Time series analysis, or the quest for quantitative measures of time dependent behavior

Alfonso M. Albano[1]*, Peter D. Brodfuehrer[1], Christopher J. Cellucci[2], Xenia T. Tigno[3], and Paul E. Rapp[2]
[1]Bryn Mawr College, Bryn Mawr, Pennsylvania, USA
[2]Uniformed Services University, Bethesda, Maryland, USA
[3]University of South Florida, Tampa, Florida, USA

We review some of the tools used to analyze time series—sequences of measurements of one or more variables pertaining to the same system. The tools include time-resolved statistics and spectra; autocorrelations, mutual information and transfer entropy; complexity, predictability and hypothesis testing. We illustrate the use of these tools by analyzing extracellular recordings from the leech, vasomotion data from monkeys, and human electroencephalograms.

I. Introduction

At one time or another, every experimental or observational scientist engages in some time series analysis. That is, given a series of measurements of a variable, one calculates quantities that may be used to characterize the data or preferably, to characterize the system that produced the data. Given two or more such measured variables, one seeks functional relationships or measures of correlation or of causal relations.

In recent decades, with the development of more precise measuring instruments, and with the increased popularity of automated, high-speed data acquisition, it has become possible to amass huge amounts of data that require analysis and interpretation. In the biological and medical sciences, unlike in some of the physical sciences, these data streams are often not predicted, and not related, by rigorous, mathematically based theory. It is the data themselves that must tell the story, and it is from analysis of the data rather than from theoretical predictions that one tries to find relationships among measured variables. This is a major task of time series analysis.

Until quite recently, time series analysis consisted almost exclusively of calculations and model-building that focused attention on linear properties of the measured variable. Often, it employed techniques that have little or no sensitivity to the time sequence of the data. One calculated the variable's statistics and measures of correlation and spectra. One built statistical or mathematical models which, when they worked, made possible the prediction of future values.

Statistics, though, are insensitive to time sequence and so are incapable of providing information about the system's time evolution—its *dynamics*, and linear measures tell only part of the story. While spectra are sensitive to time sequence, the information they provide applies to the entire epoch spanned by the data and hence cannot shed any light on changes in the system's dynamics during that epoch. Indeed, statistical as well as spectral calculations presume that no significant changes occur in the system's properties during the time spanned by the data—that is, that the data are *stationary*.

In recent decades, considerable attention has been paid to the estimation of measures that describe dynamical and nonlinear characteristics of time series. Some of these methods are extensions of familiar ones while some others have resulted from the recent resurgence of activity in nonlinear dynamics. In this review, we describe a subset of these measures, and illustrate their uses by applying them to the analysis of some biological data. The list of tools we discuss is by no means exhaustive. It only includes some of the tools that we have found to be useful and appropriate in attempting to understand our data.

In Sec. II we look at time-resolved statistics and illustrate their use in predicting behavioral responses of the medicinal leech. We find that the leech's response may depend on the activity of its nervous system prior to stimulation.

In Sec. III we use the Fourier and Gabor Transforms—the latter is essentially time-resolved Fourier Transforms—to show differences among healthy, prediabetic, and diabetic monkeys. Our results indicate that spectral measures may provide criteria for identifying subjects who are in imminent danger of developing diabetes.

In Sec. IV we discuss linear correlations and two Information Theoretic measures—mutual information and transfer entropy. Mutual information measures nonlinear correlations between data streams. Transfer entropy measures the amount of information transferred between them. We illustrate the use of these measures by studying patterns of correlation and of information transfer in 10-channel electroencephalographic (EEG) data recorded from a human subject.

Algorithmic complexity is a measure of the compressibility of symbol sequences. When applied to time series data, it is used to measure the degree of nonrandom structure in the time series. Nonlinear predictability measures how well future values of a variable can be predicted from its past. The latter provides some measure of the possibility that the data were generated by a deterministic dynamical system. We discuss these in Sec. V and again use EEG data to illustrate the kind of information provided by these quantities.

We conclude, in Sec. VI, with a discussion of hypothesis testing—the use of surrogates to test the validity of assumptions about the nature of the data. We discuss a few of the commonly used surrogates and illustrate their use by testing some hypotheses about the nature of

*To whom the correspondence should be addressed. E-mail:aalbano@brynmawr.edu
Received October 21, 2008; Accepted November 1, 2008.
Editor-in-charge: Amador C. Muriel

EEG signals and their correlations. Preliminary results suggest that an individual EEG data stream that is indistinguishable from linearly correlated noise may nevertheless have correlations and may participate in information transfer in a manner that cannot fully be accounted for by linear processes.

II. Statistics and Dynamics

II.1 Mean and variance

Given a series of values, $X=\{x_1, x_2, \ldots, x_N\}$ of a variable, x, the most common way of characterizing it is by using its mean or average value,

$$\langle x \rangle = \frac{1}{N}\sum_{i=1}^{N} x_i \qquad (1)$$

Here and in the following, we assume that N is large enough and that the series of values of x available to us is a sufficiently unbiased sampling of the population of x's so that the quantities we calculate from our sample are good enough approximations of the quantities that characterize the population as a whole.

The mean, though, is a pretty crude measure; its inadequacy illustrated by the story of the non-swimmer who drowned in a river with a mean depth of 20 cm. The mean does not convey any information about the range of values that the variable may have, and is greatly influenced by outliers—values that are either much greater or much less than the rest.

Some information concerning the spread of values about the mean is given by the variance,

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}\left(x_i - \langle x \rangle\right)^2 \qquad (2)$$

or by its square root, the standard deviation, σ.

