Analysing Perturbations and Nonstationarity in Data Series using Techniques Motivated by the Theory of Chaotic Nonlinear Dynamical Systems

D. J. Downing V. Fedorov W. F. Lawkins M. D. Morris
G. Ostrouchov

Abstract

Large data series with more than several million multivariate observations, representing tens of megabytes or even gigabytes of data, are difficult or impossible to analyze with traditional software. The shear amount of data quickly overwhelmes both the available computing resources and the ability of the investigator to confidently identify meaningful patterns and trends which may be present. The purpose of this research is, first, to give a meaningful definition to "large data set analysis" and, second, to described and illustrate a technique for identifying unusual events in large data series. The technique presented here is based on the theory of nonlinear dynamical systems.

1 Introduction

In this paper we consider the problem of identifying unusual segments of data contained in large data series, where "unusual" is also intended to mean the fraction of usual to unusual events is small. For example, suppose we are observing a process that resides in a relatively quiescent mode for significant periods of time, but occasionally becomes excited for short periods before relaxing to the background state. We describe a methodology for detecting such transients and present results for two example data series.

The methods discussed in this article are based on the theory of nonlinear dynamical systems. Due to fundamental theoretical work of Takens[13] and Mañé[11], that theory can be exploited using scalar time series data to develop nonlinear statistical methods for analysing nonlinear processes. The methods described here are examples of applications based on Takens' and Mañé's theoretical work.

We referred above to "large data series" and stated that the problem considered in this paper is related to such datasets. As a consequence of size, we assume the dataset can not be analysed at one time as a whole, but rather has to be analysed in segments. The strategy adopted here is to partition the large data series into a series of pieces and then to analyse one piece of the partition at a time for unusal segments.

The research presented here is part of a project to develop new algorithms and software for identifying meaningful events in large datasets[3, 4]. In generic terms, the goal of the analysis is to identify unusual events for the immediate purpose of saving those segments of the data series, making the assumption that the resulting collection of unusual events is ammenable to classical statistical analysis methods. As we only want to be confident unusual events are identified, it is not important that the method is accurate in the sense it is either an accurate model of the background process or the transients, but rather only that it can accurately distinguish between the two. Further, because of the large dataset assumption, it is important that the method is "fast."

We want to give a qualitative description of the technique developed here to identify unusual events. Suppose the process being observed is represented by a time dependent trajectory in state space. Further, suppose most of the time that trajectory is confined to a limited region in state space, which is referred to here as an attractor. We shall say the attractor represents the "background" process. Now, suppose occasionally the trajectory leaves that attractor and moves about in an extended region of state space before falling back to the attractor. Such a trajectory segment is called a "perturbation" and is associated with an unusual event. In addition, if that attractor is moving or changing with time, we shall say that backgound process is nonstationary. Following this description, the development of the time series analysis methodology presented here can be organized into three components. The first component is: Assuming the background process is stationary, how can we identify the attractor in state space associated with it? The second is: How do we define a perturbation? And the third is: How do we define and quantitatively measure a slow change in the background process?

The methodology developed in this article is demonstated using two example time series, one representing an atmospheric process and the other a neurophysiological process. The atmospheric process is represented by a time series collected under the auspices of the Atmospheric Radiation Measurement (ARM) project [5]. Fig. 1a displays a segment of 20,000 observations selected at random from that record. The time series used here is measurement of the liquid water content of the atmosphere. The background process in this case corresponds to a relatively clear day with dry conditions. Perturbations include cloud, rain, and fog events as well as some instrument malfunction events. The record covers a period of approximately 230 days, with observations at 20 second intervals, or 4320 observations per day, but there are gaps of varying length scattered across that period where there is no data because, for whatever reason, the instrument was turned off. There are also a lot of single point instrument malfunctions, which is manifest in Fig. 1a by a series of points along $y \approx -2.5$.

The second time series is taken from one channel of a sixteen channel electroencephalogram (EEG) record for an epileptic patient [8]. Fig. 1b displays a segment of 5,000 observations. The record covers a continuous period of 23 minutes, with 512 observations per second. EEG records typically include a great deal of "artifact," representing head movement, eye movement, muscle tension, grinding teeth, etc., in addition to unmasked neurophysiological activity. If we associate neurophysiological activity with the background process, then artifact is a perturbation relative to that background process.

The remained of this article is organized as follows. In Section 2 we describe an analysis that is part of the methodology used to model nonlinear dynamical systems. This analysis is applied to the example ARM and EEG time series. In Section 3 we describe that part of the theory of nonlinear dynamical processes that provides the foundation for using time series measurements to model such systems, and we outline the process for constructing models. In Section 4 we describe the technique addressed in this article for analysing large data series for unusual segments and demonstrate the technique using the ARM and EEG

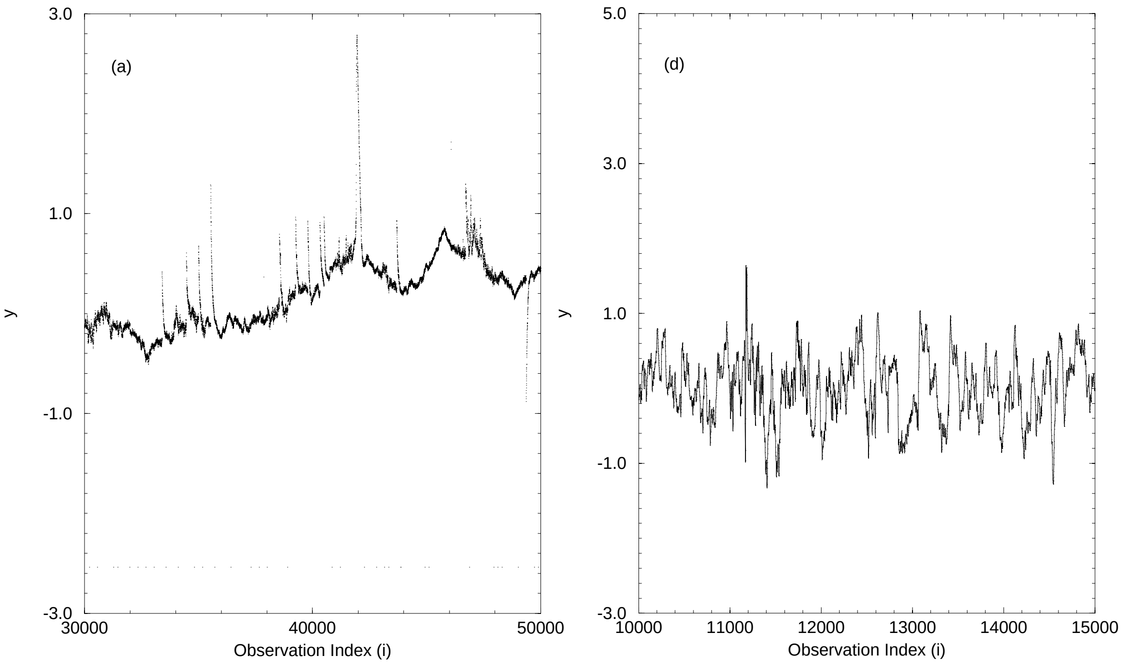

Figure 1: Example segments from the ARM and EEG datasets: (a) scaled ARM data; (b) scaled EEG data.

datasets. Finally, in Section 5 we summarize the results.

2 Data Conditioning and Preliminay Analysis

In this section we carry out a preliminary analysis that is part of the methodology that has been developed for constructing models of nonlinear dynamical processes. The analysis is applied to training sets of data that have been extracted from the ARM and EEG records using the condition that the data appears to be relatively perturbation free. Fig. 2 displays a random pair of segments from those training sets. We note that both as-measured datasets are translated so that the average value is zero and scaled so that the average squared value is one.

2.1 Low-Pass Filters

Low-pass filtering plays an important role in conditioning time series data. The significance of such conditioning for constructing empirical nonlinear dynamical models is discussed later. Here we describe the specific low-pass filter used for the analyses presented in this article.

Consider the first-order, linear low-pass filter

$$\frac{1}{\omega}\frac{dy}{dt} + y = x,$$

where x is input, y is output, $\omega = 2\pi f_c$, and f_c is the cutoff frequency. Integrating this

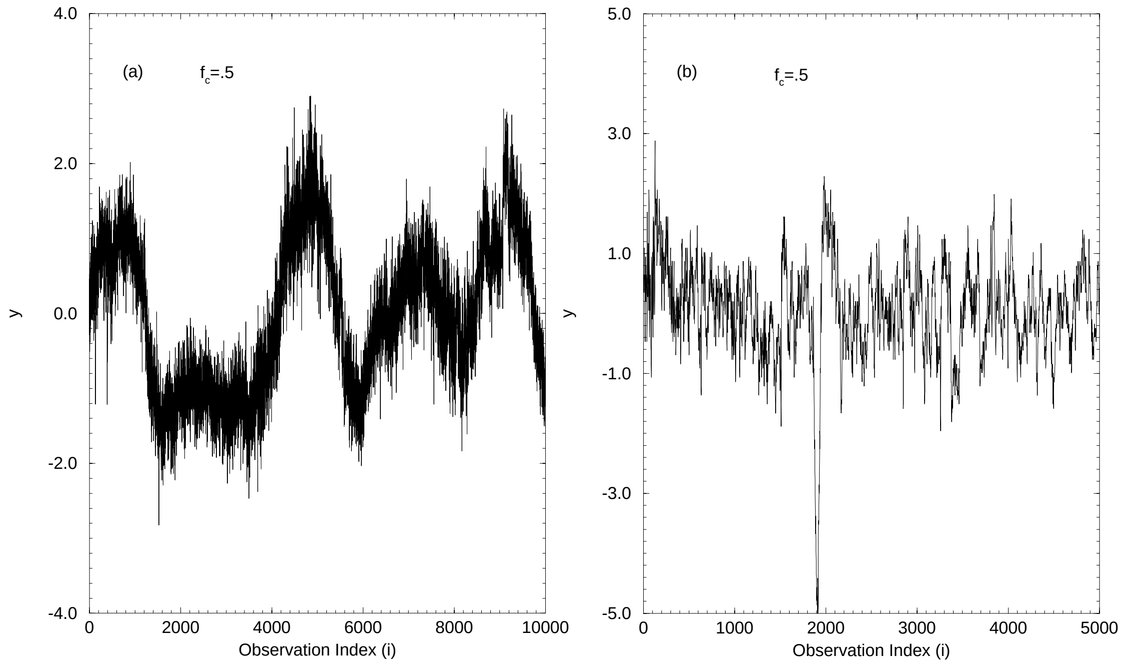

Figure 2: Example segments from the ARM and EEG training datasets: (a) scaled ARM data; (b) scaled EEG data.

equation over the interval $[t_i, t_i + t_s]$ produces

$$\begin{aligned} y_{i+1} &= y_i \exp(-\omega t_s) + x_i[1 - \exp(-\omega t_s)], \\ &= a y_i + b x_i, \end{aligned} \tag{1}$$

where $x_i, i = 0, 1, 2, ...$ is a time series obtained from $x(t)$ using the sampling time t_s. Eq. (1) is referred to in control theory as a first-order lag, or first-order infinite impulse response, low-pass filter and is frequently used to simulate the behavior of measurement instuments[10]. We use the forth-order filter that results by applying the first-order filter Eq. (1) in series four times, which is to say output from the first stage is input to the second, output from the second stage is input to the third, and so forth.

