fast decaying auto-correlations

$$\sum_{j=1}^{\infty} j\gamma_j^2 < \infty,$$

as well as periodic boundary points $x_0 = x_N$. A discussion about the additional conditions which must be met to compensate for the above difficulties can be found in [7].

2.1.2   Nonlinearity test example

In this section we will look at the performance of two different nonlinearity tests, one which specifies an alternative form for the time series and one which is open ended. For the former we will be using Keenans test [21] and for the latter we will be using the surrogate data method. For Keenans test the choice of parameter M is not critical [8], in our application we take it to be M=10, the order of the linear model that is fit. For the surrogate data method, twenty surrogate datasets were generated and the statistic that was used to capture nonlinearities was the correlation integral. The value of r was chosen to be appropriate with the scale of the data after visual inspection.

These tests will be administered to three different data sets, our model for the epileptic brain (see Section 4.1), an EEG recording from an actual epileptic brain (see Section 2.4) and a simple non-stationary time series. For the model and real epileptic data, the tests were performed on 1001 point segments of the data during the interictal state. The non-stationary data was defined as follows,

$$X_t = -\frac{1}{2}X_{t-1} + e_t \quad t = 1..500$$

$$X_t = \frac{1}{3}X_{t-1} + e_t \quad t = 501..1000$$

where $X_0 = 0$, e_t is normally distributed with mean zero and unit variance

The results of the nonlinearity tests can be found in Table 2.1.2. Both tests were able to correctly classify the model data as nonlinear, however neither test classified

Nonlinearity test results		
	Keenans Test (p-value)	Surrogate Data (p-value)
Model Data	F=1353.40 (¡0.001)	Z=31.02 (¡0.001)
Real Data	F=1.23 (0.268)	Z=0.51 (.309)
Non-Stationary Data	F=0.20 (0.655)	Z=10.06 (¡0.001)

Table 2.1: Results of nonlinearity tests

the real data as nonlinear. The non-stationary data had some very interesting results, Keenan's test classified it as linear and the surrogate data test as nonlinear. A closer look at the data provides some insight into why the test give conflicting results. Keenans test looks for a quadratic model as an alternative and since the underlying model is not quadratic it does not reject the linearity hypothesis. The surrogate data method does not look for a specific alternative and instead it is sensitive to changes in the covariance structure, the test detects the change point in the model and rejects the linearity hypothesis.

2.2 Prediction using nonlinear models

There are some difficulties with estimating nonlinear quantities from an EEG recording due to the nature of the signal. Estimation of nonlinear quantities such as correlation dimension or Lyapunov exponents require long, noise free and stationary signals none of which are guaranteed in EEG [15]. Despite these limitations many interesting results have stemmed from the nonlinear analysis of epileptic EEG and other seizure like events. Lehnertz and Elger [23] [24] have shown that in intracranial EEG recordings the correlation dimension exhibits different dynamics at the epileptic focus. They have also shown that the behaviour of the correlation dimension at the focus shows the existence of a preictal state lasting from 4-25 minutes. Ouyang et al [32] show that by replacing the hard boundary Heaviside function with a softer boundary Gaussian in estimating the correlation dimension, it becomes better at classifying the ictal states in rat EEG. Similarly, Iasemidis et al [18] [17] have shown that localization of the epileptic focus in human EEG recordings is possible using a modified 'short term' Lyapunov exponent (STL). They also show that the STL characterizes a preictal state in over 90% of seizures with a mean preictal length of over 30 minutes. Attempts to turn this method into a real time prediction algorithm with dynamic site selection resulted in 82% of seizures successfully predicted with a mean anticipation time of 71.7 minutes and 0.16 false positives per hour [19]. This method however does not offer an accurate prediction of the time until the seizure, just that there will be one in the future. A number of other nonlinear quantities such as Kolmogorov entropy, local expansion exponents and dissipation have also been shown to be able to identify a preictal state in human EEG on the scale of several minutes [28].

2.3 Alternate prediction methods

There has been some criticism of the use of nonlinear methods in seizure prediction [26]. These criticisms include the validity of nonlinearity tests, the nonstationarity of EEG signals and results being more important than model fit. This has led to the development of several statistical algorithms for seizure prediction. Statistical methods will ideally provide a more robust method for identifying the preictal state from short noisy EEG recordings. As well, these statistical methods are often less computationally intensive than nonlinear methods and the hope is that this will make online implementation much more practical.

Chiu et al [9] have been able to classify the ictal states of seizure like events (SLEs) of extracellular field recordings from in-vitro rat hippocampal slices. Using a wavelet based ANN they were able to classify the preictal states with a 83% true positive rate and a 80% true negative rate. Additionally they showed that typically ignored high frequency components in the 100-400Hz range are crucial to the classification of SLEs [11]. Attempts to use this method as a real time prediction algorithm were able to predict 70% of seizures with a mean prediction time of 36 seconds and mean error of 14 seconds [12]. Ouyang et al [25] have shown that the preictal state in rat EEG could be classified using recurrence quantification analysis. They have also produced a wavelet based algorithm which can classify ictal states with a mean anticipation time of 7 minutes and less false positives than their previous methods [33]. Winterhalder et al used measures of synchrony in human intracranial EEG to predict 60% of seizures with a false positive rate of 0.15 per hour. [43].

2.4 Real data

For an example of actual EEG data from people with epilepsy, see Figure 2.1. The first recording is sampled at 500Hz and is 1500000 points (or 50 minutes) long, it contains a single ictal event. The second recording is also sampled at 500Hz but is 841000 points (or 28 minutes) long, it also contains an ictal event. The onset of the ictal events can be seen clearly, they correspond to the regions of increased amplitude. It should be noted however that this increase is likely not due to increased brain activity but electrical activity created by muscles due to the more physical aspects of seizures. Also this is raw EEG data, there is a large amount of preprocessing which must be done before it would be ready for rigorous analysis. See Table 2.1.2 for the results of nonlinearity tests applied to this data.
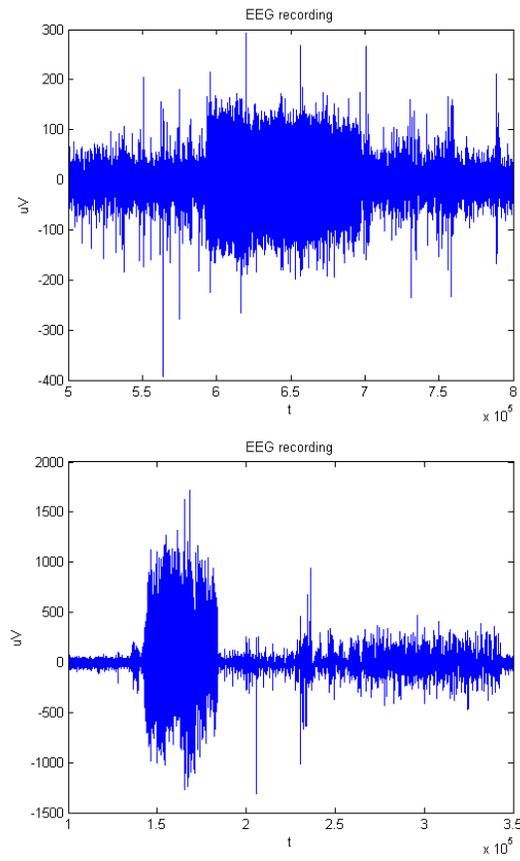
Figure 2.1: Two EEG recordings of seizures from a person with epilepsy.