II.2 Higher moments

More detailed statistical characterization of the distribution of the x's is given by averages of higher powers of x, or averages of higher powers of deviations of x from its mean, that is, the p^{th} order moments of x, defined by

$$\left\langle x^p \right\rangle = \frac{1}{N}\sum_{i=1}^{N}\left(x_i\right)^p \qquad (3)$$

or the p^{th} *order central moments of* x,

$$\left\langle \left(x - \langle x \rangle\right)^p \right\rangle = \frac{1}{N}\sum_{i=1}^{N}\left(x_i - \langle x \rangle\right)^p \qquad (4)$$

Note that the mean is the first moment of the distribution of the x's, and the variance is its second central moment. The third central moment, the skewness, measures the symmetry of the distribution about the mean, and the fourth central moment, the *kurtosis* measures the deviation of the distribution from a gaussian, or normal, distribution.

These statistical measures can only characterize the distribution of the x's, however. They are insensitive to the sequence in which different values of x appear in the series. In other words, these measures are insensitive to the dynamics of the system that produced the time series; they can only characterize the relative frequency with which the values occur. If the series of x's were arbitrarily shuffled, thereby destroying any information conveyed by their temporal order, the moments of the distribution would not change. The time sequence of dynamical signals carry important information about the systems that generate them. This level of statistical description is clearly inadequate for characterizing dynamical systems. Any characterization of dynamics must take time into account.

II.3 Time-resolved statistics

Some information on the manner in which the time series changes in time may be obtained by partitioning it into sub-epochs or windows and calculating the statistics of each window. The windows may or may not overlap. The behavior in time of the statistics characterizing this sequence of windows then provides some dynamical information. The resulting quantities are sometimes called time-resolved statistics. We illustrate their use with some data from the leech.

II.3a Leeches: to swim or not to swim

When an interneuron in one of the brains of the medicinal leech (*Hirudo medicinalis*) is electrically stimulated (yes, a leech has two brains, one at each end!), sometimes it shows signs of swimming, sometimes it shows signs of crawling, sometimes it does not show any overt response (Brodfuehrer et al. 2008 and references therein).

Figure 1 shows electrical signals propagating along the ventral nerve cord before, during, and after stimulation. The data were recorded extracellularly between the 9th and 10th ganglia of an isolated leech nervous system. Signals were digitized at 4.0 kHz. The topmost trace is

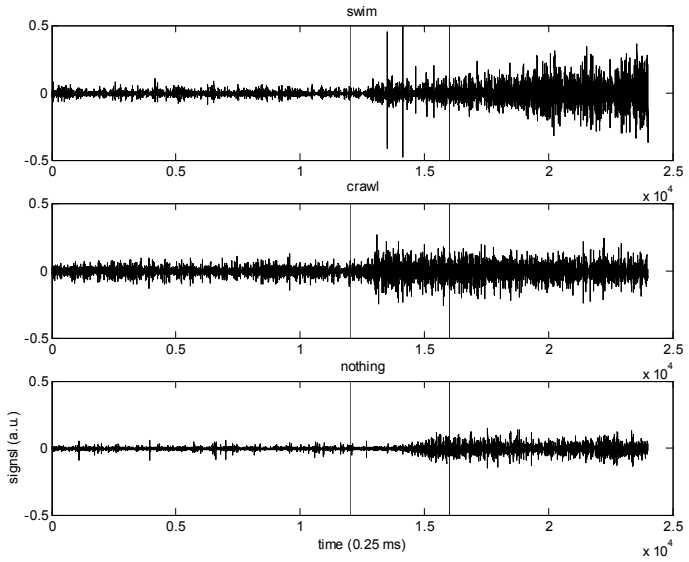

Figure 1. Electrical signals propagating in the ventral nerve cord of a leech before, during, and after electrical stimulation. The time axis is in units of 0.25 ms. The top signal preceded fictive swimming, the middle preceded fictive crawling, the bottom signal did not elicit either swimming or crawling. The beginning and the end of the stimulation are marked by the vertical bars.

a signal in a situation when the leech subsequently generated neural signals associated with swimming (*fictive swimming*), the middle trace is when the leech subsequently showed fictive crawling, the bottom trace is when the leech did nothing. The vertical lines (at $t=1.2\times10^4$ and $t=1.6\times10^4$ in units of 0.250 ms) mark the beginning and the end of the stimulation. The stimulation in each case consisted of the same amount of electrical current injected during a 1 s time interval.

Figure 1 shows that before stimulus onset, the data appear to be stationary, indicating that the usual, time-independent statistics may be sufficient. This, however, is not the case after the stimulus begins, when the time dependence of the signals clearly demands some time resolution.

Figure 2 shows the time-resolved variance of the signals for the three cases, calculated by using overlapping 400-point (100 ms) epochs. The starting points of successive epochs differed by 40 points (10 ms).

It is clear from Fig. 2 that before stimulus onset, the signals are relatively stationary relative to variance. The mean values for the three cases are all significantly different from each other, ordered as: *swim >crawl>no response* (*t*-test, $p<10^{-4}$ for all pairs). The differences in the ventral cord activity prior to stimulation could well determine the leech's subsequent response (Albano et al. 2006).

After stimulus onset, the crawl and swim variances rise to an essentially common maximum, attaining it in 290 ms. During this period, they are *not* statistically distinguishable. Past their maxima, the two signals begin to differ, eventually diverging at approximately 630 ms after stimulus onset. The possibility that the signals are indeed indistinguishable during the immediate post stimulus-onset period lends further support to the idea that it is the pre-stimulus state of the animal's nervous system that determines its eventual response.