Figure 3 displays the same segments of ARM and EEG data as shown in Fig. 2, but after applying the low-pass filter. The filter cutoff values displayed in Fig. 3 are normalized by the respective sampling frequencies. The normalized filter cutoff value shown in both frames of Fig. 2 is $f_c = .5$, which is half the sampling frequency. That value for the cutoff frequency is commonly known as the 'Nyquist' frequency and is used in measurement instrumentation to prevent 'aliasing.' The sampling time and sampling frequency for each of those time series, together with the dimensional value of f_c shown in Fig. 3, are:

$$\begin{array}{llll} \text{ARM}: & t_s = 20s & f_s = t_s^{-1} = .05\,Hz & f_c = .0025\,Hz \\ \text{EEG}: & t_s = \frac{1}{512}s & f_s = 512\,Hz & f_c = 50\,Hz \end{array} \tag{2}$$

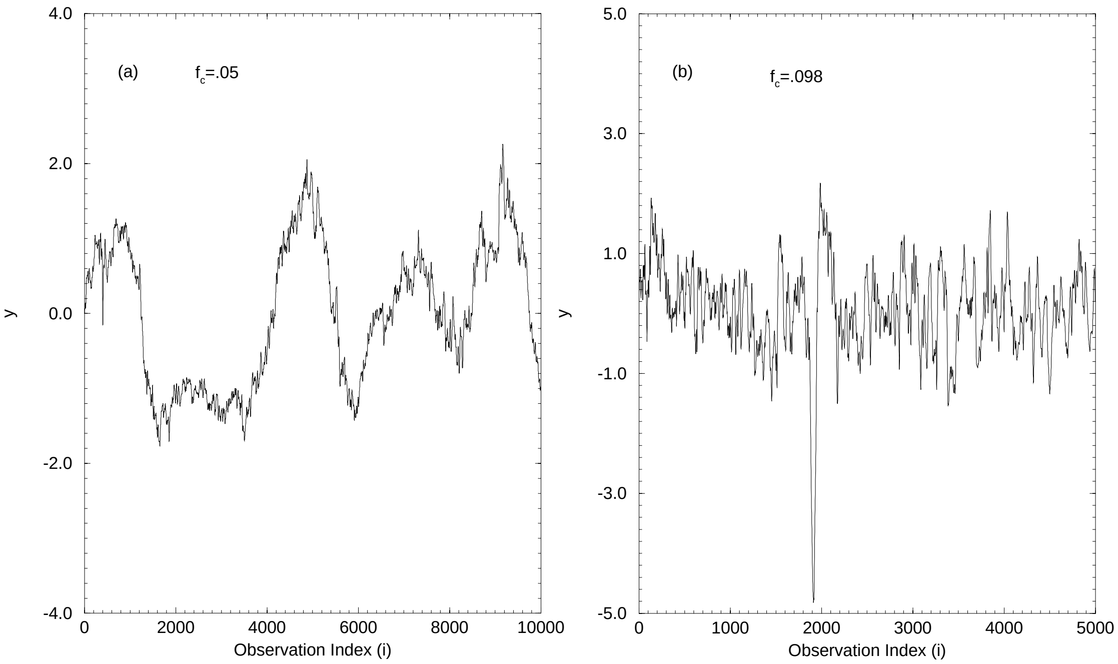

Figure 3: Example segments from the ARM and EEG training datasets: (a) scaled and filtered ARM data ($f_c = .05$); (b) scaled and filtered EEG data ($f_c = .098$).

2.2 Power Spectra and Autocovariance

Power specta analysis of discrete time series is well established and we refer to Blackman and Tukey [1] for detailed information on that topic. Here, we want to briefly review the relationship between the power spectrum and autocovariance function of a time series $y = \{y_i\}_{i=1}^{N}$. The autocovariance estimate for lag k is

$$c_k = \frac{1}{N} \sum_{i=1}^{N-k} (y_i - \bar{y})(y_{i+k} - \bar{y}) \ , \ k = 0, 1, ..., N-1 \ , \tag{3}$$

where $\bar{y}$ is the mean value of y. Suppose N is odd. Then we can write $N = 2q + 1$. The time series ϕ can be decomposed into Fourier modes according to the equations

$$y_i = \alpha_0 + \sum_{j=1}^{q} (\alpha_j \cos(2\pi f_j i) + \beta_j \sin(2\pi f_j i)) \ , \ i = 1, ..., N, \tag{4}$$

where $f_j = j/N$ and

$$\begin{aligned} \alpha_0 &= \bar{y} \ , \\ \alpha_j &= \frac{2}{N} \sum_{i=1}^{N} y_i \cos(2\pi f_j i) \ , \\ \beta_j &= \frac{2}{N} \sum_{i=1}^{N} y_i \sin(2\pi f_j i) \ . \end{aligned} \tag{5}$$

The power spectrum for the time series y is

$$P(f_j) = \frac{N}{2}(\alpha_j^2 + \beta_j^2)\ ,\ j = 1, 2, ..., q. \tag{6}$$

Finally, the sample spectrum $P(f)$ can be calculated from the sample autocovariance by the cosine transformation

$$P(f_j) = 2[c_0 + 2\sum_{k=1}^{N-1} c_k \cos(2\pi f_j k)]\ ,\ f_j = \frac{j}{N}\ . \tag{7}$$

In general, most of the variation in the autocovariance function occurs for relatively small values of lag k, so that it follows most of the energy in the power spectrum $P(f)$ for high frequencies derives from the autocovariance function for small values of k.

Figure 4 displays power spectra for the ARM and EEG data sets, both for the as-measured data and for the conditioned time series obtained using the low-pass filter described in the previous section. The choice of low-pass filter cutoff value for either example time series is guided by the desire to preserve the power in the dominant low frequency band and to minimize power above that band. The magnified frames in Fig. 4 support the view that those conditions are achieved by the respective choice of filter cutoff value. The time series segments shown in Fig. 2 and Fig. 3 illustrate the effect of data conditioning that results from reducing power in the spectrum above the dominant band.

2.3 Mutual Information

Mutual information is a nonlinear measure of the extent to which one random variable is a function of another[12]. Let (x, y) be an $\mathbb{R}^2$-valued random variable. Further, let $\rho(x, y)$ be the joint probability density of (x, y), and let $\rho(x), \rho(y)$ be the probability densities of x, y, respectively. Then, the mutual information of the random variables x, y is, by definition,

$$M(x, y) = \int_{\mathbb{R}^2} m(x, y)dxdy, \tag{8}$$

where

$$m(x, y) = \rho(x, y)ln\frac{\rho(x, y)}{\rho(x)\rho(y)}. \tag{9}$$

We note several properties of mutual information M. First, by definition, mutual information is symmetric in its arguments. Second, $M(x, y) \geq 0$ and $M(x, y) = 0$ if and only if x, y are independent. Third, if y is a function of x, then $M(x, y)$ is unbounded.

Consider now the time-series $\{y_i\}_{i=1}^N$, and suppose y is a measurement from a stationary stochastic process. For an arbitrarily fixed value of the time delay k, we define the bivariate $\mathbb{R}^2$-valued random variable (y_i, y_{i+k}) and the mutual information function associated with that time series by

$$M(k) = M(y_i, y_{i+k}). \tag{10}$$

The function $M(k)$ is a nonlinear measure of the dependence of two observations from the time series y separated by the time delay k. In general, mutual information is regarded as a more appropriate measure of independence versus dependence for a nonlinear process

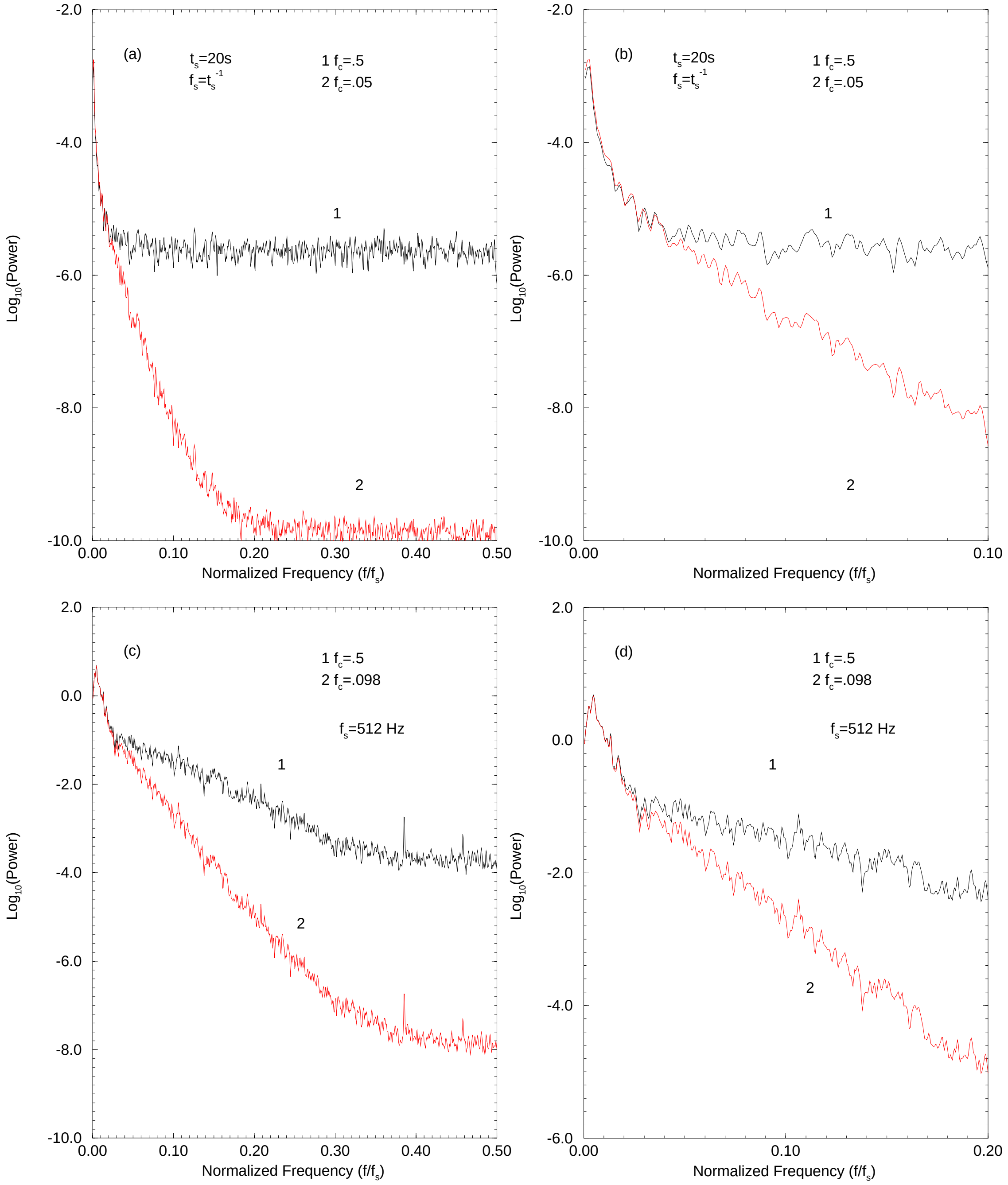

Figure 4: Power Spectra for ARM and EEG training datasets: (a) scaled ARM data; (b) magnified view of scaled ARM data; (c) scaled EEG data; (d) magnified view of scaled EEG data.

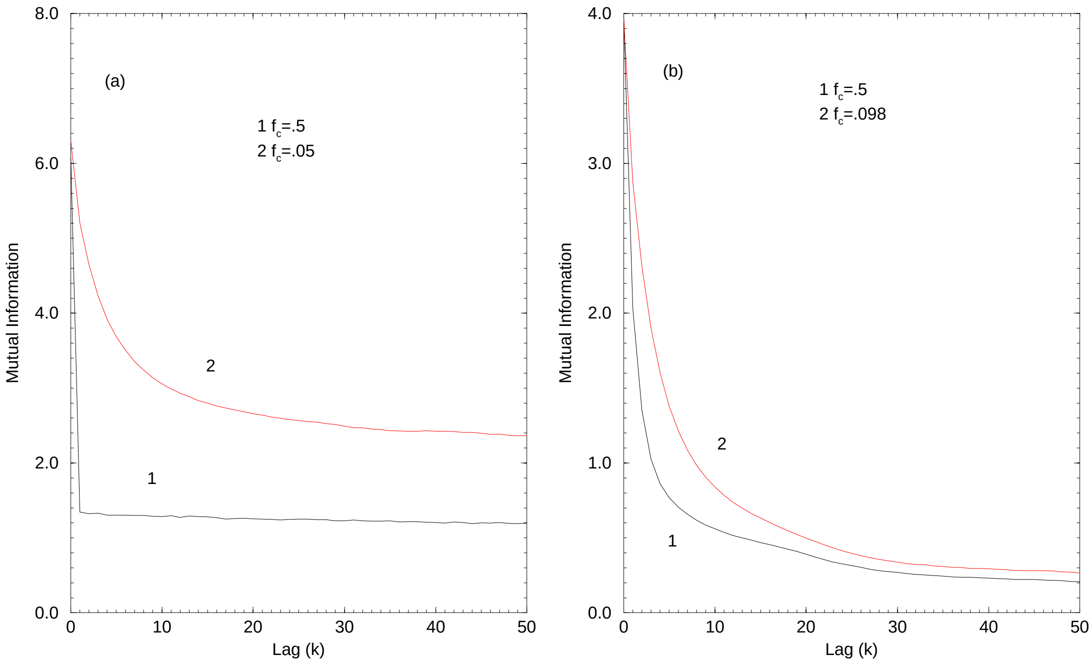

Figure 5: Mutual information for the ARM and EEG training datasets: (a) ARM data; (b) EEG data.

than the autocovariance function Eq. (3) [6, 7]. Like the autocovariance function, $M(k)$ in general varies most rapidly for small values of delay k, and, referring to the relationship between the autocovariance and power spectrum, the time scales associated with the most rapidly varying segment of mutual information translate to relatively high frequencies that are not resolved in the nonlinear process reconstruction described below.