Chapter 3

Methods

In many real life situations we do not observe the full system evolving in its phase space and all we have is a finite amount of data sampled from an observable in order quantify the dynamical system. An observable of a dynamical system is any smooth function of the trajectory of the system. Such analyses are made possible by an important theorem by Floris Takens which allows us to reconstruct the phase space of a dynamical system from an observable [41].

Theorem 3.0.1 *Let P be a compact manifold of dimension D_P. For pairs (φ, y), $\varphi{:}P{\to} P$ a smooth diffeomorphism and $f{:}P{\to} \mathbb{R}$ a smooth observable then the map $\Phi_{(\varphi,f)}{:}P{\to} \mathbb{R}^{2D_P+1}$ defined by: $\Phi_{(\varphi,f)}(x) = \{f(x), f(\varphi(x)), \; ... \; , f(\varphi^{2m+1}(x))\}$ is an embedding.*

In this context the meaning of smooth is a continuous second derivative and embedding is an injective function from $P \to \mathbb{R}^{2D_P+1}$ such that Φ is a diffeomorphism between P and $\Phi(P)$.

3.1 Method of delays

Takens theorem is the basis for a method to reconstruct the phase space called 'The Method of Delays'. The time series $\{x_t\}_{t=1}^n$ is a set of measurements of the observable at equally spaced time steps. We define a set of measurements of $\Phi_{(\phi,y)}$ as $y_t = [x_t, x_{t+\tau}, ..., x_{t+(d_0-1)\tau}]$, the points $\{y_t\}_{t=1}^{n-d_0}$ make up our phase space reconstruction. The value d_0 is the minimum embedding dimension of the object we're looking to embed, it is bounded above by the unknown quantity $2D_P + 1$. The value τ is a time lag which represents an iteration of φ. It should be noted that the points $\{y_t\}$ only represent the *active regions* [1] of the phase space, meaning it

---

[1]The active region of the phase space is the points which are visited by the current trajectory
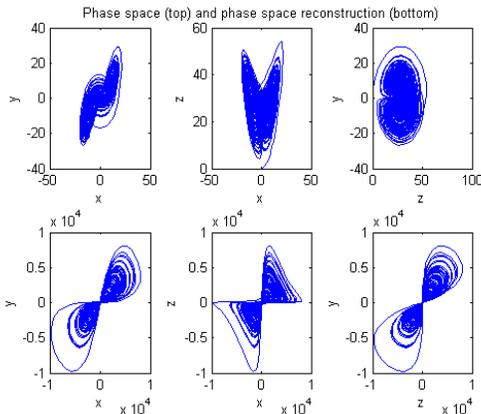
Figure 3.1: Projections of the phase space of the Lorenz equations onto the XY, XZ and ZY planes (top). Projection of the reconstructed phase space onto the XY, XZ and ZY planes (bottom). The phase space is reconstructed using the observable $f(x, y, z) = sin(y) - x^3 + 2y$ and the parameters $\tau = 5$ and $d_0 = 3$

may only reconstruct a portion of the phase space. It is important that our reconstruction contains all relevant regions of the phase space, such as an attractor, so that quantities describing these features can be accurately estimated. In order for this to happen, it is required that the dynamical system be ergodic. See Figure 3.1 for plots of the phase space of the Lorenz equations, as well as a reconstruction of the phase space from the observable $f(x, y, z) = \sin(y) - x^3 + 2z$ with parameters $\tau = 5$ and $d_0 = 3$. This reconstruction is done with 10 000 data points in three dimensions, however when dealing with smaller amounts of data and/or higher dimensional reconstructions the quality of the reconstruction may decrease. Also, it should be noted that the axes of the original phase space do not correspond to specific axes in the reconstruction and a rotation or scaling transformation may be required to make the spaces more visually similar.

3.1.1   Estimating τ

When estimating the parameters for the method of delays, τ is typically estimated first. This is because most methods for estimating d_0 are dependent on τ. In the case of an infinite amount of noise free data, any value of τ would suffice, however this is rarely the case. By carefully choosing the value of τ to maximize the amount of information about the attractor contained in each vector (or by minimizing the redundant information) we can reduce the minimum embedding dimension d_0. However, we must be careful not to chose τ so big that successive points become independent or we will lose the structure of the object that we are trying to embed.

There are several methods of estimating τ including one based on the auto-correlation function [38] as well as one based on Fourier analysis [18]. The method that is used in this paper is based on mutual information, a concept from information theory developed by Shannon [40].

Mutual Information Method

If the set $\{\alpha_1, \alpha_2 ... \alpha_K\}$ defines the possible states of a time series X_t, with associated probabilities $Pr_X(\alpha_i)$ then the entropy of the time series is defined as

$$H(X) = -\sum_{i=1}^{K} Pr_X(\alpha_i) \log_2(Pr_X(\alpha_i)).$$

The entropy of a time series is a measurement of its uncertainty and is measured in bits. Similarly we can define the joint entropy of two signals X, Y by

$$H(X,Y) = -\sum_{i,j=1}^{K_X, K_Y} Pr_{\{X,Y\}}(\alpha_i, \beta_j) \log_2(Pr_{\{X,Y\}}(\alpha_i, \beta_j)).$$

From here we define the mutual information (MI) of two signals, it measures the reduction in entropy of the signal X given knowledge about the signal Y. In other words it measures the amount of shared information between the two signals, it is defined as:

$$MI(X,Y) = H(X)+H(Y)-H(X,Y) = \sum_{i,j=1}^{K_X, K_Y} Pr_{\{X,Y\}}(\alpha_i, \beta_j) \log_2\left(\frac{Pr_{\{X,Y\}}(\alpha_i, \beta_j)}{Pr_X(\alpha_i)Pr_Y(\beta_j)}\right).$$

The mutual information method takes τ to be the first local minimum of the lagged mutual information function $MI_\tau(X_t, X_{t+\tau})$.