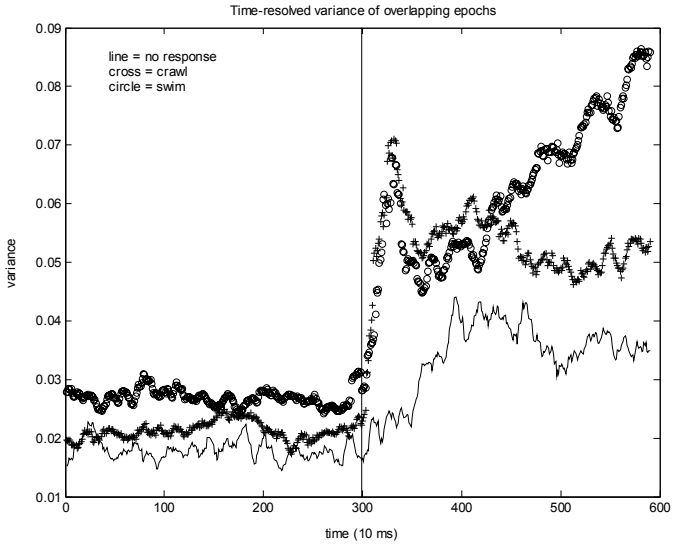

Figure 2. Time-resolved variance of signals propagating in the leech ventral cord before, during, and after electrical stimulation. These values were calculated using non-overlapping 100.0-ms windows. The time axis is in units of 10 ms. Results shown are averages of 5 trials in which the stimulus elicited no response, 13 trials that led to crawling, and 21 trials that led to swimming. Solid line = no response, crosses = crawl, circles = swim. Vertical line marks stimulus onset.

III. Fourier and Gabor

III.1 The Discrete Fourier Transform (DFT) (Press et al. 2002)

The Fourier Transform, $h(f)$, of a continuous variable, $x(t)$, is defined by

$$h(f)=\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{2i\pi ft}x(t)\,dt. \qquad (5)$$

When dealing with a time series, $\{x_1, x_2, \ldots, x_N\}$, a sequence of values of x measured at discrete values of time, $t_1=\tau$, $t_2=2\tau$, ..., $t_k=k\tau$,..., with sampling time, τ, the above definition needs to be modified. Instead of $h(f)$, which is a function of a continuous variable, f, we get a *Discrete Fourier Transform (DFT)*, h_n, a sequence of values corresponding to discrete frequencies, $f_n=n/N\tau$. The h_n's are given by,

$$h_n=\sum_{k=1}^{N}x_k e^{2i\pi kn/N}, \quad n=-N/2,-N/2+1,\ldots,0,\ldots N/2-1,N/2. \qquad (6)$$

For real-valued x's (which is all we care about here), $P(n)=c|h_n|^2$, $n=0, 1, \ldots, N/2$, is a measure of the amount of "energy" carried by x between the frequencies f_n and f_{n+1}. (This terminology arose from the use of the Fourier Transform in electrical circuits, in which x is either a voltage or a current and electrical power is proportional to the square of either of these quantities).

The constant, c, is a normalization factor, of which there are quite a few that are in common use (see, *.e.g*, Press et al 2002, Chap 13). Sometimes, the $P(n)$'s are called either the *power spectrum* or the *power spectral density (PSD)* depending on the value of c. Here, we use the value, $c=\Sigma_n|h_n|^2$ (see Eq. 7), so that the resulting $P(n)$'s represent the fraction of the energy in the frequency range (f_n, f_{n+1}) and, following Press et al 2002, we will use the two terms interchangeably.

III.1a. Discontinuities, detrending windowing, and aliasing

The Discrete Fourier Transform tries to do quite a lot. It uses a finite number of sampled points, $\{x_1, x_2, \ldots, x_N\}$, to get an approximation of the behavior of a function, $x(t)$, that depends on a continuous variable, t, which has an infinite domain, $-\infty<t<\infty$. This can work only if either x is nonzero just in the time interval sampled, or if the sample used is "typical" of all other intervals.

If one takes the view that $x(t)$ is nonzero only in the interval sampled, it means that if $x_1\neq0$ or $x_N\neq0$, then $x(t)$ changes drastically at one or the other boundary of the interval. If one takes the view that the

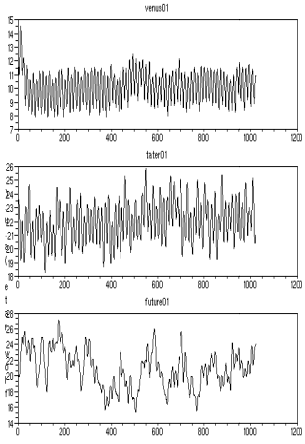

Figure 3. Flow rates for a normal (top), a prediabetic (middle) and a diabetic (bottom) monkey. The ordinate are in arbitrary flow rate units, the time axis (abscissae) spans 31.7 s in steps of 0.031 s.

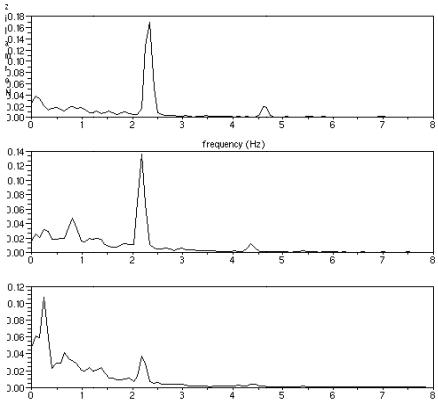

Figure 4. Normalized power spectral densities (PSD) of the time series shown in Fig. 3. Top: normal, middle: pre-diabetic, bottom: diabetic.

sample used is typical, then one needs to imagine $x(t)$ as an infinite sequence of repetitions of $\{x_1, x_2, \ldots, x_N\}$. In this case, if $x_1 \neq x_N$ then $x(t)$ changes discontinuously at both boundaries of the interval. If there is a linear trend in the data, then, in general, $x_1 \neq x_N$. Any discontinuous change at the boundaries results in anomalously large high-frequency contributions to the DFT.

This artifactual enhancement of high-frequency components because of edge effects can be mitigated by subtracting the mean and the linear trend (*detrending*) and by forcing the detrended time series to approach zero gently at both ends while keeping most of its values in the interior of the interval relatively unchanged. The latter can be done by multiplying the time series with a function that approaches zero gently at both ends, but which rises close or equal to one in most of the interior points. Such a function is called a *window function* (see Press et al. 2002 for some commonly used window functions). The DFT is then calculated for the time series that has been detrended, and then windowed (the order of the operations matters).