Figure 5 displays mutual information for the ARM and EEG data, both for the as-measured and conditioned data sets. Correlation in the ARM as-measured data appears to be masked by white noise, but correlated structure is revealed in the conditioned data. There is some difference in the observed correlated structure between the as-measured and conditioned EEG data sets, but that difference is small compared to the ARM data. For both cases, the dominant contibution to high frequency content in the power spectra appears to be restricted to Lag values $k < 20$.

3 Nonlinear Processes and Time Series

In this section we review the definition of a nonlinear dynamical process and describe the theory that provides for constructing an approximate representation of such a process using time series. In addition, we review the methodology then used for determining a reconstruction.

3.1 Nonlinear Dynamical Processes

Using $E^{\bar{n}}$ to denote $\bar{n}$-dimensional Euclidean space, let

$$f : E^{\bar{n}} \to E^{\bar{n}}$$

be a diffeomorphism. Suppose $M \subset E^{\bar{n}}$ is a compact, $\bar{n}$-dimensional differentiable manifold and that f restricted to M is a diffeomorphism of M. Let A be a compact subset of M such that f maps A onto A. Further, let $U \supset A$ be an open subset of M such that

$$\lim_{i \to \infty} f^i(U) = A.$$

Finally, let p be an ergodic probability measure on A with respect to the transformation f, which is to say that, if $V \subset A$ is measurable with respect to p, then $p(f(V)) = p(V)$.

We refer to $E^{\bar{n}}$ as "state space" and to A as an "attractor." The point of view taken here is that nonlinear dynamical processes, including chaotic systems, are represented by triples of the form (A, f, p), that is, an attractor in state space, a diffeomorphic mapping of that attractor onto itself, and an ergodic probability measure with respect to that attractor and diffeomorphic mapping.

Suppose $a \in A$ is the state of the experimental system at the instant we begin to observe it. Then, $\{f^i(a)\}_{i=0}^{\infty}$ is a time series of states visited by the process. Now, suppose $y : A \to \mathbb{R}$ is a real-valued function, or observable, on A. Then,

$$y_i = y(a_i) = y(f^i(a)), \quad i = 0, 1, 2, \ldots,$$

is a real-valued time series of the experimental system.

The following is a theorem due to Takens[13] that constitutes the foundation for the time series analysis methods described here. As used in the statement of the theorem, "smooth" means at least C^2.

Theorem 1 *(Takens) Let $M \subset E^{\bar{n}}$ be a compact, differentiable manifold of dimension $\bar{n}$. Further, let $f : M \to M$ be a smooth diffeomorphism of M and let $y : M \to \mathbb{R}$ be a smooth, real-valued function on M. It is a generic property that the map*

$$\Phi : M \to E^{2\bar{n}+1}$$

defined by

$$\Phi(a) = (y(a), y(f^1(a)), \ldots, y(f^{2\bar{n}}(a))^T, \quad a \in M,$$

where T is matrix transpose, is a smooth embedding.

For some choice of values for the positive integers (n, k), consider the object $\tilde{A}_y \subset E_y^n$, where

$$\tilde{A}_y = \{\mathbf{y}_i\}_{i=0}^{\infty} \tag{11}$$

and $\mathbf{y}_i$ is defined by

$$\mathbf{y}_i = (y_i, y_{i+k}, y_{i+2k}, \ldots, y_{i+(n-1)k})^T. \tag{12}$$

Provided $n \geq 2\bar{n} + 1$, Takens' reconstruction theorem implies the mapping

$$\tilde{\Phi} : A \to \tilde{A}_y$$

defined by

$$\tilde{\Phi}(a_i) = \mathbf{y}_i, \quad i = 0, 1, 2, \ldots,$$

produces a smooth embedding $\tilde{\Phi} : A \to E_y^n$. Define

$$\begin{aligned} \tilde{f}_y &= \tilde{\Phi} \circ f \circ \tilde{\Phi}^{-1} \\ \tilde{p}_y &= p \circ \tilde{\Phi}^{-1}. \end{aligned}$$

Then, the triple $(\tilde{A}_y, \tilde{f}_y, \tilde{p}_y)$ constitutes a faithful representation of the dynamical system (A, f, p).

The pair of integers (n, k) used to define the vector $\mathbf{y}_i$ Eq. (12) are referred to as the embedding dimension and time lag, respectively. Choosing appropriate values for (n, k) is a primary objective for an analysis leading to a model $(\tilde{A}_y, \tilde{f}_y, \tilde{p}_y)$ of a nonlinear dynamical process.

3.2 Nonlinear Dynamical Process Model

Here we review a procedure introduced by Broomhead and King[2] as a mechanism for implementing the reconstruction process in an optimal manner. The application of that procedure, augmented by the use of low-pass filters, is described in detail by Lawkins, Daw, Downing, and Clapp [9].

Recall that the primary task is to determine appropriate values for the reconstruction parameters (n, k), where n is the embedding dimension and k is the time delay. Given values for those parameters, we set

$$\tilde{A}_y = \{\mathbf{y}_i\}_{i=1}^N \subset E_y^n \,, \tag{13}$$

where $y = \{y_i\}_{i=1}^N$ is the measured time series and

$$\mathbf{y}_i = (y_i, y_{i-k}, y_{i-2k}, \ldots, y_{i-(n-1)k})^T, \; i = 0, 1, 2, \ldots, N. \tag{14}$$

Before continuing, we define the following notation.

$$\begin{aligned} t_s &= \text{sampling time} \\ &= \text{time between } y_i \text{ and } y_{i+1} && (15) \\ t_w &= \text{window time scale} \\ &= (n-1) \times k \times t_s && (16) \\ f_w &= \text{window frequency} \\ &= t_w^{-1} && (17) \end{aligned}$$

Let the matrix Y be defined by $Y_i = \mathbf{y}_i$, that is, the i-th column of Y is $\mathbf{y}_i$. Then, define the matrix

$$K = \frac{1}{N} Y Y^T = \frac{1}{N} \sum_{i=1}^N \mathbf{y}_i \mathbf{y}_i^T.$$

The matrix K is symmetric, nonnegative. We denote the eigenvalue, eigenvector pairs of K by $\{(\lambda_\alpha, \psi_\alpha)\}_{\alpha=1}^n$, where

$$\lambda_1 \geq \lambda_2 \geq \ldots \lambda_n \geq 0 \ . \tag{18}$$

Define the matrix Ψ by

$$\Psi = (\psi_1, \psi_2, \ldots, \psi_n), \tag{19}$$

meaning that the j-th column of Ψ is the eigenvector ψ_j. Then the equation

$$\theta = \Psi(\mathbf{y}) = \Psi^T \mathbf{y} \tag{20}$$

defines an orthonormal transformation

$$\Psi : E_y^n \rightarrow E_\theta^n \tag{21}$$

that transforms

$$\tilde{A}_y = \{\mathbf{y}_i\}_{i=1}^N \ \stackrel{\Psi}{\rightarrow} \ \tilde{A}_\theta = \{\theta_i\}_{i=1}^N \ .$$

We want to make two observations regarding the transformation Eq. (21). The first is that, in general, the principal mode ψ_j is expected to have a total of j relative maxima and minima. Thus, we expect the j-th mode to have approximately $j/2$ waves over the time span of a point $\mathbf{y}_i$, which is the window time scale t_w Eq. (16). Consequently, the j-th coordinate in E_θ^n, which is the inner product of $\mathbf{y}_i$ with ψ_j according to Eq. (20), corresponds approximately to information in the time series resolved by the frequency

$$\begin{aligned} f_j &= [\frac{t_w}{j/2}]^{-1} \ , \\ &= j \times \frac{1}{2t_w} \ , \\ &= j \times \frac{1}{2} f_w \ . \end{aligned} \tag{22}$$

The second observation is that the eigenvalues Eq. (18) are estimates of the second moments of each coordinate value in the set $\tilde{A}_\theta$ and, as such, determine length scales in phase space E_θ^n for $\tilde{A}_\theta$. Those length scales are

$$\sqrt{\lambda_j} \ , \ j = 1, \ldots, n \ . \tag{23}$$

We select values for the reconstruction parameters (n, k) based primarily on the following two conditions. First, referring to Eq. (22), we want the lowest order frequencies f_j, $j = 1, \ldots, m$, where $m \leq n$, to resolve information in the time series, as represented by the power spectrum, considered to be significant. Second, we want the corresponding distribution of eigenvalues, λ_j, $j = 1, \ldots, m$, to resolve the state-space range of length scales associated with the frequencies f_j, $j = 1, \ldots, m$. Lawkins, et al. [9] examined those conditions using the correlation integral in addition to the power spectrum and distribution of eigenvalues to judged the quality of a reconstruction based on a pair of values (n, k). The two conditions cited above are evaluated through the correlation integral, first, by convergence of the

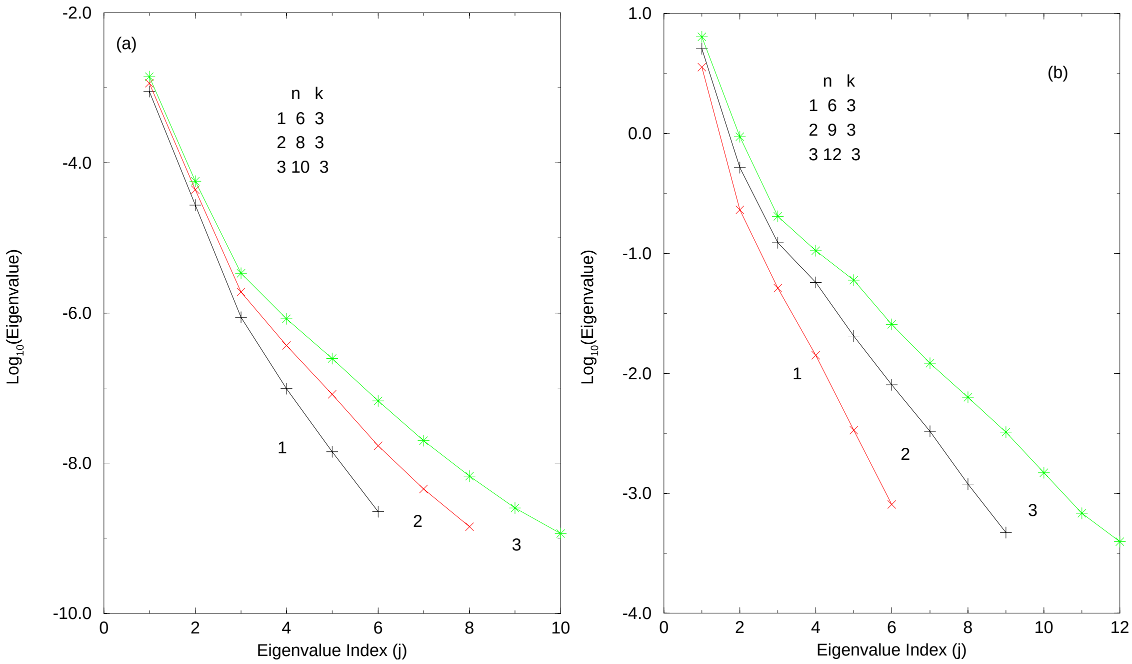

Figure 6: Eigenvalues for the ARM and EEG training datasets: (a) ARM data; (b) EEG data.

correlation dimension and, second, by the range of state-space length scales over which fractal structure is observed.