Implementation

To estimate the joint probability distribution $Pr_{\{X,Y\}}$ we first discretized the time series by dividing the range of each signal into eight equal sized bins for a total of 64 bins and denote the number of observations from the (i,j) bin by $N_{i,j}$. The number of bins is important for getting good estimates of the probability distribution. If there are too many bins there will not be enough data to get a reliable estimate,
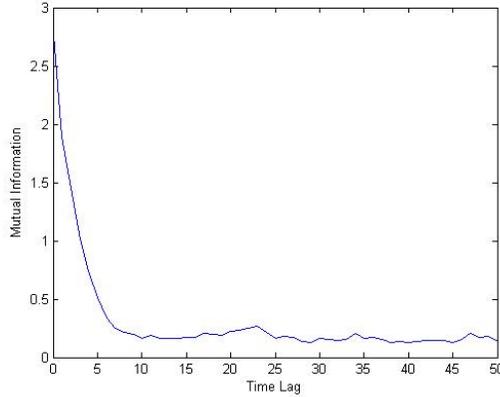
Figure 3.2: A plot of the time lagged mutual information $MI(X_t, X_{t+T})$ against T. the first local minimum of this function is used as an estimate for τ. In this example the value of τ would be 10.

while if the number of bins is too small important features of the distribution will be averaged over and lost. We then define a simple frequency based estimate of the probability distribution by

$$\widehat{Pr_{\{X,Y\}}}(\alpha_i, \beta_j) = \frac{N_{i,j}}{N}, i,j = 1..8.$$

We similarly define the marginal distributions by

$$\widehat{Pr_X(\alpha_i)} = \sum_{i=1}^{8} \frac{N_{i,j}}{N}, i = 1..8.$$

Using these estimates of the probability distributions we then estimate $MI(X_t, X_{t+\tau})$ for values of τ ranging from 0 to 50. The value of τ for the method of delays is taken to be the first local minimum of $MI(X_t, X_{t+\tau})$. See Figure 3.2 for an example of this method.

3.1.2   Estimating d_0

The estimation of d_0 has been done using a variety of different methods. One technique called the invariant method is to measure an invariant property of the attractor for several values of d_0. The idea is that the invariant property will not be invariant for values of d_0 which are too small and will be a constant for values of d_0 which are large enough. The minimum embedding dimension is thus taken to be the smallest value of d_0 such that the invariant is constant [16]. Another more statistical method for determining d_0 is based on estimating the dimensionality of the attractor using singular value decomposition [4]. There are also methods
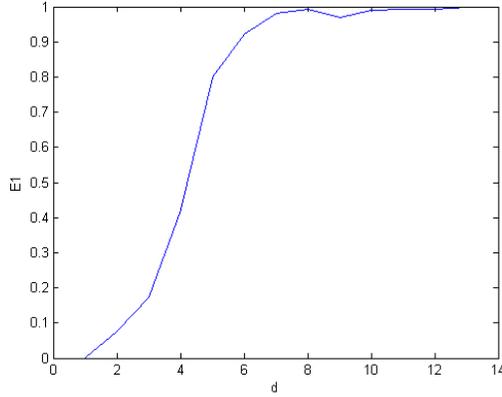
Figure 3.3: A plot of the value E1(d) against d used in the Cao method. The value of d_0 is estimated to be the value for which E1(d) stops changing. For this example we take the value of d_0 to be 7.

which deal with a concept of false neighbors. The idea is that if you are projecting points into a space which is too small then points will be forced together, creating false neighbors which would not actually be close together in a large enough space. These methods monitor false neighbors as the dimension d_0 increases, and chose the value of d_0 where false neighbors no longer exist [22]. The method that will be used in this paper is an extension of the nearest neighbors method called Cao's method [5].

Cao's Method

We denote the ith vector in our $d_0 = d$ dimensional phase space reconstruction as $y_i(d)$ and the index of the nearest neighbour to $y_i(d)$ as n(i,d). With this notation we define the quantity

$$a(i, d) = \frac{||y_i(d + 1) - y_{n(i,d)}(d + 1)||}{||y_i(d) - y_{n(i,d)}(d)||}$$

which will be large for false neighbours but relatively small for true neighbours. We then estimate the spatial average of this quantity

$$E(d) = \frac{1}{N - \tau d} \sum_i a(i, d)$$

and monitor its behaviour as we increase the dimension. If the value of

$$E1(d) = \frac{E(d + 1)}{E(d)}$$

24

stops changing for all $d \geq k$ then the value of the minimum embedding dimension is taken to be $d_0 = k + 1$. Figure 3.3 shows a sample plot of E1(d).

3.2 Estimating Lyapunov exponents

In order to estimate Lyapunov exponents from our time series, we will use the popular Wolf algorithm [44]. The ideal way to estimate the maximum exponent is to monitor the longterm behaviour of a pair of trajectories in the phase space, however since we are dealing with observational data we only have one trajectory in the phase space. The algorithm deals with this issue by considering two points in the phase space to be from different trajectories if their temporal separation is large enough. Another problem which arises in monitoring longterm behaviour is that trajectories can undergo what is called a 'folding process' where two separated trajectories are brought back together. These folding processes are a result of the exponential separation of trajectories on a bounded attractor [44]. To deal with this problem the algorithm goes through a replacement step just before folding occurs, recording the current trajectories' contribution to the estimate and then replacing one of the trajectories. Replacement trajectories are chosen to be as close as possible to the remaining trajectory while trying to maintain the orientation of the original trajectories. Keeping these things in mind we denote the time of the ith replacement as t_i, the initial separation of the trajectories after the ith replacement as $L_0(i)$ and the separation of the trajectories at the end of the ith replacement as $L_1(i)$. With this notation we define our estimate

$$\hat{\lambda}_1 = \frac{1}{N} \sum_i \log \frac{L_0(i)}{L_1(i)}$$

where the range of summation depends on the number of replacement steps. For a visualization of the algorithm see Figure 3.4.