If the signal contains frequencies greater than $f_{N/2}=\frac{1}{2}\tau$, the *Nyquist critical frequency*, then the DFT folds these high-frequency contributions into the interval, $(0, f_{N/2})$. This phenomenon is called *aliasing*. Unfortunately, no post-measurement manipulations can mitigate aliasing. It can only be eliminated by low-pass filtering the data as they are measured (Press et al. 2002).

III.1b. Monkeys: vasomotion in sickness and in health

Vasomotion is the spontaneous rhythmic oscillation of vessel diameter, flow, or pressure in the peripheral blood vessels. Derangement of vasomotion has been previously suggested as a possible cause of diabetic neuropathy (Stansberry et al. 1996). We illustrate the use of the DFT on biological time series by analyzing vasomotion data from normal, prediabetic, and diabetic monkeys.

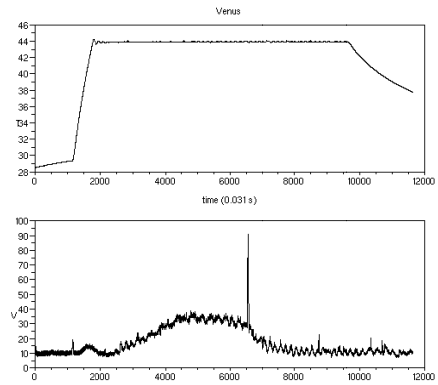

Figure 5. Time dependence of flow rate for a normal monkey subjected to a "heat treatment." Top panel shows the temperature (°C), bottom panel shows flow rate.

Figure 3 shows 400-point epochs of flow rate data for a normal (top), a prediabetic (middle), and a diabetic (bottom) monkey. The data were obtained using laser Doppler fluximetry from the dorsum of the foot of anesthetized monkeys after an overnight fast. The sampling time was 0.031 s. Even in these time series, there are visible differences in the predominant frequencies for the three metabolic states, differences which we quantify using the DFT.

In order to compare signals with different total energies, we normalize the PSD:

$$P_n = |h_n|^2 / \sum_{k=1}^{N/2} |h_k|^2, \tag{7}$$

with the h_k's as defined in Eq. (6). Fig. 4 shows normalized PSD's for the three time series shown in Fig. 3. The predominant spectral peaks are seen to be below 2.5 Hz. As the metabolic state changes from normal to prediabetic to diabetic, the largest peak moves to lower frequencies. This shift may represent a shift in the distribution of blood from nutritional (capillaries) to non-nutritional (arterio-venous shunts) vessels with progression of diabetes (Tigno et al. 2009)

III.2 The Gabor Transform

For data such as those shown in Fig. 3, the use of a single PSD to characterize the entire time series is likely to be adequate. Each of the time series shows oscillatory behavior, but there do not appear to be major qualitative changes in the character of the data. This is certainly not the case for the data shown on the lower panel of Fig. 5 where there are obvious changes not only in the magnitude, but also in the predominant frequency of the signal as time progresses. One way to keep track of these changes is to perform time-resolved spectral calculations using a moving window such as that used for calculating time-resolved statistics in Sec. II.3. The resulting sequence of Fourier Transforms is called a short-time Fourier Transform which provides a characterization of the data in both the time and frequency domains. It is particularly useful if the characteristic frequencies of the time series change in time. This procedure for obtaining a time-frequency representation of the signal was initially proposed by Gabor (Gabor 1946) using a gaussian window to evaluate the DFT's of the individual data segments. The result is usually called a Gabor transform, although the term is also used to refer to results obtained using other windows (Bastiaans and Geilen 1996). The Gabor transform is to spectra what time-resolved statistics are to statistics.

III.2a Thermally stressed monkeys

It is known that different metabolic risk groups respond differently to thermal stimulation (Tigno et al. 2003). An example of heat treatment data is shown in Fig. 5. The upper panel shows the temperature to which the monkey is subject while the lower panel shows the measured flow rates. In this example, vasomotion recording was started with the monkey at a temperature of 29.5°C. The temperature was then increased linearly until it reached some 44°C, and then held at that level for some 243 s, after which the monkey was allowed to cool down.

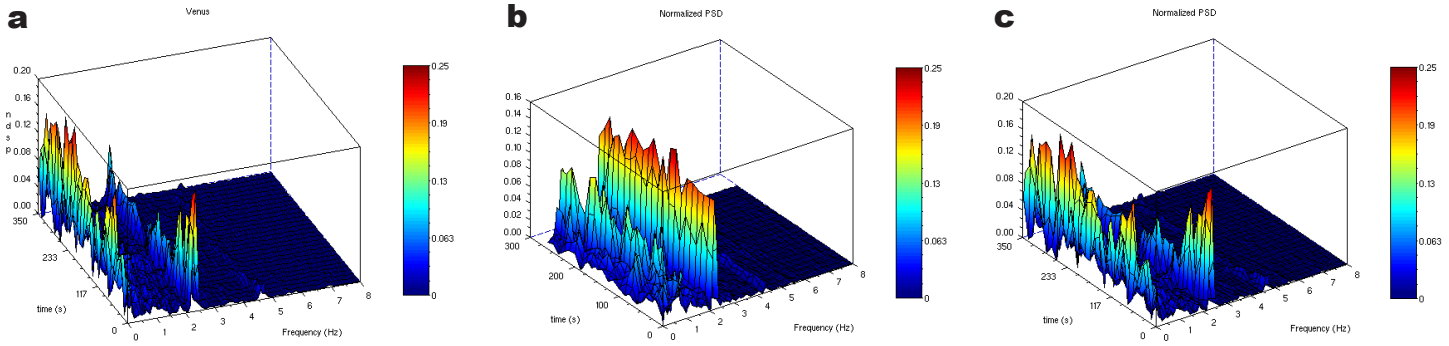

Figure 6. PSD surface plots. a, obtained from the Gabor Transform of the data from a normal monkey shown on the bottom panel of Fig. 5. The temperature ramp lasts from 34 s to 57 s; the temperature plateau ends at 298 s; **b**, from the Gabor Transform of data from a pre-diabetic monkey undergoing heat treatment; and **c,** from the Gabor Transform of data from a diabetic monkey undergoing heat treatment.