Recall the ordering Eq. (18). If $\sqrt{\lambda_{m+1}}$ is sufficiently small, the projection

$$P^m : E_\theta^n \to E_\theta^m \ , \tag{24}$$

where

$$E_\theta^m = L\{\psi_j\}_{j=1}^m$$

is the linear subspace of E_θ^n spanned by to the first m eigenvectors Eq. (19), will result only in eliminating small scale, noisy detail in the highest order coordinates. Referring again to Lawkins, et al. [9], the effect of that projection on the correlation integral is on state-space length scales less than the lower bound of observed fractal structure, and consequently, there is no lose in observed dimensionality of the process. We refer to the resulting reconstruction by the ordered triple of parameters (n, k, m).

In general, the low-pass filter described in Section 2.1 is used to reduce the frequency content of the time series for frequencies greater than the upper bound of the frequency band in the power spectrum considered to include significant information about the process of interest. Recall from Section 2.2 that the extent of the most rapidly varying segment of mutual information also provides an estimate of the upper bound of that frequency band. The impact of such low pass filtering is to reduce the potential noise-like effect of high frequencies on the reconstruction and, hence, on the analyses that depend on the reconstruction.

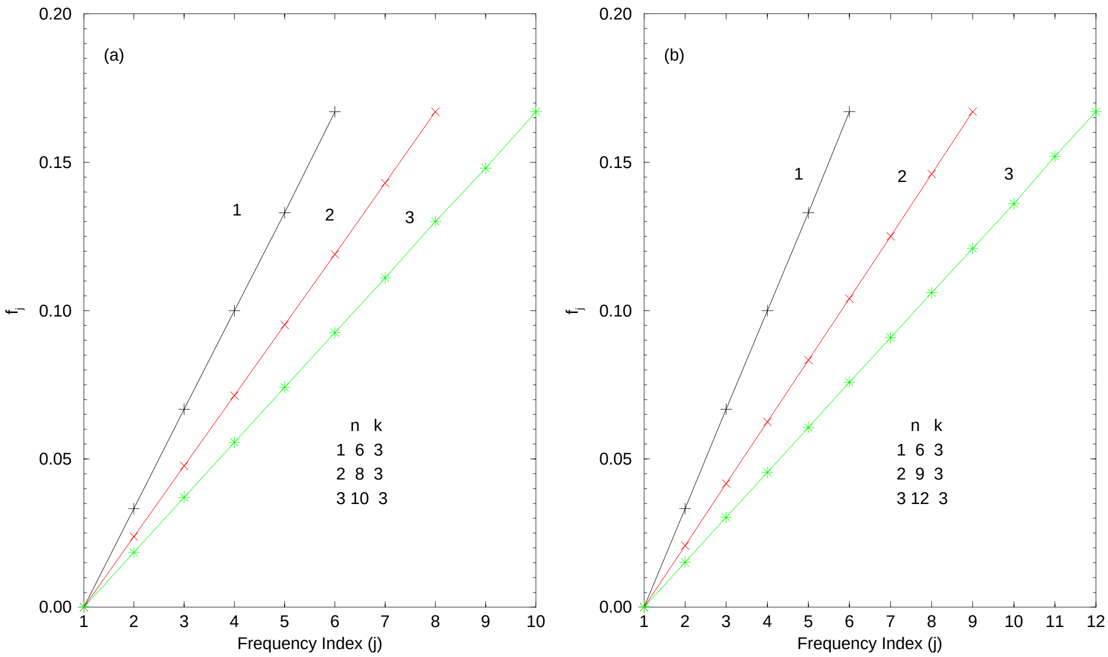

Figure 7: Coordinate frequencies corresponding to Fig. 6: (a) ARM; (b) EEG.

4 Perturbation Analysis Technique

In this section we present the methods developed to analyse perturbations and nonstationarity. We begin by determining approximate models to represent the ARM and EEG datasets. Then, we describe the technique used to analyse perturbations, and finally, we consider stationarity.

4.1 Approximate Models for the ARM and EEG Datasets

We repeat the qualitative description given in the Introduction for identifying significant transients in time series. We suppose the process being observed is represented by a trajectory in state space. Further, we suppose most of the time that trajectory is confined to a limited region in state space, which is referred to as an attractor. We suppose that occasionally the trajectory leaves that attractor and moves about in an extended region of state space before falling back to the attractor. Such a trajectory segment is a “perturbation ” and is associated with an unusal segment of the time series.

In this section we want to determine models for the ARM and EEG datasets that serve to define the so-called attractors associated with the two background processes. The method for determining these models is like that for determining a model for a nonlinear dynamical process as described in Section 3.2. However, the method described here is different because we only need a model sufficiently accurate to distinguish perturbations from the background process.

In Section 2 we presented results of preliminary analyses of training datasets assumed to be representative of the background processes corresponding to the ARM and EEG time series. Fig. 6 displays the eigenvalues Eq. (18) that result using an array of values for the modeling parameters (n, k) for each of the training datasets, and Fig. 7 displays the coordinate frequencies Eq. (22) corresponding to those modeling parameter values. The training datasets used here are the conditioned datasets corresponding to the preliminary analyis results presented in Figs. 3,4,5.

Referring to Fig. 6a, which corresponds to the ARM case, note that the first three eigenvalues are relatively stable over the selection of parameter values (n, k). From Fig. 7a, we observe that the first three coordinate frequencies associated with those parameter values fall in the low frequency band highlighted in the power spectrum Fig. 4b. As an aside, note that as f_w, which is defined by Eq. (17), decreases, the eigenvalues of the higher order coordinates, say $j = 4$ for example, increase. This is consistent with the decrease in the coordinate frequency $f_j(n, k)$, $j = 4$, observed in Fig. 7b.

The discussion just given relative to the ARM dataset also holds for the EEG case, although not quite as strongly. Referring to Fig. 6b and Fig. 7b, the first three transformed coordinates, θ_j, $j = 1, 2, 3$, Eq. (19), appear to capture the dominant structure in the EEG dataset, particularly for the latter two selected values for the parameter pair (n, k).

We want to summarize our results to this point. For each case, we have a "training" set of data, $\{y_i\}_{i=1}^N$, representative of the background process, and we have gone through a partial procedure of constructing a representation $(\tilde{A}_\theta, \tilde{f}_\theta, \tilde{p}_\theta)$ of the process corresponding to that training set. Recall that in constructing that representation we select values for the parameter pair (n, k), then define the trajectory

$$\{\mathbf{y}_i\}_{i=1}^N \subset E_y^n,$$

where

$$\mathbf{y}_i = (y_i, y_{i+k}, y_{i+2k}, \ldots, y_{i+(n-1)k})^T,$$

and then carry out the transformation Eq. (20)

$$\Psi : E_y^n \to E_\theta^n,$$

thus producing the trajectory

$$\{\Psi(\mathbf{y}_i)\}_{i=1}^N = \{\theta_i\}_{i=1}^N \subset E_\theta^n.$$

The transformation Ψ is equivalent to the regression model

$$\mathbf{y}_i = \sum_{\alpha=1}^{n} \theta_{i\alpha} \psi_\alpha \, .$$

For appropriately chosen values of the parameter triplet (n, k, m), that regression model can be replaced by the approximate model

$$\mathbf{y}_i = \sum_{\alpha=1}^{m} \theta_{i\alpha} \psi_\alpha + \epsilon_i, \tag{25}$$

where $m < n$ is the range of significant eigenvalues and ϵ_i is a random noise term used to model the insignificant eigenvalue components. For the two cases examined here, we find the background process can be approximately modeled as follows.

$$\begin{array}{rcl} ARM: \ (n, k, m) & = & (8, 3, 3) \\ EEG: \ (n, k, m) & = & (9, 3, 3) \end{array} \tag{26}$$

4.2 Identifying Perturbations

We are interested in developing techniques to analyse large datasets. One consequence of the "large" condition is that the dataset can not be analysed at one time as a whole. The general strategy is to partition the large dataset into pieces that are manageable and to separate the unusual from the usual over one piece at a time. In this section we describe how a piece is analysed.

The problem at hand is to identify perturbations relative to some background process. To accomplish that end, we propose to do what is equivalent to using a weak regression model of the process. Referring to Eq. (25), we propose to use the regression model

$$\mathbf{y}_i = \sum_{\alpha=1}^{m} \theta_{i\alpha}\psi_\alpha \tag{27}$$

with the parameter values (n, k, m) displayed in Eq. (26) for the ARM and EEG datasets.

The next task is to formalize the meaning of "perturbation." First, we generalize the definition of the projection P^m Eq. (24) by defining the projection

$$P_\theta^{m_1,m_2} : E_\theta^n \to E_\theta^{m_2-m_1+1}$$

by

$$P_\theta^{m_1,m_2}(\theta) = (\theta_{m_1}, \ldots, \theta_{m_2})^T \in L\{\psi_j\}_{j=m_1}^{m_2}, \tag{28}$$

where $L\{\psi_j\}_{j=m_1}^{m_2}$ is the linear subspace of E_θ^n spanned by the eigenvectors $\{\psi_j\}_{j=m_1}^{m_2}$. Then, define

$$\tilde{B} = P_\theta^{m_1,m_2}(\{\theta_i\}_{i=1}^N). \tag{29}$$

Thus, $\tilde{B}$ is the projection into the $(m_2 - m_1 + 1)$-dimensional subspace $L\{\psi_j\}_{j=m_1}^{m_2}$ of the trajectory $\{\theta_i\}_{i=1}^N$ in E_θ^n constructed from the training set $\{y_i\}_{i=1}^N$. We assume that the background process projects to a relatively small, dense region $\tilde{B}$ for a small value of $m_2 - m_1 + 1$.

The assumptions described above concerning the background process versus perturbations implies that the region $\tilde{B}$ is a concentrated region in state space associated with usual data segments and that unusual segments of the time series will produce trajectory segments in $L\{\psi_j\}_{j=m_1}^{m_2}$ that move outside $\tilde{B}$. Let T_s be a characteristic time scale associated with the background process. We define a background event, or usual segment of the observed time series by conditions on the reconstructed trajectory segment in E_θ^n. Define

$$\Gamma_j = \{\theta_i\}_{i=i_j}^{i_j+l_j-1}, \tag{30}$$

so that l_j is the length in time steps of Γ_j. If

$$\begin{array}{lcl} P_\theta^{m_1,m_2}(\theta_{i_j-1}) & \notin & \tilde{B}, \\ P_\theta^{m_1,m_2}(\Gamma_j) & \subset & \tilde{B}, \\ P_\theta^{m_1,m_2}(\theta_{i_j+l_j}) & \notin & \tilde{B}, \\ l_j \times t_s & \geq & T_s, \end{array} \tag{31}$$

where t_s Eq. (16) is the sample time for the time series, then the trajectory segment Γ_j is a usual event and the time series segment corresponding to Γ_j is a usual segment. In turn, we define the trajectory segment

$$\Delta_j = \{\theta_i\}_{i=i_j+l_j}^{i_{j+1}-1} \tag{32}$$

separating Γ_j and Γ_{j+1} to be the j-th perturbed segment. The length in time steps of Δ_j is

$$p_j = (i_{j+1} - 1) - (i_j + l_j) + 1 = i_{j+1} - (i_j + l_j).$$

Note that this definition allows a perturbed trajectory segment to pass through the region $\tilde{B}$ so long as the time it takes is less than the time scale T_s. The time series segment corresponding to Δ_j is, by definition, a perturbation, or unusual segment. Define

$$l(i) = \left\{ \begin{array}{lllll} 1 & , & \theta_i \in \Gamma_j & , & \text{some j}\,, \\ 0 & , & \theta_i \notin \Gamma_j & , & \text{any j}\,, \end{array} \right. \tag{33}$$

and define the function that accounts for the cumulative observation of usual segments

$$L(i) = \sum_{k=1}^{i} l(k)\,. \tag{34}$$

Referring to Eq. (24), we determined that the ARM and EEG background processes can be weakly approximated using the regression model Eq. (27) with the parameter values Eq. (26). For both cases, $m = 3$. If there is a long term trend in the background process, we expect that to be reflected primarily in the first coefficient, θ_1, of the regression model. Consequently, we use the coefficients (θ_2, θ_3), which is equivalent to setting

$$(m_1, m_2) = (2, 3) \tag{35}$$

in the projection $P_\theta^{m_1,m_2}$ Eq. (28). Thus, $\tilde{B}$ is the projection into the two dimensional subspace $L\{\psi_j\}_{j=2}^3$ of the trajectory $\{\theta_i\}_{i=1}^N$ in E_θ^n constructed from the training set $\{y_i\}_{i=1}^N$. Finally, we find that appropriate time scales for defining usual events are

$$\begin{array}{rcl} \text{ARM}: \ T_s & = & 180 \times t_s = 1hr\,, \\ \text{EEG}: \ T_s & = & 100 \times t_s \approx .2s\,. \end{array} \tag{36}$$