3.2.1 Implementation details

The above section is just an outline of how the Wolf algorithm works, however in practice there are a number of details which need to be ironed out before it can be applied. The following implementation details were specifically tailored to the model we will use to emulate the epileptic brain (see Section 4.1). The first step in the algorithm is to the choose the point u_0, this trajectory will be the fiducial trajectory which does not get replaced. Next we need to choose a second point which lies on a different trajectory and is as close to the first as possible. In order to do this we need to define what exactly makes a point on a different trajectory. Wolf considers two vectors to be from different trajectories if they are separated by at least one orbital period [44]. One method to estimate the mean orbital period (MOP) is to choose a time corresponding to the most powerful frequency present in spectral
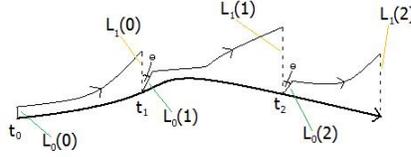
25

Figure 3.4: A diagram outlining the steps involved of the Wolf algorithm.

analysis, for our model we will take the value $MOP = 10$ based on the average time until folding seen in Figure 3.5. Also when choosing a second point we must have an acceptable range of initial separations. Due to computer precision limitations we must define an minimum acceptable separation, this stops us from choosing two points which the computer cannot distinguish from each other. Based on the software used for our simulations we will set this value equal to $DISTMIN = 10^{-5}$. Similarly we must set a maximum allowable separation between trajectories, this is because with finite amounts of data it is not always possible to find points as close as we would like. By looking at a plot of separation as a function of time (see Figure 3.5) we choose this value to be $DISTMAX = 0.25$, a value large enough to help the algorithm run smoothly.

The next step is to propagate both of the trajectories until just before folding occurs. This step is best done manually by a trained researcher, however this is not always practical. Since our eventual goal is to be able to predict seizures in real time we will be using an automated alternative of initializing replacements. The algorithm will automatically trigger a replacement every $\Delta t = 9$ steps. To choose a value for Δt, the average time between 'folding' was estimated visually from Figure3.5.

Once the decision has been made to make a replacement, the next step is to find a replacement point. When we are looking for a replacement point we use the same parameters MOP, Δt, DISTMIN and DISTMAX used in choosing the second trajectory. Wolf implements an additional constraint on the maximum allowable change in orientation, however there are reports that this is unnecessary when only estimating the maximum exponent in the presence of an ergodic measure [39].
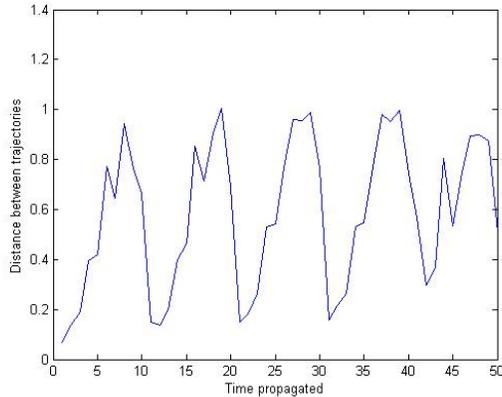
Figure 3.5: A plot of the distance between two trajectories over time. This plot is used to tune parameters in the Wolf algorithm. By looking at the amount of time between folding we can determine how long to let systems propagate between replacements. Also by looking at the distance between trajectories after folding, we can determine an appropriate value for the maximum allowable distance between points during replacement.

This replacement process is repeated until the fiducial trajectory reaches the end of the data. One other implementation detail that must be looked at is the possibility that no suitable replacement point is found. In this situation we go back several time steps to a point where there will hopefully be an acceptable replacement. For our implementation we will simply go back a single time step at a time until we reach a time when a replacement is found.

3.2.2 Alternate Interpretation

Now we consider an alternate interpretation of this estimate which leads to a more complete understanding of the exponents. First we define a set of points $\{d_i\}_{i=0}^{t}$ where d_i is the distance between the trajectories at time i. Based on the definition of Lyapunov exponents, the long term behaviour of these exponents should follow the relation

$$d_i = d_0 e^{\lambda i}.$$

Next we perform variance and mean stabilizing transformations, giving us a new set $\{w_i\}_{i=1}^{t}$ defined by

$$w_i = \log(d_i) - \log(d_{i-1}).$$

Based on the relation between d_t and λ we see that the long term average value
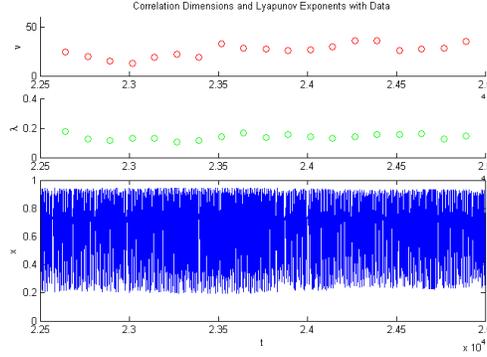
27

Figure 3.6: Plot of simulated model data along with estimates of the correlation dimension (top) and Lyapunov exponents (middle) calculated on 250 point windows.

of w_i should be $\bar{w} = \lambda$. However it can be shown that an estimate based on this average would be the same as the Wolf estimate (in the case of no replacements)

$$\lim_{t\to\infty} \frac{1}{t} \sum_{i=1}^{t} w_i = \lim_{t\to\infty} \frac{1}{t} \left(w_t - w_0\right) \tag{3.1}$$

$$= \lim_{t\to\infty} \frac{1}{t} \left(\log(d_t) - \log(d_0)\right) \tag{3.2}$$

$$= \lim_{t\to\infty} \frac{1}{t} \log\left(\frac{d_t}{d_0}\right) \tag{3.3}$$

3.3 Estimating Correlation Dimension

In order to estimate the correlation dimension ν, we estimate the slope of the linear relation between $\log(C(r))$ and $\log(r)$. The first step is to chose the scaling region, we do this by looking at plots of the $\log(\widehat{C(r)})$ against $\log(r)$. The scaling region is chosen to be the values of r for which the linear relationship between $\log(\widehat{C(r)})$ and $\log(r)$ appears to hold, for our model the region was between 0.05 and 0.25. We then estimate the correlation integral $C(r)$ for twenty evenly spaced values of r in the scaling region. We define our estimate of the correlation integral, sometimes called the correlation sum as

$$\widehat{C(r)} = \frac{1}{N^2} \sum_{i,j}^{N} \rho(r - ||x_i - x_j||).$$

We then estimate the slope of the linear relationship using the least squares estimator. See Figure 1.2 for a sample plot of $\log(C(r))$ and $\log(r)$.