Figure 6a shows a surface plot of the normalized PSD from the Gabor Transform of the data shown in Fig. 5, calculated using contiguous, non-overlapping 400-point (12.4 s) windows and a Hanning window function. The temperature ramp lasts from 34 s to 57 s. At the start, the peak at ~2 Hz dominates the spectrum. As the temperature is increased, it diminishes as the peak at ~1 Hz gains prominence. Both peaks suffer a dip around the mid-point of the temperature plateau, and then both increase again with the 1 Hz peak still dominant. As the monkey cools off, the 2 Hz peak eventually becomes larger than the 1 Hz peak, but the former does not attain its starting value.

Figures 6b and c are surface plots of normalized PSD's from the Gabor transforms of a prediabetic and a diabetic monkey, respectively. The differences between these and Fig. 6a are quite dramatic. In the prediabetic monkey, the 2 Hz peak dominates throughout, while in the diabetic case, it is the 1 Hz peak that dominates although there are considerable variations in the heights of the two major peaks. Analysis of additional data is underway to determine if these behaviors are indeed typical.

IV. Correlation and Causality

Given a series of simultaneously measured values of two variables pertaining to the same system,

$$X = \{x_1, x_2, \ldots, x_N\} \text{ and } Y = \{y_1, y_2, \ldots, y_N\}, \tag{8}$$

there are a number of quantities that can be used to measure the extent to which the variables are correlated or to determine if there exists a causal relationship between them. Here, we restrict ourselves to a brief discussion of two measures of correlation and one of causality, and illustrate their use with human electroencephalographic (EEG) signals.

IV.1 Correlations

IV.1a Pearson's *r*

The *cross-correlation coefficient, r* (Pearson's *r*), has been long and widely used to measure correlation. It is a measure of how two variables track each other. The cross-correlation coefficient of *X* and *Y* (Eq. 8) is estimated by,

$$r = \frac{1}{N}\sum_{k=1}^{N}\frac{(x_k - \langle x \rangle)(y_k - \langle y \rangle)}{\sigma_x \sigma_y} \tag{9}$$

where $\langle x \rangle$ and $\langle y \rangle$ are the respective means of *X* and *Y*, and σ_x and σ_y their respective standard deviations. This measure, however, is limited by that fact that it is insensitive to nonlinear correlations. (Mars and Lopes da Silva 1987)

IV.1b Mutual Information

The *Mutual Information*, *I*(*X*,*Y*), of *X* and *Y*, is the amount of information common to *X* and *Y* (Shannon and Weaver 1949) and is estimated by

$$I(X,Y) = \sum p_{X,Y}(x_i, y_i)\log\left(\frac{p_{X,Y}(x_i, y_i)}{p_X(x_i)p_Y(y_i)}\right), \tag{10}$$

where $p_{X,Y}(x,y)$ is the joint probability distribution of *X* and *Y* and $p_X(x)$ and $p_Y(y)$ are their respective marginal probabilities. *I*(*X*,*Y*) is a measure of the amount of information about *X* given a measurement of *Y* and *vice versa*. More precisely, it is the average number of bits of *Y* that can be predicted by measuring *X*, where it can be shown that this is symmetrical *I*(*X*,*Y*)=*I*(*Y*,*X*). It is a generalization of measures of cross-correlation such as Pearson's *r* to the extent that unlike *r*, it is sensitive to nonlinear correlations (Fraser and Swinney 1986, Mars and Lopes da Silva 1987).

IV1c. Transfer Entropy

Correlations, linear or nonlinear, only indicate the extent to which two variables are similar. They say nothing about causal relationships. Neither can they indicate if two variables are similar not because they interact with each other, but because they are both driven by a third. These issues can be addressed by measures of causality or of information transfer. One such measure is Schreiber's Transfer Entropy (Schreiber 2000) defined by,

$$T_{Y\to X} = \sum_k p(x_{k+1}, x_k, y_k)\log_2\frac{p(x_{k+1}\,|\,x_k, y_k)}{p(x_{k+1}\,|\,x_k)} \tag{11}$$

where $p(x|y)$ are conditional probabilities. It is a measure of the influence of the sequence, *Y*, on the evolution of *X*. Hlavácková-Schindlera et at. 2007 showed that this is equivalent to the *conditional mutual information*, the mutual information of $\{x_1, x_2, \ldots, x_{N-1}\}$ and $\{x_2, x_3, \ldots, x_N\}$ given $\{y_1, y_2, \ldots, y_{N-1}\}$.