To illustrate the perturbation detection technique, we have selected example segments from the ARM and EEG datasets, each example segment including the training segment that has been used to this point. The length in number of observations and time of those example segments is

$$\begin{array}{lll} ARM: & \delta i = 50,000 & \delta t \approx 11.6\, days\,, \\ EEG: & \delta i = 46,080 & \delta t = 90.0\, s\,. \end{array} \tag{37}$$

The training set for the ARM example dataset is the initial set of $\delta i = 10,000$ observations, which is approximately 2 days. For the EEG case, the training set covers approximately a $20\,s$ period over the final phase of the example dataset.

Fig. 8 displayes the result of the projection P^{m_1,m_2}, where $(m_1, m_2) = (2, 3)$, for the ARM example dataset, and Fig. 9 displays the corresponding results for the EEG example segment. We note two points about those figures. Ordinarily, in making state-space plots of trajectories we connect successive points, usually by straight line segments, so that the trajectories can be observed. We have not done that here, so those figures simply display the locus of projected points from the respective reconstructed state-space trajectories. The second point is that those figures include only every 10-th point from the projection, which is done so that the plots are not too cluttered. Figures like Fig. 8 and Fig. 9 are sometimes referred to as scatter plots.

Consider the ARM case. Fig. 8a displays the scatter plot for only the training dataset, and Figs. 8b,c display the scatter plot for the example segment. Note that the axes in Fig. 8b are the same as those in Fig. 8a. Fig. 8c is a telescopic view of example segment scatterplot that includes all the points in the locus. Note the difference in scales used for the axes in Fig. 8c compared to those in Fig. 8a,b. The high density region observed in Figs. 8a,b is little more than point size in Fig. 8c.

We consider further the ARM case. In the next section we describe how the region $\tilde{B}$ used above to define usual trajectory segments is quantitatively estimated. For now, we state that the analysis produces the result

$$ARM: \; \tilde{B} = \left\{ \begin{array}{c} -.12E-01 \leq \theta_2 \leq +.12E-01 \\ -.13E-01 \leq \theta_3 \leq +.25E-02 \end{array} \right\} . \tag{38}$$

Note that this prescription for $\tilde{B}$ roughly approximates the rectangular region used for the scatter plots Figs. 8a,b, although it appears to be quite conservative relative to the observed high density areas in those figures. Referring to the difference in axes scales for Fig. 8c compared to Figs. 8a,b, perturbations in the ARM case include length scales that range over several orders of magnitude. Fig. 10a displays the cumulative time of usual trajectory segments for the example segment Eq. (34). Referring to the definition of usual segments, the slope of the curve in Fig. 10a is $slope = 1$ during periods of minimum duration T_s that the projected trajectory is inside $\tilde{B}$, while the slope is $slope = 0$ during periods when the projected trajectory moves about outside $\tilde{B}$. The total number of usual segments is nineteen, which means the total number of perturbations is approximately nineteen, also. Fig. 10a displays three clearly defined plateaus, the second one occurring at approximately 4.0 days actually being made up of two pieces, but several other relatively short segments of zero slope can also be observed. The total fraction of usual events, according to this analysis, is $F \approx 66\,\%$.

Consider now the EEG dataset results. Note in Fig. 9 that the same rectangular region is used for both frames. The EEG example segment is significantly different than the ARM dataset in that perturbations of greatly different length scales than the background process, as observed in the (θ_2, θ_3) linear subspace, do not occur. Our analysis yields the prescription

$$EEG: \; \tilde{B} = \left\{ \begin{array}{c} -.12E+01 \leq \theta_2 \leq +.12E+01 \\ -.57E+00 \leq \theta_3 \leq +.57E+00 \end{array} \right\} . \tag{39}$$

In the case of the EEG dataset, the perceived high density region from Fig. 9 corresponds closely to the analysed region $\tilde{B}$. Fig. 10b displays the cumulative time of usual segments Eq. (34). We note from the analysis that the total number of usual segments is ninety-four. From Fig. 10b, we estimate that the total fraction of usual segments for the EEG example

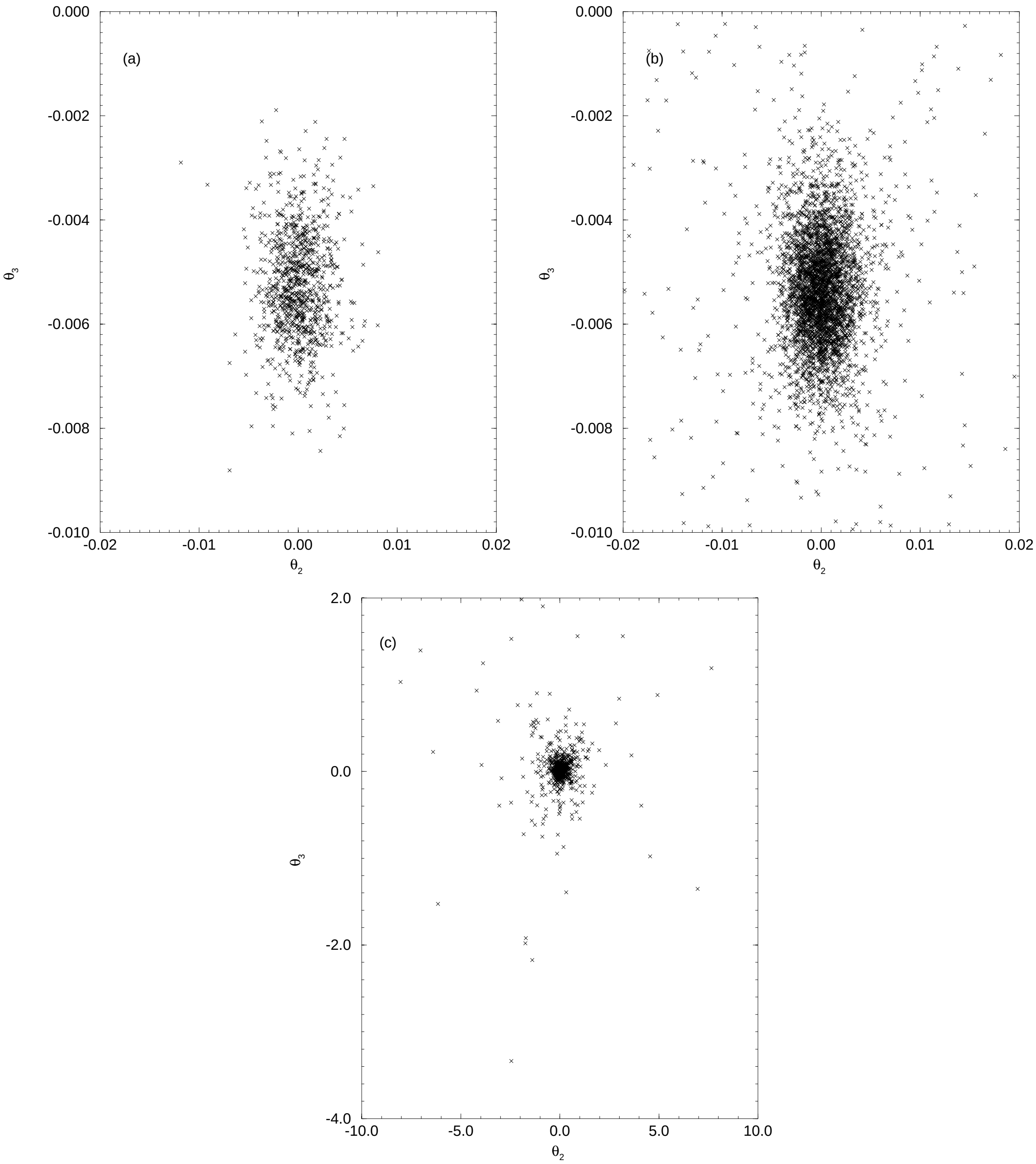

Figure 8: Trajectory projection into (θ_2, θ_3) linear subspace for ARM data: (a) training dataset; (b) 11.5 day datasegment; (c) telescopic view of frame (b).

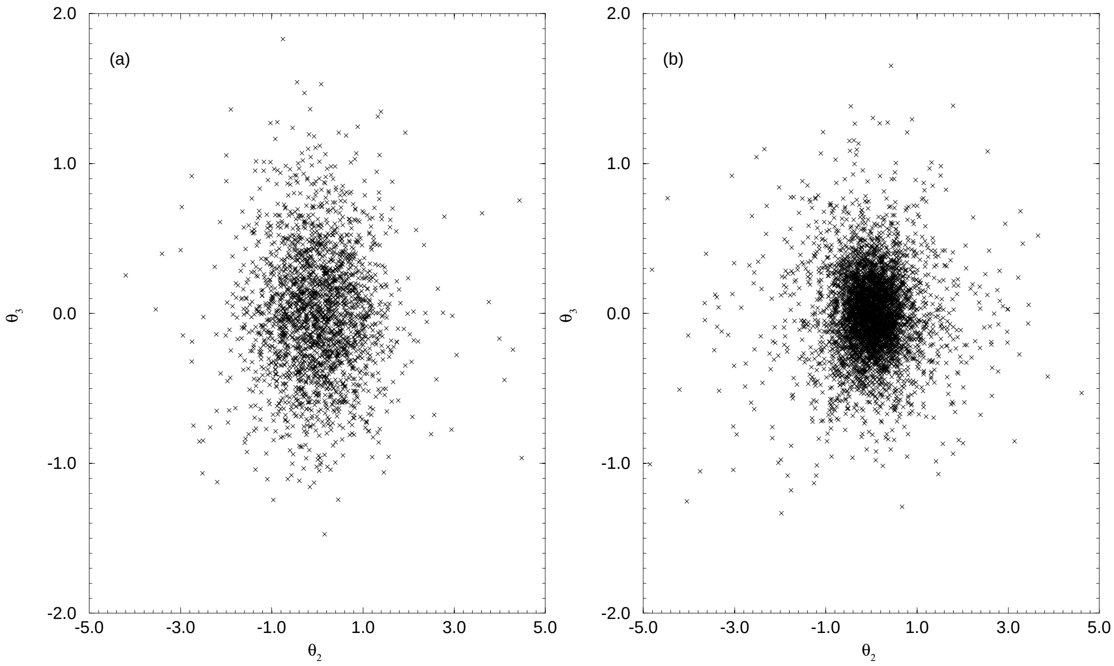

Figure 9: Trajectory projection into (θ_2, θ_3) linear subspace for EEG data: (a) training dataset; (b) 90s datasegment.

dataset is $F \approx 66\,\%$, but that the majority of perturbations occur in the first $40s$ segment. In fact, the training dataset covers approximately the last $20s$ piece of the example segment.