28

Chapter 4

Simulations

The purpose of the study is to create a simple mathematical model which mimics the temporal dynamics of the epileptic brain. The model will have a time dependent parameter which spend most of its time in a range which will exhibit chaotic dynamics, this will be the interictal state. Occasionally the parameter will slowly wander towards a value which will turn the system into a lower complexity, this will be the preictal and ictal states. We will then look at the ability of Lyapunov exponents and correlation dimension to classify the preictal state as well as accurately predict the onset of the ictal state.

4.1 Model

The model that we used to model the epileptic brain is built around the logistic equation

$$x_{t+1} = k_t x_t (1 - x_t),$$

with $0 \leq x_0 \leq 1$. This system converges to a fixed point for values of k_t between 0 and 3. It starts to experience period doubling from 3 until approximately 3.57. At k_t=3.57 the system start to exhibit chaotic behaviour for most values, although certain values still exhibit periodic behaviour. For our model we use a time dependent parameter defined as

$$k_0 = 3.8,$$

$$k_{t+1} = \begin{cases} 3.8 & k_t > 3.9, k_t = 1.8 \\ 1.7 & 1.8 < k_t < 3.7 \\ k_t + Normal(0, 0.1) & 3.7 \leq k_t \leq 3.9 \\ k_t + 0.02 & k_t < 1.8 \end{cases}.$$
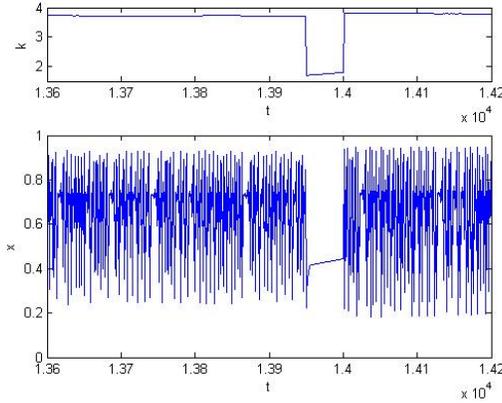
Figure 4.1: A sample plot of the data (bottom) and the parameter (top) of our model when a SLE occurs.

For values of k_t between 3.75 and 3.9 the model is considered to be in the interictal state, for k_t between 3.7 and 3.75 the model is considered to be in the preictal state and when k_t drops below 3.7 the model enters the ictal state which will be called a seizure like event (SLE). To see a plot of data simulated from this model see Figure 4.1. We generate 30 datasets using this model for t=1..10000 and values of x_0 generated from a uniform[0,1] distribution. The number of SLEs that occurred in each set of data ranged from one to six, with a median value of two.

4.2 Phase Space Reconstruction

The first step towards estimating the maximum Lyapunov exponent and correlation dimension is to reconstruct the phase space using the 'Method of Delays' described in Section 3.1. In order to do this we must estimate the parameters d_0 and τ.

In order to choose a value of τ for this model we use the mutual information method from Section 3.1.1 on N=250 point segments of interictal data. In total there were 4605 segments of interictal data with a mean estimated value of 10.6 and a median of 10. For our model we take the value of the time lag τ to be 10.

Now that we have fixed our value of τ we use Cao's Method to estimate the minimum embedding dimension. Similar to when we estimated τ we will estimate d_0 for many segments of interictal data, except this time we will use segments that are N=1000 points long. In order to determine the minimum embedding dimension plots of E1(d) were looked at individually and the value d_0 was manually chosen.
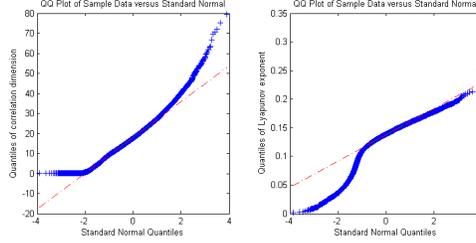
Figure 4.2: QQ plots of the correlation dimensions and Lyapunov exponents. The correlation dimension does not deviate much from the normal quantiles, however the Lyapunov exponents are quite different for the low quantiles.

Analysis of 100 segments gives a mean value of 6.4 and a median of 6, we take the value of our minimum embedding dimension to be $d_0 = 6$.

4.3 Classification

Classifying preictal states is actually quite different from predicting SLEs. When we attempt to classify states, we perform a post hoc analysis of the data with complete knowledge of the model parameters. On the other hand, prediction is done using only current and past data with no knowledge of the parameters. While these analyses are not directly useful in the prediction of SLEs, they can be interpreted as a measure of predictability. By performing these classifications using different methods, we can determine which method will be the best at predicting SLEs.

In order to classify the preictal states we will use Lyapunov exponents and correlation dimensions, calculated on 250 and 500 point windows. Let ψ be the quantity we are calculating to classify states, then define the sample mean and standard deviation of the quantity during the interictal period as μ_ψ and σ_ψ respectively. Figure 3.6 contains a plot of interictal data along with the corresponding estimates of the correlation dimensions and Lyapunov exponents. The estimates of the correlation dimension seem to be quite large, this is likely an artifact created by the small amounts of data that is being used in the computations. However this issue is not addressed because it is the ability of the estimates to predict seizures that is of interest, not their actual value. We assume that during the interictal state ψ is distributed as $Normal(\mu_\psi, \sigma_\psi)$. For each SLE the preictal event is defined to be all the window of data just before the SLE. We define our test statistic as

$$Z_{obs} = \frac{\left| \mu_\psi - \bar{\psi} \right|}{\sigma_\psi}.$$

31

Classification Results

	Within Trial	Between Trail
LE (250)	2%	2%
LE (500)	0%	0%
CD (250)	19%	23%
CD (500)	23%	14%

Table 4.1: Results of preictal classification

Under our assumptions this quantity follows $Normal(0,1)$, and hence it is said to classify the preictal state if $Z_{obs} > Z_{crit} = 1.645$. The normal quantile plots in Figure 4.2 suggest the assumption may not be valid. The correlation dimension deviates from normality at the tails and the since it is strictly positive there is some quantile bunching around zero. The Lyapunov exponents appear to be normally distributed on the upper half of the quantiles, however they deviate greatly from normality below the median. The classifications were run twice, the first with μ_ψ and σ_ψ calculated for each of the 30 datasets (within subjects) and then calculated based on all the data (between subjects). The results of the classifications can be found in Table 4.3, from the table we can see that the correlation dimension on 250 point windows does the best job classifying the preictal states. Since there was no measurement of false positives in this classification, it may be that they do not do a good job predicting SLEs or that there is much room for improvement by parameter tuning.