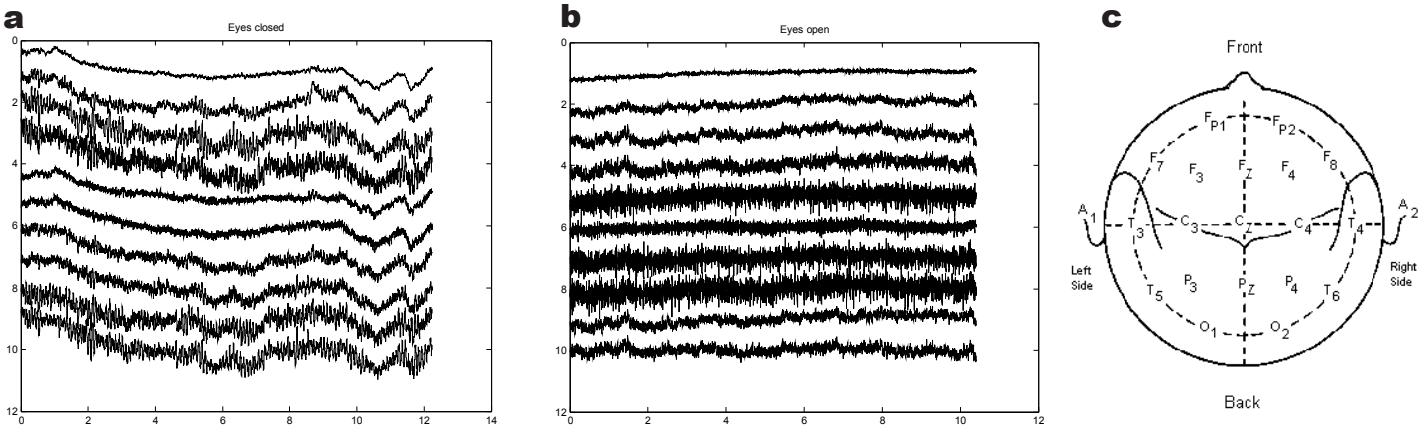

Figure 7. Electroencephalographic activity of a human subject. a, with eyes closed taken from scalp locations Fz, Cz, Pz, Oz, F3, F4, C3, C4 P3 and P4 (top to bottom); **b**, EEG of the same subject from the same scalp locations as Fig7a but with eyes open. **c**, Electrode locations on the scalp used for EEG measurements ("The 10-20 system"). Oz is on the mid-line between O1 and O2.

IV2. People: seeing vs. not

Vision takes up a lot of the human brain's resources, and this is indicated in part by differences in the free-running electroencephalographic (EEG) activity of a subject with eyes closed and that of the same subject with eyes open. Figs. 7a and b are 10-channel EEG recordings. These records are voltages measured at the scalp locations shown on Fig. 7c relative to linked earlobes.

Visual inspection of these graphs provide a lot of information to trained encephalograpers, but quantitative information about the data and about correlations and causal relations can only be provided by computational tools such as those described above.

Figures 8a and b show cross-correlation coefficients for all scalp site pairs of the data shown in Figs. 7a and b. In Fig. 8a, the color of a cell represents the cross-correlation coefficient of the two scalp sites corresponding to its row and column. Fig. 8b shows the cross-correlation coefficients of all scalp site pairs as a single line plot for each condition. With eyes closed, all scalp site pairs are highly correlated. In contrast, with eyes open, most are relatively uncorrelated except for Fz-Cz, Cz-Pz, Pz-Oz, P3-Cz, P3-Oz, and P3-Fz (points 1, 3, 6, , 30, 31, and 32 in Fig. 8b).

Figures 9a and b show the corresponding results for Mutual Information, calculated using an algorithm due to Cellucci (Celluci et al. 2005). These figures tell a story that is similar to that told by the cross-correlation coefficient. The patterns of linear and nonlinear correlations are similar.

Patterns of information transfer, however, are completely different from those of the correlations. Figs. 10a and 10b display values of *net* Transfer Entropy. That is, in Fig. 10a, the value associated with a cell at the m^{th} row and n^{th} column is the difference between the Transfer Entropy from m to n and that from n to m. A positive value indicates that information is being transferred from m to n, a negative value is information transfer is the other way around. In Fig. 10, note that the color bar scales are different.

In the eyes closed condition, the Transfer Entropy oscillates in the range [-0.325, 0.228] with a mean value of -0.064. In the eyes open condition the corresponding values are [-1.0243, 0.8783] and -0.254. The largest Net Transfer Entropy values for the eyes open condition are over five times greater than the standard deviation of the eyes closed condition. There is, indeed, much greater information transfer among brain areas when one sees.

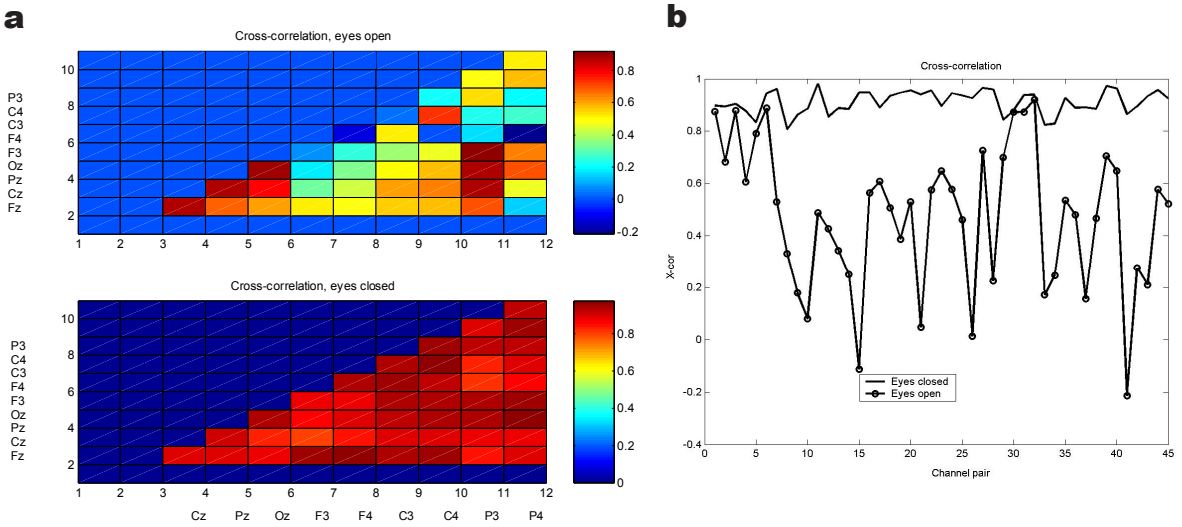

Figure 8. Cross-correlation coefficients (Pearson's r) of the EEGs shown in Figs. 7a and 7b. a, Top: eyes open, bottom: eyes closed. Rows and columns are identified by the 10-20 system (Fig 7c) designations of scalp electrode sites. The color of each cell represents, using the scale on the color bar on the right, the cross-correlation coefficient of the sites identified by the row and the column of the cell; **b**, The cross-correlation coefficients of Fig. 8a shown as a single line plot Each point corresponds to a cell in the arrays shown in Fig. 8a. Solid line = Eyes closed, Circles = Eyes open.