To complete this section, we present a rudimantary form of cluster analysis for the two example analyses. Figs. 11,12, display the ARM and EEG example segments, where perturbations are colored coded according to duration. For both cases there are six bins, or time intervals, and a perturbation is colored according to which bin its duration belongs. The background, which determines usual segments, is colored *black*. Table 1 describes the color code used for those figures. We note that for the ARM case, the absolute maximum value of amplitude for perturbations increases with duration, so that for the scales used in Fig. 11 short duration perturbations may not appear to have structure. However, in a magnified view, they are clearly distinguishable from the background.

		ARM	EEG
0	black		
1	red	$0.0\,hr \leq p_j \leq 4.0\,hr$	$0.0\,s \leq p_j \leq 0.1\,s$
2	gold	$4.0\,hr \leq p_j \leq 8.0\,hr$	$0.1\,s \leq p_j \leq 0.2\,s$
3	yellow	$8.0\,hr \leq p_j \leq 12.0\,hr$	$0.2\,s \leq p_j \leq 0.3\,s$
4	green	$12.0\,hr \leq p_j \leq 16.0\,hr$	$0.3\,s \leq p_j \leq 0.4\,s$
5	purple	$16.0\,hr \leq p_j \leq 20.0\,hr$	$0.4\,s \leq p_j \leq 0.5\,s$
6	sea green	$20.0\,hr \leq p_j$	$0.5\,s \leq p_j$

Table 1: Color codes for unusual segments in ARM and EEG example datasets displayed in Figs. 11,12.

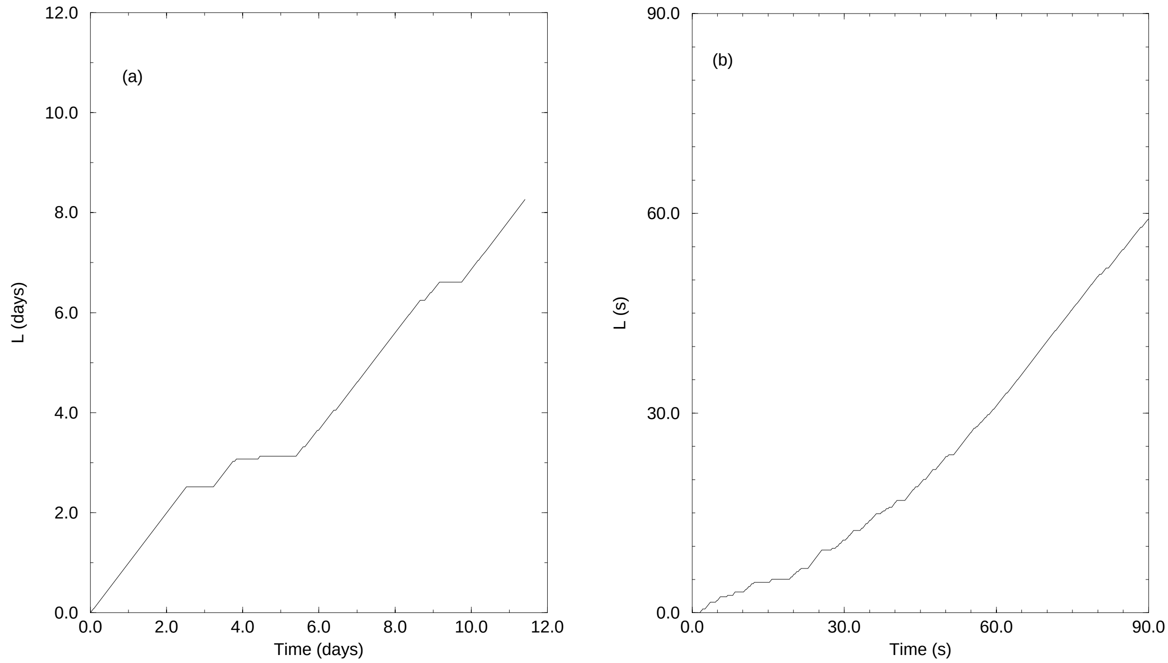

Figure 10: Cumulative time of usual segments for the example segments: (a) ARM example dataset; (b) EEG example dataset.

4.3 Analysing Nonstationarity

As described in the introductory remarks to the previous section, we are looking for techniques to analyse datasets that are too large to be computationally manageable as a whole at one time. The strategy is to partition the dataset into manageable pieces and analyse one piece of the partition at a time. In the previous section we looked at a method for analysing one piece. In this section we complete the description of the technique discussed in this article for analysing large datasets by describing how the analysis moves from one piece to the next.

At issue is the prescription of the region $\tilde{B}$ Eq. (29) used to define usual segments Eqs. (30),(31). In the previous section we tacitly assumed the procedure for identifying $\tilde{B}$ depends on having a training set of data. Because of the large dataset condition, which also implies there are many pieces in the partition of that dataset, it is practical to assume the analysis of the large dataset can be preceded by a preliminary analysis to initialize it. The initialization procedure may reasonably include identifying a training set to model the background process. However, in moving from one piece of the partition to the next, it is not practical for the analyst to interrupt the analysis in order to identify a new training set.

We are going to describe a method for objectively specifying $\tilde{B}$ Eq. (29) for an arbitrarily fixed member of the large dataset partition. If $\tilde{B}$ is a constant function of the partition index, then we say the process is "stationary;" otherwise, the process is "nonstationary."

We are gonig to assume the background process does not experience a fundamental, qualitative change during the time of observation. Thus, we are assuming the initial reconstruction analysis for the background process that resulted in values for the reconstruction parameters (n, k, m) given in Eq. (26) remains valid and that the subsequent analysis which

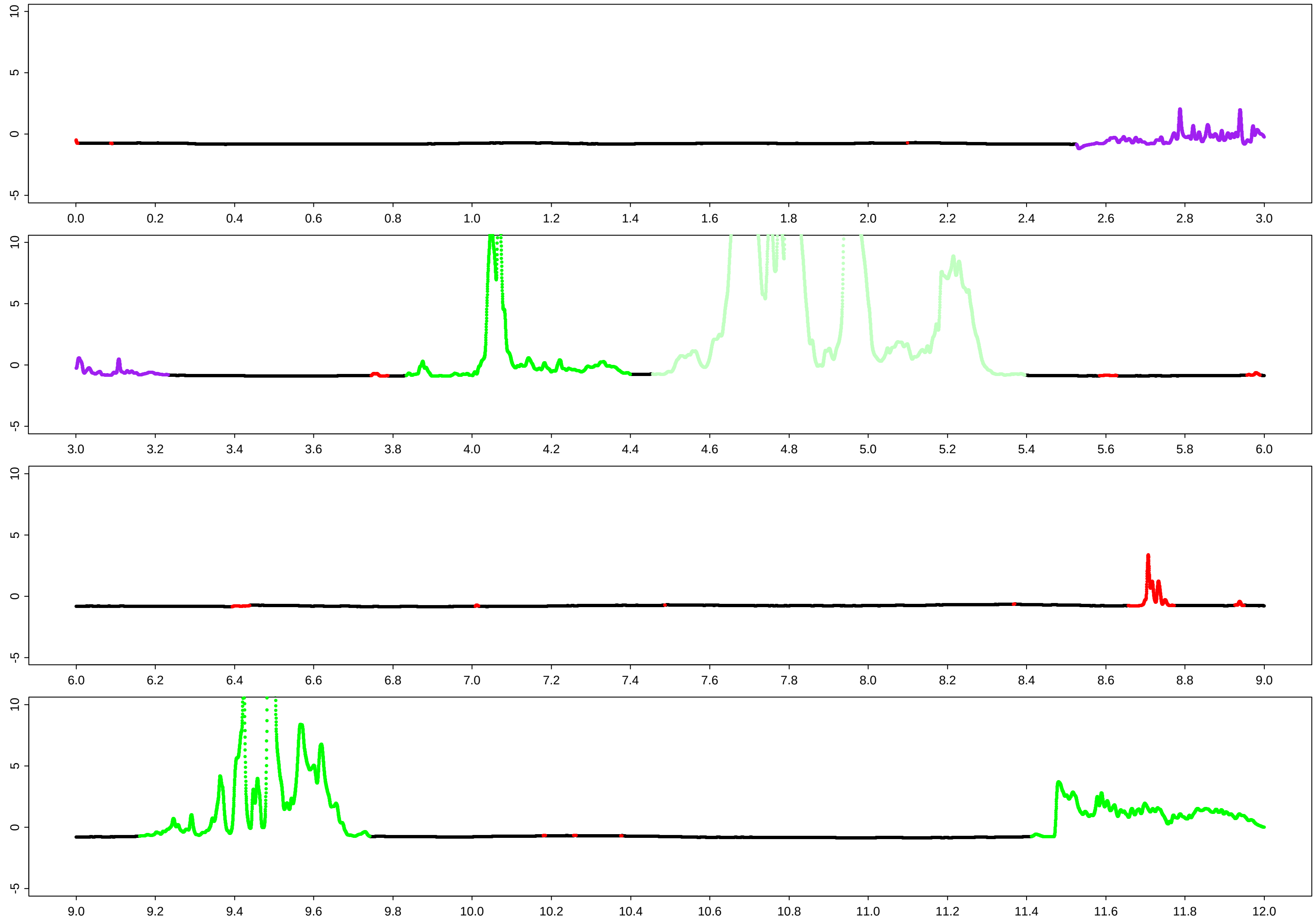

Figure 11: ARM example data series with perturbations color coded by duration according to the description given in Table 1: ordinate = data(y); abscissa = time(days).

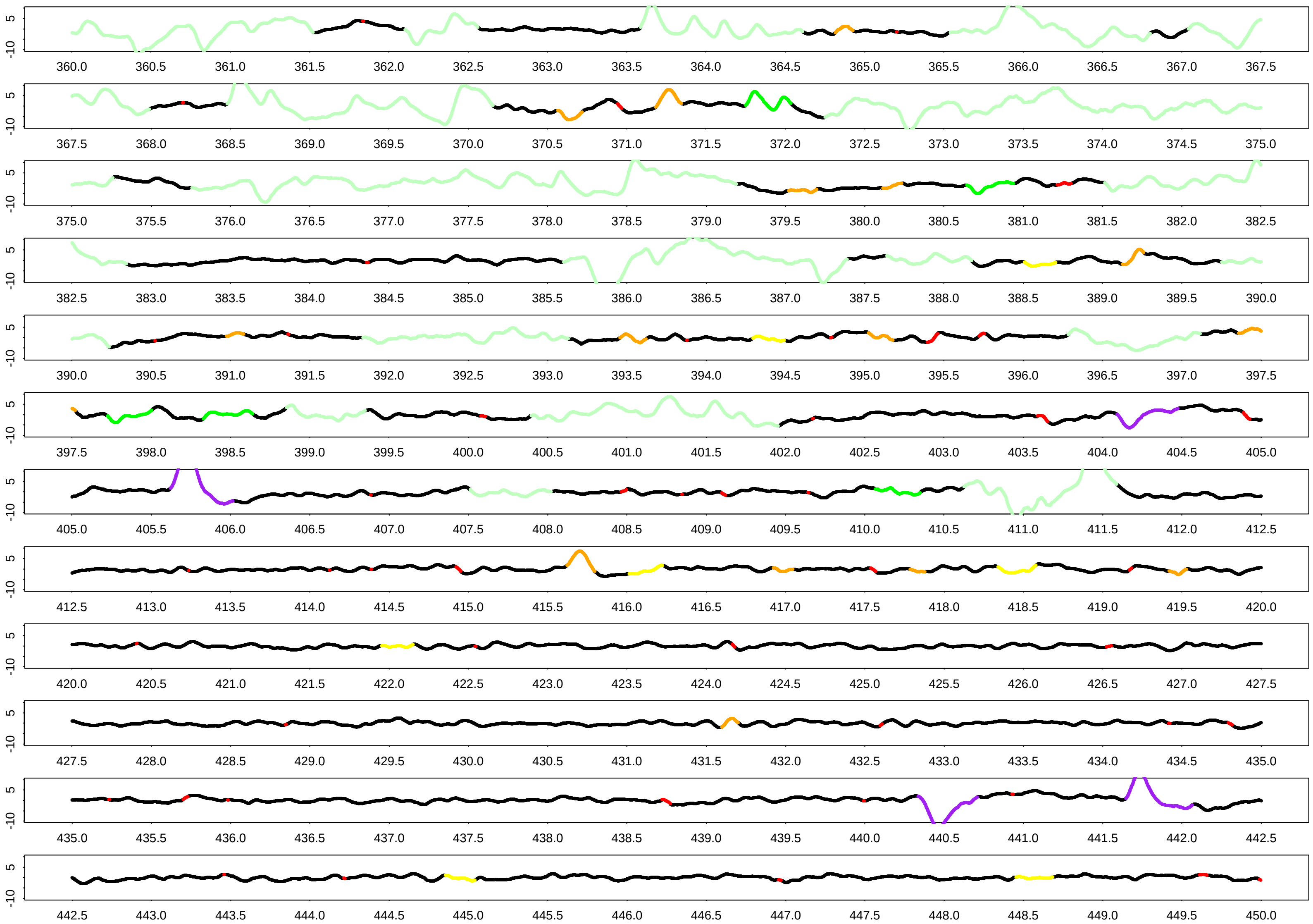

Figure 12: EEG example data series with perturbations color coded by duration according to the description given in Table 1: ordinate = data(y); abscissa = time(seconds).

lead to values for the projection parameters (m_1, m_2) given in Eq. (35) used to determine $\tilde{B}$ also remains valid.