4.4   Prediction

Based on the results of the classification analyses, we will attempt to predict the SLEs using correlation dimensions calculated on 250 point windows. The first time series will be used in order to establish an initial estimate of μ_ψ and σ_ψ. We shall identify windows as being preictal in the same way we did in the classification section, except that we take $Z_{crit} = NUMBER$. A SLE will be predicted if we classify a preictal state in 3 consecutive windows. A prediction is said to be a false positive if there are 2 consecutive windows that are classified as interictal, before a SLE has occurred. This algorithm successfully predicted 34 out of 68 (or 50%) SLE's, however it falsely predicted seizures 372 out of 406 (or 91%) of the time.

Chapter 5

Conclusions

Using nonlinear quantities to perform predictions on complex time series is a method which has many complications, limiting the class of problems where it can be effectively implemented. There are difficulties that arise when calculating nonlinear quantities as well as a very large number of parameters which must be tuned in order to get reliable estimates. Moreover, ergodicity plays a large role in estimating these nonlinear quantities and it is very difficult to test for it in time series data. Another important aspect of these analyses is having a useful criteria for deciding when they are appropriate, namely low dimensional nonlinear systems. Tests which claim that data is generated by a nonlinear process can be unreliable and are not sufficient to validate the use of nonlinear methods because they do not address dimensionality.

There are a number of different tests for nonlinearity, however they all share some common problems. A linear process can be arbitrarily complicated and hence the null hypothesis that a process is linear is infinite dimensional. This means that for some linear models a test will have very little power to reject the null hypothesis with only a finite amount of data. Another potential complication with these tests is that they assume that the process is stationary. If this assumption is not valid, then depending on the nature of a nonstationarity it could cause the test to falsely classify the process as nonlinear. There is also a decision to be made about the form of the alternative hypothesis. Tests which specify a model in the alternate hypothesis end up restricting the types of nonlinearities they can identify, however tests which do not specify an alternate model often require large amounts of data and have higher computation times.

Ergodicity is instrumental for computing statistics on dynamical systems. It plays the role of a law of large numbers, allowing us to estimate the spatial mean of the system despite the time dependence in the data. Ergodicity also allows us to

generalize results across the entire dynamical system instead of restricting them to a single trajectory. When reconstructing the phase space of a dynamical system, it is ergodicity that ensures us that we are reconstructing the important regions of the system. All these reasons combined with the fact that it is very difficult to determine whether a system is ergodic from time series data means that ergodicity must always be a concern when predicting using nonlinear quantities.

There are many problems and parameters which must be dealt with in order to estimate nonlinear quantities such as Lyapunov exponents and correlation dimensions. The first step in estimating either quantity is to reconstruct the phase space of the dynamical system, this involves choosing the parameters τ and d_0 from the method of delays. When estimating Lyapunov exponents, geometric 'folding' and the need for two trajectories need to be dealt with. To do this algorithms like Wolf's [44] must be used, however this adds a number of other parameters which must be estimated for each model. When estimating the correlation dimension, the scaling region is another parameter which must be estimated. It takes a lot of data and computation time to get good estimates of the correlation integrals for many values of r.

Due to these complications with a nonlinear dynamics approach, perhaps a more statistical approach to prediction in complex time series may give better results. The statistical properties of less complex quantities such as variance are much better understood than those of nonlinear quantities. As well, estimates of these quantities are much easier to compute and they have lower estimate variance. Another benefit of simple models is that the effects of noise on such predictors are better understood, this is especially important in EEG data where noise is commonplace. Finally, it is important to remember that when dealing with prediction simpler models often get better results, even when more complicated models seem to fit better.

APPENDICES

Appendix A

Matlab Code

A.1 Phase Space Reconstruction

```
function PSpace=getPhase(D,t,d)
%Input is (data, tau, d0)

PSpace=zeros(length(D)-(d-1)*t,d);
for i=1:d
  PSpace(:,i)=D(1+(i-1)*t:length(D)-(d-i)*t);
end
end
```

A.2 Wolf Algorithm

```
function [L] = wolfLE(D)
% D is a vector of data

tau=10; % Parameters for method of delays
d0=6;

distMAX=0.4; % Parameters for algorithm implementation
distMIN=0.0001;
MOP=10;
dt=8;

P=getPhase(D,tau,d0); % Phase space reconstruction
N=length(P(:,1));
M=floor((N-1)/dt);

t=1; % Initial values
```

```matlab
L=0;
n=0;
dcrit = 0;

% Finding initial point of second trajectory
for j=1:N-dt
    if norm(P(t,:)-P(j,:)) < distMAX && norm(P(t,:)-P(j,:)) > distMIN
     && abs(t-j) > MOP
        dcrit=norm(P((t),:)-P(j,:));
        n=j;
        break
    end
end
for j=n+1:N-dt
    if norm(P(t,:)-P(j,:)) < dcrit && norm(P(t,:)-P(j,:)) > 0.0001
        n=j;
        dcrit=norm(P(t,:)-P(j,:));
    end
end

% Performs algorithm
for i=1:M
    L1=norm(P(t,:)-P(n,:));
    n=n+dt;
    t=t+dt;
    L2=norm(P(t,:)-P(n,:));
    L=L + log2(L2/L1);
    n=replacement(P,n,t,0);
end
L=L/(M*dt);
end




function n=replacement(P,nOld,t,h)
% P is the phase space
% nOld is the point being replaced
% t is the point of the fiducial trajectory
% h is a parameter for when a replacement point can't be found

dt=8;    % Algorithm parameters
distMIN=0.0001;
distMAX=0.4;
MOP=10;
```

```matlab
N=length(P(:,1)); %Initial values
dcrit=0;
n=0;

%Find repalcement point
for j=1:N-dt
    if norm(P((t-h),:)-P(j,:)) < distMAX && norm(P((t-h),:)-P(j,:)) > distMIN
     && abs((t-h)-j) > MOP
        dcrit=norm(P((t-h),:)-P(j,:));
        n=j;
        break
    end
end
for j=n+1:N-dt
    if norm(P((t-h),:)-P(j,:)) < dcrit && norm(P((t-h),:)-P(j,:)) > distMIN
     && abs((t-h)-j) > MOP
        dcrit=norm(P((t-h),:)-P(j,:));
        n=j;
    end
end

%If no replacement found, move t back two points
if dcrit == 0
    n=replacement(P,nOld,t,h+2);
end
end
```