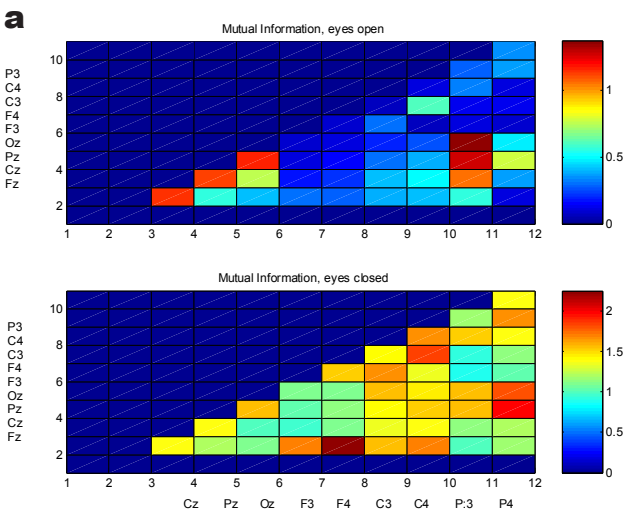

Figure 9. Mutual Information of the data shown on Figs. 7a and 7b. a, Top: eyes open, bottom: eyes closed. See the caption of Fig. 8a for details of the color coding. **b**, Mutual Information values from Fig. 8a shown as a single line plot. Each point corresponds to a cell in the arrays shown in Fig. 8a. Solid line = Eyes closed, Circles = Eyes open.

Figure 10a shows that the largest amounts of Information Transfer are from F3 and F4 to Fz, Cz and Pz and from F3, F4, C3 and C4 to P3. The contrast between these results and those for cross-correlation and Mutual Information shows that two highly correlated data streams may not in fact be participating in a net transfer of information. They may well be receiving the same information from a common source.

V. Complexity and Predictability

V.1 Complexity

The algorithmic complexity of a symbol sequence is the length of the shortest instruction set needed to reconstruct it. If a symbol-based complexity measure is to be applied to a time series, the time series must be converted to a symbol sequence. The easiest way to do this is to map the original real value of the time series to symbol "1" if its value is greater than the median and to symbol "0" if its value is less than the median—a binary partition. The choice of the median rather than the mean is important (Rapp et al. 1994). For a series of a given length, a random sequence has the highest complexity because there are no rules that would reproduce it. An ordered sequence is less complex than a random one—the more ordered, the less complex. There are many ways of estimating complexity (see, *e.g.*, Rapp and Schmah 1996). Regardless of how it is estimated, it provides a sequence-dependent measure of how much the structure of the time series differs from that of a random sequence with the same distribution of symbols (Watanabe et al. 2003).

In the following, we use a definition by Lempel and Ziv (Lempel and Ziv 1976). To calculate the Lempel-Ziv complexity, the time series is first mapped into a binary partition. Then, the resulting binary symbol sequence is expressed as a unique set of subsequences using a procedure we describe below. The Lempel-Ziv complexity is equal to the number of these subsequences.

We illustrate the Lempel-Ziv procedure by adapting an example in Watanabe et al. 2003. Given a time series that has been reduced to the binary symbol sequence,

$$X=\{x_1,x_2,\ldots,x_N\} = 0001101001000101, \qquad (12)$$

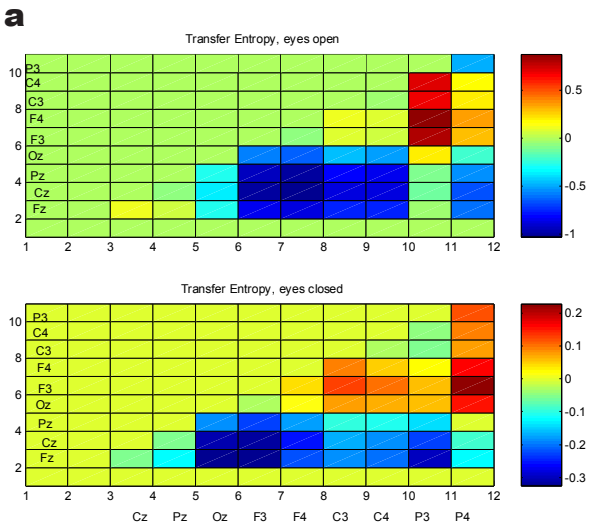

Figure 10. Net Transfer Entropy of the data shown on Figs. 7a and b. a, Top: eyes open, bottom: eyes closed. See the caption of Fig. 8a for details of the color coding. Note the difference between the color scales of the two panels; **b**, Net Transfer Entropy values from Fig. 9a shown as a single line plot. Each point corresponds to a cell in the arrays shown in Fig. 9a. Solid line = Eyes closed, Circles = Eyes open.