The method we use for determining $\tilde{B}$ depends on estimating distributions for the relavent θ vector components. We want to use the first and second moments of the estimated distributions to define $\tilde{B}$. The difficulty stems from the fact that the tails of the global distributions of the relavent θ components can produce values for those moments that are poor estimates of the high density region we want to associate with $\tilde{B}$. A two step procedure is introduced, first, to locate the high density region and, second, to measure its extent.

We are going to describe the method for estimating $\tilde{B}$ using the ARM example dataset introduced in the preceding section. Consider Fig. 8. Recall that Figs. 8a,b display the high density region in the (θ_2, θ_3) linear subspace for the example ARM dataset. For both θ_2 and θ_3, the interval $[-.1, +.1]$ is larger by an order of magnitude than the interval used for either variable in those figures. Fig. 13a displays densities for θ_2, θ_3 restricted to the interval $[-.1, +.1]$, and Fig. 13b shows the cumulative distribution functions for those densities. Fig. 13c displays approximations of the second derivative of those cumulative distribution functions. Letting F represent either cumulative distribution function, a limited region $[\theta_{min}, \theta_{max}]$ for θ is determined by the conditions

$$\begin{aligned} \frac{d^2F}{d\theta^2}(\theta_{min}) &= +\alpha \,, \\ \frac{d^2F}{d\theta^2}(\theta_{max}) &= -\alpha \,, \end{aligned} \tag{40}$$

where θ_{min} and θ_{max} are the minimum and maximum values of θ, respectively, satisfying these equations. Next, using the resulting intervals for the two variables, we construct the density functions limited to those intervals. Setting

$$\alpha = .05 \tag{41}$$

in Eq. (40) leads to the densities displayed in Fig. 13d.

Given the real random variable x with density $\rho(x)$, the mean and squared variance are defined, respectively, by

$$\mu_x = \int x\rho(x)dx \,, \tag{42}$$

$$\sigma_x^2 = \int (x-\mu_x)^2\rho(x)dx \,. \tag{43}$$

The means and variances corresponding to the densities displayed in Figs. 13a,d are summarized in Table 2. Referring to the dense region in Fig. 8, note that the means and variances estimated from the densities displayed in Fig. 13d serve to quantify that region very well.

As a result of numerical experimentation, we find that the rule

$$\theta = \mu_\theta \pm 2 \times \sigma_\theta \tag{44}$$

for defining the extent of $\tilde{B}$ in each coordinate direction works well. The prescription of $\tilde{B}$ for the ARM example dataset given by Eq. (38) follows from the values for mean and variance given in Table 2 using this rule.

Exactly the same method used to analyse $\tilde{B}$ for the ARM dataset is used for the EEG dataset. The initial density estimates for θ_2, θ_3 are made using the intervals $[-4.0, +4.0]$, $[-2.0, +2.0]$, respectively, which appear to be reasonable choices according to Fig. 9. The same value for α Eq. (41) used in Eq. (40) for the ARM dataset is used for the EEG dataset. The resulting densities are displayed in Fig. 14 and the resulting means and variances are given in Table 3. As for the ARM case, the means and variances in Table 3 describe the high density region for the EEG example dataset, which is illustrated in Fig. 9, very well. The prescription for $\tilde{B}$ given by Eq. (39) corresponds to the formula Eq. (44) used with the means and variances in Table 3.

The estimate of $\tilde{B}$ for an arbitrary piece of the partition of the large data series is made using the methodology described above. This methodology requires knowledge of an interval on which to make an initial estimate of a density. For the first piece, we can do a preliminary analysis that provides intervals like those found above for the example ARM and EEG datasets, which led to the densities displayed in Figs. 13a,14a. Beyond the first piece, we use the results from the preceding piece to provide those intervals for making initial estimates of the densities. Using $[\theta_{min}, \theta_{max}]$ defined by Eq. (40), we define an initial interval for each θ by magnifying that interval by a factor about its center. That is, setting

$$\begin{aligned} \theta_c &= .5 \times (\theta_{min} + \theta_{max}) \quad , \\ \theta_d &= (\theta_{max} - \theta_c) \quad , \end{aligned}$$

we use

$$\theta = \theta_c \pm \beta \times \theta_d \tag{45}$$

to define the initial interval for estimating the density of θ. For the results displayed below, we use the magnification factor value $\beta = 4.0$.

Fig. 15 displays the cumulative time of usual segments for excerpts from the ARM and EEG records, each sufficiently long to illustrate the large data series analysis technique. The initial 15 pieces of the partition of each record are used for this purpose, where each piece corresponds in length to the example datasets used earlier Eq. (37). For convenience, we repeat that prescription here.

$$\begin{aligned} ARM: &\quad \delta i = 50,000 \quad \delta t \approx 11.6\, days \\ EEG: &\quad \delta i = 46,080 \quad \delta t = 90.0\, s \end{aligned}$$

Each frame of Fig. 15 has two curves, one corresponding to making a new estimate of $\tilde{B}$ for each piece, referred to in the figure as "dynamic attractor", and the other corresponding to using the estimate of $\tilde{B}$ made on the initial piece for all pieces, which is labeled "fixed attrator." For both cases, the dynamic specification of $\tilde{B}$ leads to a cumulative function that is approximately linear. The difference in results between the dynamic and fixed specifications of $\tilde{B}$ is evident in both cases, but is particularly dramatic for the EEG case. We note, however, that for a case of sufficiently long duration, which is not given here, the ARM results display the same character with respect to fixed versus dynamic specification of $\tilde{B}$ as the EEG results shown in Fig. 15b.

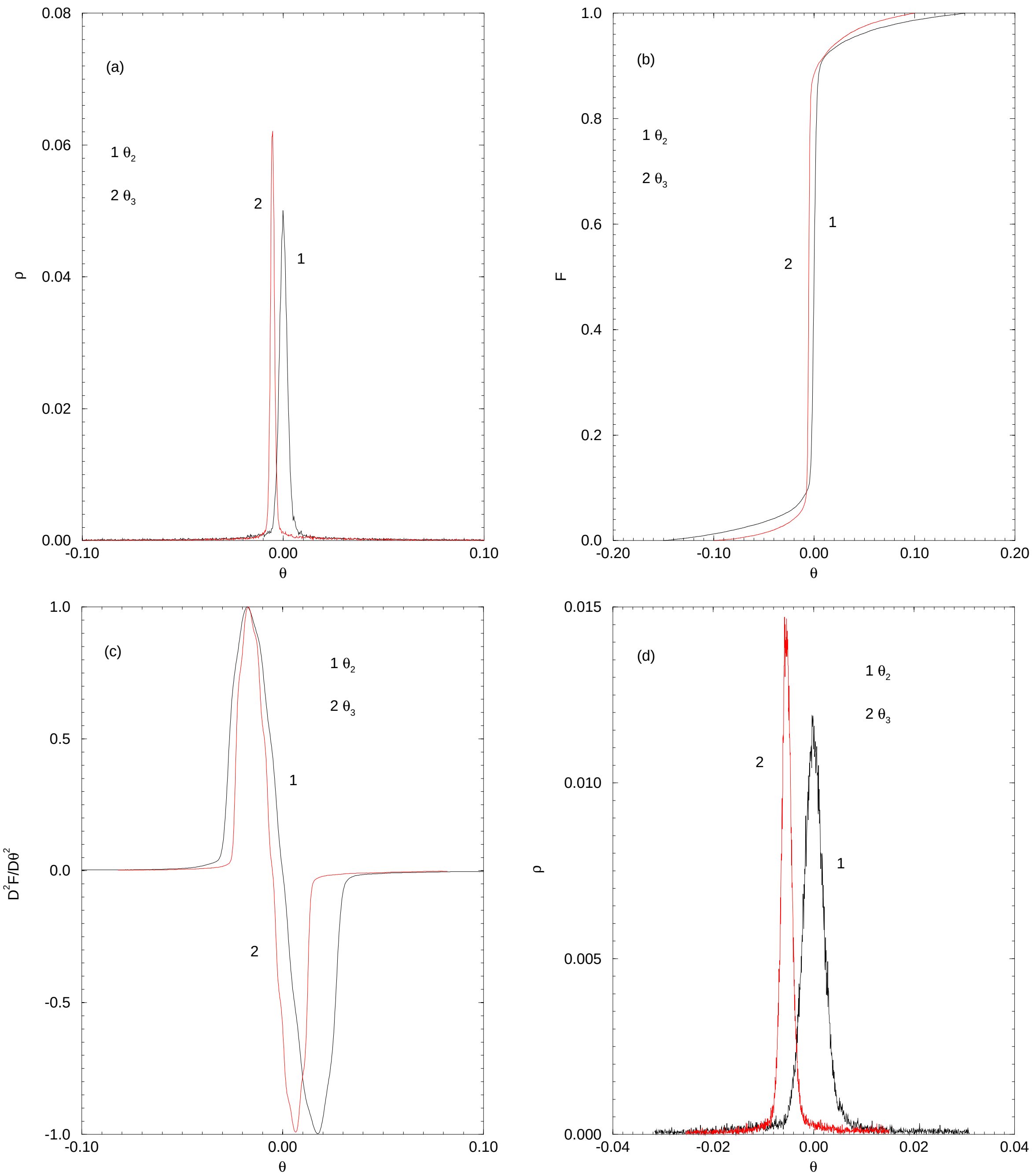

Figure 13: Density analysis of θ_2, θ_3 for ARM example segment: (a) initial density estimate; (b) cumulative frequency function of initial density; (c) second derivative of cumulative frequency function of initial density; (d) resulting density estimate.