A.3   Estimating correlation dimension

```matlab
function B=CorrDim(D,t,d)
%Input is data, tau, d0

P=getPhase(D,t,d); % Initial values
r=0.05:0.02:0.25;
N=length(r);
c=zeros(1,N);

%Calculate correlation dimension for all values of r
for i=1:length(r)
    c(i)=C(P,r(i));
end

% Least squares estimate of slope
```

```matlab
B=(N.*sum(log(c).*log(r)) - sum(log(r)).*sum(log(c)))/(N.*sum(r.^2)-sum(r).^2);
end




function x=C(P,r)
% Input is phase space and r value

N=length(P(:,1)); % Initial values
count=0;

%Count points which are closer than r units
for i=1:N
    for j=1:N
        if norm(P(i,:)-P(j,:))<r
            count=count+1;
        end
    end
end
x=count/N.^2;
end
```

A.4   Surrogate Data

```matlab
function s=getSurrogate(x)
% Input original data

% Properties of original data
N=length(x);
m=(N-1)/2;
xBar=mean(x);
F=fft(x)/sqrt(2.*pi.*N);
A=zeros(1,m);
for j=1:m
    A(j)=abs(F(j+1));
end

%Randomized phase
p=rand(1,N).*(2*pi);

%Create surrogate data
s=zeros(1,N);
for j=1:N
```

```
        count=0;
        for k=1:m
            count=count+2.*A(k).*cos(2.*pi.*j.*k/N+p(k));
        end
        s(j)=xBar + sqrt(2*pi/N).*count;
end
```

A.5   Surrogate Data Test

```
function [Z]=Surrogate(x,M,r)
%Input is data and number of surrogate datasets
% and value of r in correlation integral
%Output is test statistic

N=length(x); %Initial values
d=6;
t=10;
y=zeros(M,N);
Q=zeros(1,M);

%Generate surrogate data
for i=1:M
    y(i,:)=getSurrogate(x);
end

%Calculate R for real data
P=getPhase(x,t,d);
R=C(P,r);

%Calculate R for surrogate data
for i=1:M
    P=getPhase(y(i,:),t,d);
    Q(i)=C(P,r);
end

%Calculate test statistic
Z=abs(R-mean(Q))/sqrt(var(Q));
```

A.6   Keenan's Test

```
function Z=Keenan(Data,M)
%Input is data and order of models
```

```
%Output is test statistic

N=length(Data); %Initial Values
X=zeros(N-M,M+1);

%Fitting linear model and
% estimating residuals
for i=M+1:N
    X(i-M,:)=[1,Data(i-M:i-1)];
end
Y=Data(M+1:N);
B=inv(X'*X)*X'*Y';
Yhat=X*B;
e=Y'-Yhat;

%Fitting quadratic model and
% estimating residuals
Y2=Yhat.^2;
B2=inv(X'*X)*X'*Y2;
Yhat2=X*B2;
e2=Y2-Yhat2;

%Fitting linear residuals to
% quadratic ones
Y3=e;
X3(:,1)=zeros(1,N-M)+1;
X3(:,2)=e2;
B3=inv(X3'*X3)*X3'*Y3;

%Calculate test statistic
n=B3(2).* sum(e2.^2).^(1/2);
F=(n.^2.*(N-2*M-2))/(sum(e.^2)-n.^2);
```

References

[1] F. Balibrea. *Chaos, Periodicity and Complexity in Dynamical Systems.* Springer Berlin, 2006. 3

[2] G. Benettin, L. Galgani, A. Giorgilli, and J. Strelcyn. Lyapunov characteristic exponents for smooth dynamical systems and for hamiltonian systems; a method for computing all of them. part 1: Theory. *Meccanica*, pages 9–20, 1980. 6

[3] W.A Brock, J.A Scheinkman, W.D Dechert, and B LeBaron. A test for independence based on the correlation dimension. *Econometric Reviews*, 15:197–235, 1996. 14

[4] D.S. Broomhead and G.P. King. Extracting qualitative dynamics from experimental data. *Physica D: Nonlinear Phenomena*, 20:217–236, 1986. 23

[5] L. Cao. Practical method for determining the minimum embedding dimension of a scalar time series. *Physica D*, 110:43–50, 1997. 24

[6] M Casdagli, L Iasemidis, R Savit, R Gilmore, S Roper, and J.C. Sackellares. Non-linearity in invasive eeg recordings from patients with temporal lobe epilepsy. *Electroencephalography and clinical Neurophysiology*, 102:98–105, 1997. iii, 15

[7] K. Chan. On the validity of the method of surrogate data. In C. Cutler and D. Kaplan, editors, *Nonlinear Dynamics and Time Series*, pages 77–97, 1997. 13, 16

[8] W.S Chan and H Tong. On tests for non-linearity in time series analysis. *Journal of Forecasting*, 5:217–228, 1986. 14, 16

[9] Alan W.L. Chiu, Sarit Daniel, Houman Khosravani, Peter L. Carlen, and Berj L. Bardakjian. Prediction of seizure onset in an in-vitro hippocampal slice model of epilepsy using gaussian-based and wavelet-based artificial neural networks. *Annals of Biomedical Engineering*, 33(6):798–810, 2005. 12, 18

[10] C.I Chueshov. *Introduction to the Theory of Infinite-Dimensional Dissipative Systems.* ACTA, 2002. 1

[11] Alan W L Chui, Shokrollah S Jahromi, Houman Khosravani, Peter L Carlen, and Berj L Bardakjian. The effects of high-frequency oscillations in hippocampal electrical activities on the classification of epileptiform events using artificial neural networks. *Journal of Neural Engineering*, 3:9–20, 2006. 18

[12] Alan W.L. Chui, Eunji E. Kang, Miron Derchansky, Peter L. Carlen, and Berj L. Bardakjian. Online prediction of onsets of seizure-like events in hippocampal neural networks using wavelet artificial neural networks. *Annals of Biomedical Engineering*, 34(2):282–294, 2006. 18