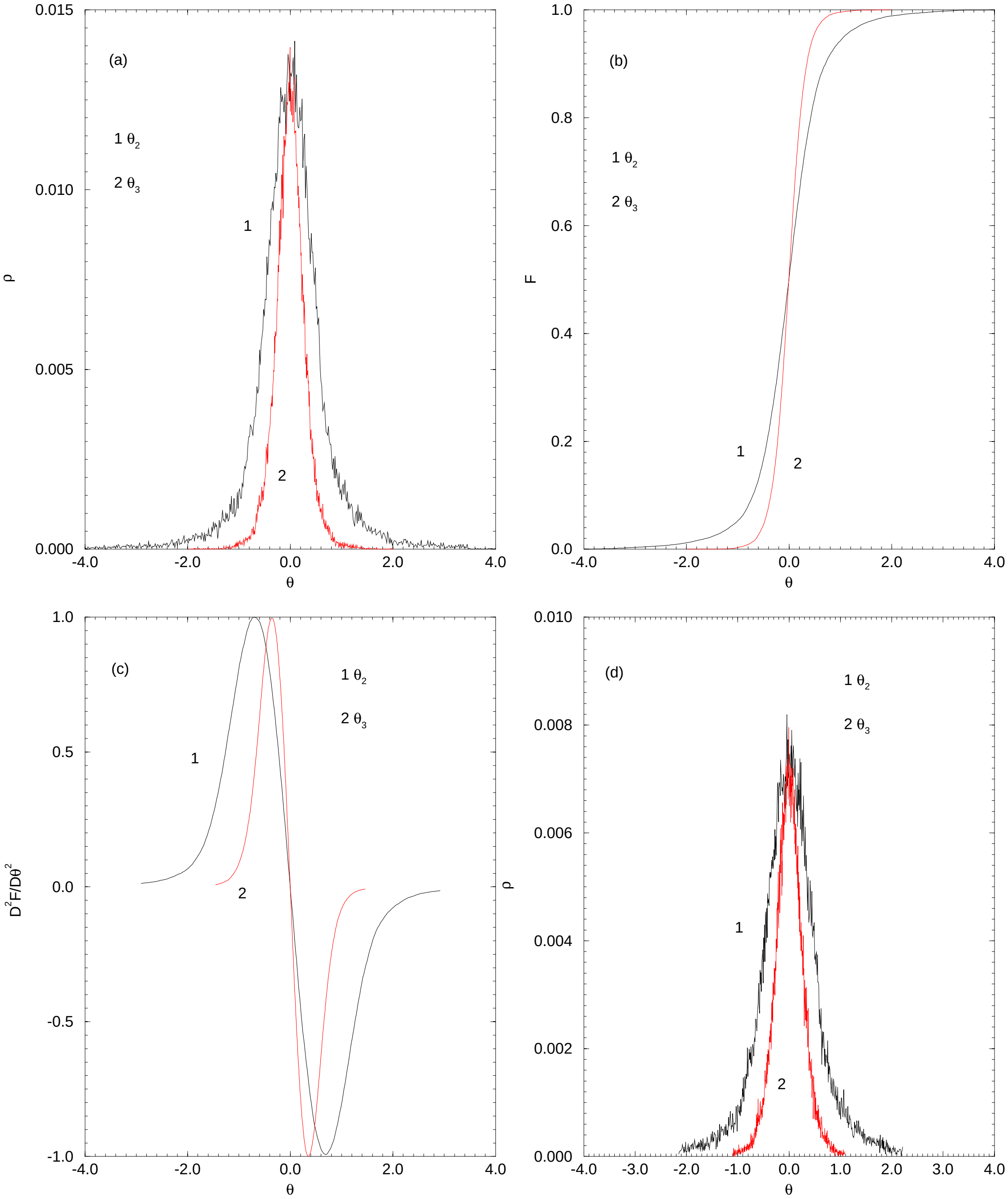

Figure 14: Density analysis of θ_2, θ_3 for EEG example segment: (a) initial density estimate; (b) cumulative frequency function of initial density; (c) second derivative of cumulative frequency function of initial density; (d) resulting density estimate.

Fig. 13a			Fig. 13d	
	θ_2	θ_3	θ_2	θ_3
μ	+.27E−03	−.32E−02	−.11E−04	−.51E−02
σ	+.27E−01	+.17E−01	+.59E−02	+.38E−02

Table 2: Mean and variance values corresponding to densities displayed in Figs. 13a,d for ARM example dataset.

Fig. 14a			Fig. 14d	
	θ_2	θ_3	θ_2	θ_3
μ	+.40E−02	−.17E−03	+.79E−02	−.19E−02
σ	+.70E+00	+.30E+00	+.59E+00	+.29E+00

Table 3: Mean and variance values corresponding to densities displayed in Figs. 14a,d for EEG example dataset.

5 Summary and Conclusions

This article outlines a technique for finding unusual segments in large data series. Further, the technique is demonstrated using two examples of such data sets. The essence of the technique is, first, to partition the large data series into a series of manageable pieces, and second, to construct a simple model that can distinguish between the usual and unusual, where "simple" is intended to imply "fast". That model is applied to one piece of the partition at a time.

The technique is based on the qualitative idea that most of the measured data corresponds to observing some background process, but that occasionally, due either to external forces or internal conditions, perturbations take place. We refer to data corresponding to the background process as "usual" and to that corresponding to perturbations as "unusual."

In Section 2 we discuss preliminary analyses used, first, to identify the information content of the data series in frequency space and, second, to condition the the data series by minimizing energy outside the meaningful band. This band determines the range of time scales we want to model.

In Section 3 we outline that part of the theory of nonlinear dynamical processes that provides the foundation for modeling nonlinear systems using time series measurements. Further, we outline the methodology used to construct models. That methodology depends on the information gained about the time series doing the analyses described in Section 2, especially the information about relavent time scales.

In Section 4 the technique for analysing large data series is described and illustrated using the two example time series. In particular, we describe how to construct a "simple" model using the more general ideas for modeling described in Section 3. Further, we define the meaning of "usual" and, in turn, we define "unusual" to be the complement of all that is usual.

We want to make some observations concerning the assumptions on which the technique is based.

1. It is assumed the large data set can be partitioned serially so that each member,

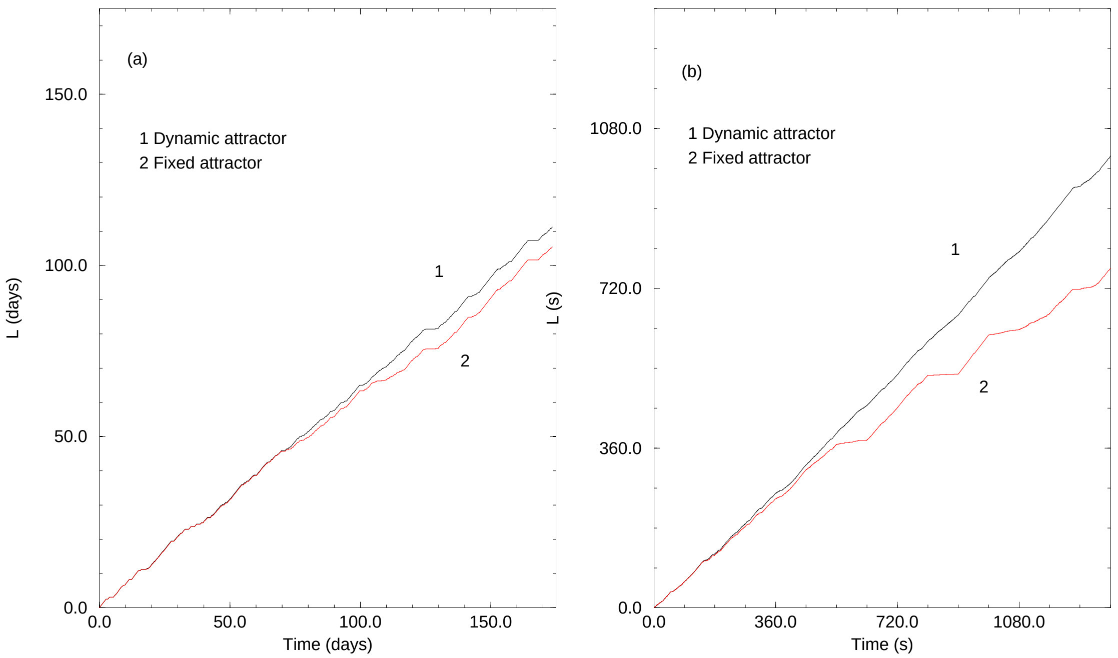

Figure 15: Cumulative time of usual segments: (a) ARM dataset; (b) EEG dataset.

or piece, of the partition is manageable and so that each of the longest time-scale unusual events of interest is expected to be contained in a member of the partition. By "manageable," we mean that data can be accomodated on a workstation generally available to researchers. As the large dataset may be something observed over a long period of time, some knowledge of the process, other than purely empirical, is needed to determine a time scale for the pieces of the partition. Futher, each piece is assumed to contain sufficient information about the background process to model it.

2. Referring to power spectra, we assume both the background process and the perturbations of interest correspond to the same frequency band. Part of the methodolgy includes isolating a frequency band of information from above by low-pass filtering. In effect, the length of a piece isolates the band from below. If the measurement instrumentation includes a high-pass filtering component, that value together with physical information about the process should be taken into consideration in choosing the length scale for a piece.

In the examples used to illustrate the technique we have not tried to use meaningful time scales for the pieces of the two partitions. Further, referring to what is described as a "characteristic time scale", T_s Eq (31), we did not discuss how to select a value for it, and the values we used for the two examples Eq. (36) are not the result of some analysis of the two background processes associated with the examples. However, the choice of a characteristic time scale can be expected, in general, to reflect the dominant frequency in the frequency band of information for the background process.

We did not discuss the results of the analyses of the ARM and EEG time series from the standpoint of scientific value. That is not the purpose of this article. However, on the basis of discussions with Dr. Jim Liljegren, Pacific Nothwest Laboratory, Richland, WA, an atmospheric instrumentation scientist who is knowledgeable about the ARM data used in this article, we can say tentatively that virtually all instrumentation malfunctions as well as observed atmospheric liquid water events, like cloud, rain, and fog, are identified by the technique. Futher, on the basis of discussions with Dr. Michael Eisenstadt, Neurologist, Knoxville Neurology Clinic, Knoxville, TN, we can say that virtually every instance of clinically defined artifact, like muscle tension, head movement, and eye movement, is identified as a perturbation.

References

[1] R. B. Blackman and J. W. Tukey. *The Measurement of Power Spectra.* Dover Publications, Inc., New York, NY, 1958.

[2] D. S. Broomhead and G. P. King. Extracting qualitative dynamics from experimental data. *Physica*, 20D:271, 1986.

[3] D. J. Downing, V. V. Fedorov, W. F. Lawkins, M. D. Morris, and G. Ostrouchov. Large datasets: Segmentation, feature extraction, and compression. Technical Report ORNL/TM-13114, Oak Ridge National Laboratory, Oak Ridge, TN 37831, 1995.

[4] D. J. Downing, W. F. Lawkins, M. D. Morris, and G. Ostrouchov. A method for detecting changes in long time series. Technical Report ORNL/TM-12879, Oak Ridge National Laboratory, Oak Ridge, TN 37831, 1995.

[5] D. O. E. Atmospheric radiation measurement program plan. Technical Report DOE/ER-0441, U. S. Department of Energy, Office of Health and Environmental Research, Atmospheric and Climate Research Division, National Technical Information Service, 5285 Port Royal Road, Springfield, Virginia 22161, 1990.

[6] A.M. Fraser and H.L. Swinney. Independent coordinates for strange attractors from mutual information. *Physical Review A*, 33:1134–1140, 1986.

[7] M. Fraser. Reconstructing attractors from scalar time series: A comparison of singular systems and redundancy criteria. *Physica D*, 34:391–404, 1989.

[8] L. M. Hively, N. E. Clapp, C. S. Daw, W. F. Lawkins, and M. L. Eisenstadt. Nonlinear analysis of eeg for epileptic seizures. Technical Report ORNL/TM-12961, Oak Ridge National Laboratory, Oak Ridge, TN 37831, 1995.

[9] W. F. Lawkins, C. S. Daw, D. J. Downing, and N. E. Clapp. The role of low-pass filtering in the process of attractor reconstruction from experimental chaotic time series. *Physical Review E*, 47:2520–2535, 1993.

[10] Lawrence R. Rabiner and Bernard Gold. *Theory and Application of Digital Signal Processing.* Prentice-Hall, Inc, Englewood Cliffs, New Jersey, 1975.

[11] R. Mañé. On the dimension of the compact invariant sets of certain nonlinear maps. In D. A. Rand and L.S. Young, editors, *Proceedings of the Warwick Symposium*, pages 230–242, New York, 1981. Springer.

[12] C.E. Shannon and W. Weaver. *The Mathematical Theory of Communication.* The University of Illinois Press, Urbana,Illinois, 1949.

[13] F. Takens. Detecting strange attractors in turbulence. In D. A. Rand and L.S. Young, editors, *Proceedings of the Warwick Symposium*, pages 366–381, New York, 1981. Springer.