[13] D.M Durand and M. Bikson. Suppression and control of epileptiform activity by electrical stimulation: A review. *Proc IEEE*, 89:1065–1082, 2001. 13

[14] C.E Elger and K. Lehnertz. Seizure prediction by non-linear time series analysis of brain electrical activity. *European Journal of Neuroscience*, 10:786–789, 1998. 13

[15] G.W. Frank, T. Lookman, M.A.H. Nerenberg, C. Essex, J. Lemieux, and W. Blume. Chaotic time series analyses of epileptic seizures. *Physica D*, 46:427–438, 1990. 17

[16] P. Grassberger and I. Procaccia. Characterization of strange attractors. *Phys. Rev. Lett.*, 50:346–349, 1983. 7, 23

[17] L.D. Iasemidis, P. Pardalos, J.C. Sackellares, and D.S. Shiau. Quadratic binary programming and dynamical system approach to determine the predictability of epileptic seizures. *Journal of Combinatorial Optimization*, 5(1):9–26, 2001. 12, 17

[18] L.D. Iasemidis, J.C. Sackellares, H.P. Zaveri, and W.J. Williams. Phase space topography and the lyapunov exponent of electrocortigrams in partial seizures. *Brain Topography*, 2(3):187–201, 1990. 17, 22

[19] L.D. Iasemidis, D. Shiau, W. Chaovalitwongse, J.C. Sackellares, P.M. Pardalos, J.C. Principe, P.R. Carney, A. Prasad, B. Veeramani, and K. Tsakalis. Adaptive epileptic seizure prediction system. *IEEE Transactions on Biomedical Engineering*, 50(5):616–627, 2003. 17

[20] B.H. Jansen. Quantitative analysis of electroencephalograms: Is there chaos in the future? *Int J Biomed Comput*, 27:95–123, 1991. 13

[21] D.M Keenan. A tukey nonadditivity-type test for time series nonlinearity. *Biometrika*, 72:39–44, 1985. 13, 16

[22] M.B. Kennel, R. Brown, and H. Abarbanel. Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Physical Review A*, 45:3403–3411, 1992. 24

[23] K. Lehnertz and C.E Elger. Spatio-temporal dynamics of the primary epileptogenic area in temporal lobe epilepsy characterized by neuronal complexity loss. *Electroencephalography and clinical Neurophysiology*, 95:108–117, 1995. 12, 17

[24] K. Lehnertz and C.E Elger. Can epileptic seizures be predicted? evidence from nonlinear time series analysis of brain electrical activity. *Phys. Rev. Lett.*, 80:5019–5022, 1998. 17

[25] Xiaoli Li, Gaoxiang Ouyang, Xin Yao, and Xinping Guan. Dynamical characteristics of pre-epileptic seizures in rats with recurrence quantification analysis. *Physics Letters A*, 333:164–171, 29 November 2004. 18

[26] P.E McSharry, L.A Smith, and L. Tarassenko. Prediction of epileptic seizures: are nonlinear methods relavent? *Nature Medicine*, 9, 2003. 18

[27] J. Milnor. On the concept of attractor. *Comm. Mathematical Physics*, 99:177–195, 1985. 3, 4

[28] H.R. Moser, B. Weber, H.G. Wieser, and P.F. Meier. Electroencephalograms in epilepsy: analysis and seizure prediction within the framework of lyapunov theory. *Physica D*, 130:291–305, 1999. 17

[29] David G. Myers. *Psychology*. Worth Publishers, seventh edition, 2004. 8, 9

[30] M. Oscar-Berman and K. Marinkovic. Alcoholism and the brain: An overview, 2004. 10

[31] V. Oseledec. A multiplicative ergodic theorem. lyapunov characteristic numbers for dynamical systems. *Trans. Moscow Math. Soc.*, 19, 1968. 6

[32] Gaoxiang Ouyang and Xiaoli Li. Nonlinear similarity analysis for epileptic seizures prediction. *Nonlinear Analysis*, 64(8):1666–1678, 2006. iii, 12, 17

[33] Gaoxiang Ouyang, Xiaoli Li, Yan Li, and Xinping Guan. Application of wavelet-based similarity analysis to epileptic seizures prediction. *Comput. Biol. Med.*, 37(4):430–437, 2007. 18

[34] J.P. Eckmann P. Collet. *Iterated maps on the interval as dynamical systems*. Boston: Birkhauser, 1980. 3

[35] M. Palus. Testing for nonlinearity using redundancies: quantitative and qualitative aspects. *Physica D*, 80:186–205, 1995. iii, 14, 15

[36] M. Palus. Nonlinearity in normal human eeg: cycles, temporal asymmetry, nonstationarity and randomness, not chaos. *Biological Cybernetics*, 75:389–396, 1996. iii, 14, 15

[37] Z. Psaradakis and N. Spagnola. Power properties of nonlinearity tests for time series with markov regimes. *Studies in nonlinear dynamics and econometrics*, 6, 2002. 13, 14

[38] P. Rapp, T. Bashore, J. Martinerie, A. Albano, I. Zimmerman, and A. Mees. Dynamics of brain electrical activity. *Brain Topography*, 2:99–118, 1989. 22

[39] M.T Rosenstein, J.J Collins, and C.J De Luca. A practical method for calculating largest lyapunov exponents from small data sets. *Physica D*, 65:117–134, 1993. 26

[40] C.E. Shannon. A mathematical theory of communication. *The Bell Systems Technical Journal*, 27:623–656, 1948. 22

[41] F Takens. Detecting strange attractors in turbulence. In D.A. Rand and L.S. Young, editors, *Proceedings of the Symposium on Dynamical Systems and Turbulence. Lecture Notes in Mathematics*, pages 366–381, Berlin, 1980. Springer-Verlag. 20

[42] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, and J.D. Farmer. Testing for nonlinearity in timeseries: the method of surrogate data. *Physica D*, 58:77–94, 1992. 14, 15, 16

[43] M. Winterhalder, B. Schelter, T. Maiwald, A. Brandt, A. Schad, A. Schulze-Bonhage, and J. Timmer. Spatio-temporal patient-individual assessment of synchronization changes for epileptic seizure prediction. *Clinical Neurophysiology*, 117:2399–2413, 2006. 18

[44] A. Wolf, J.B. Swift, H.L. Swinney, and J.A. Vastano. Determining lyapunov exponents from a time series. *Physica D*, 16:285–317, 1985. 6, 15, 25, 34