We want to express X as

$$X = X_1X_2\ldots X_c, \tag{13}$$

where $X_1=x_1$, the first element of X, and the X_k, k=2, …, c are unique subsequences of X obtained as follows:

1. Start building X_2 by taking the next element of X, *i.e.*, x_2=0: X_2=0.
2. Concatenate X_1 and X_2 to get X_1X_2=00.
3. Remove the last element of X_1X_2 to get $(X_1X_2)_\pi$=0. If the current value of X_2 appears in $(X_1X_2)_\pi$, then append the next element of X to X_2, otherwise, start building X_3. In the present case, the current value of X_2=0 is contained in $(X_1X_2)_\pi$=0, so we append the next value, x_3=0, to X_2 to get a new subsequence, X_2=00.
4. Repeat 2 and 3 to get the new values, X_1X_2=000, $(X_1X_2)_\pi$=00. X_2 is contained in $(X_1X_2)_\pi$, so we continue building X_2 by appending x_4=1, and get the new X_2=001.
5. Repeat 2 and 3 to get the new values, X_1X_2=0001, $(X_1X_2)_\pi$=000. X_2 is no longer contained in $(X_1X_2)_\pi$, so X_2 is complete, and we start building X_3.
6. Start by taking $X_3=x_5$=1, concatenate X_1, X_2, and X_3 to get $X_1X_2X_3$=00011; remove the last element to get $(X_1X_2X_3)_\pi$=0001. We note that X_3=1 is contained in $(X_1X_2X_3)_\pi$ so we add the next element, x_6=0, to get X_3=10. We get new values of $X_1X_2X_3$ and $(X_1X_2X_3)_\pi$ and note that the current X_3=10 is not contained in $(X_1X_2X_3)_\pi$, so X_3 is complete, and we start building X_4.

We continue with the process until we come to the end of X with the result that we can write X as the sequence of subsequences,

$$X = (0)(001)(10)(100)(1000)(101). \tag{14}$$

Thus, the complexity of X is 6.

The Lempel-Ziv complexity is sensitive to the length of the time sequence, so care must be taken when comparing sequences of different lengths. A generalization that is not epoch length dependent is *redundancy*, defined in (Rapp et al. 2001) as

$$R = 1 - \frac{C_{orig}}{<C_o>} \tag{15}$$

where C_{orig} is the complexity of the data and $<C_o>$ is the average complexity of a number of "equiprobable surrogates" of the same length as the original. In the case of the binary partition used here, an equiprobable surrogate is just a random sequence of 0's and 1's in which there are as many 0's as there are 1's.

V.1a Vision is complex

Evaluating the Lempel-Ziv complexity of the EEG data shown in Figs. 7a and b using a binary partition requires some care. Long-term trends, as in the case of the eyes closed data, could be due to drifts in the electronics of the measuring instruments. Performing a binary partition about the median of an entire time series with a linear trend could, in the worst case, put the first half of the time series below the median and the second half above. In the presence of long-term trends, rather than calculating the complexity of an entire time series, it is prudent to calculate, instead, the average complexity of a number of epochs that sample the whole time series.

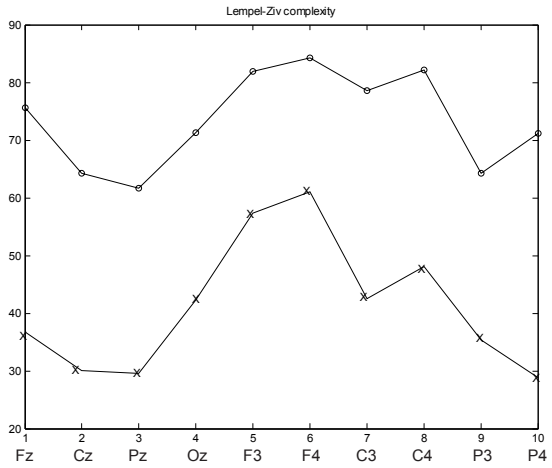

Figure 11. Average Lempel-Ziv complexity of ten 1000-point epochs of the EEG data shown in Figs. 9 and 10 at each scalp site. x : eyes closed, dots: eyes open.

To estimate the complexity of the EEG data of Figs. 7a and b, we used ten equally-spaced non-overlapping 1000-point epochs that span the entire time series. Each epoch is detrended before its Lempel-Ziv complexity is calculated. The results are shown in Fig. 11. At each scalp site, the complexity of the EEG signal with eyes open greatly exceeds that for eyes closed. This is in agreement with the redundancy calculations reported by Watanabe et al 2003.

V.2 Predictability

If a time series is generated by a deterministic dynamical system, then it ought to be possible to predict its future from its past, at least in the short run. This notion serves as the basis of a measure that can be used for testing the possibility that a time series is deterministic —that is, to find out how well its future can be predicted from its past. A commonly used algorithm makes use of a procedure similar to one that used to be popular in weather forecasting. To predict tomorrow's weather, look for a day in the past when the conditions were closest to today's, and then predict that tomorrow's weather will be like the day that followed *that* day in the past.

Predicting values of a time series, $X=\{x_1, x_2, \ldots, x_N\}$, is just slightly more complicated than that old weather forecaster's trick. Instead of trying to match x_{k-1} in order to predict x_k, one may try to match the pair, (x_{k-1}, x_{k-2}). This way, one uses an approximation not just of the value of x, but also of its first derivative; or one may try to match $(x_{k-1}, x_{k-2}, x_{k-3})$, to include the value of x as well as its first and second derivatives, etc.

Predicting the k^{th} element of X, then, involves the following steps:

1. Take the m previous elements, $X^m_k=\{x_{k-m}, x_{k-m+1}, \ldots, x_{k-1}\}$. The value of m depends on how many of x's derivatives one wishes to include.[1]

[1]The m-tuples, $X^m_k=(x_{k-m}, x_{k-m+1}, \ldots, x_{k-1})$ are vectors of an m-dimensional time-delay embedding of X. There exists a considerable literature on criteria for choosing m. See, *e.g.* Abarbanel 1996 and Cellucci et al. 2003.

2. For all previous sequences of m elements, $X^m_j=\{x_{j-m}, x_{j-m+1}, \ldots, x_{j-1}\}$, $m+1<j<k$, calculate the distance,

$$d_{k,j}=\left(\sum_{p=1}^{m}\left(x_{k-p}-x_{j-p}\right)^2\right)^{1/2}.$$

Let X^m_Q be that sequence for which $d_{k,Q}=\min(d_{kj})$. X^m_Q is the nearest neighbor of X^m_j.

3. The predicted value of x_k, $(x_k)_{\text{pred}}=x_Q$.

We get a trivial prediction if X^m_Q is the sequence immediately preceding X^m_k, or is one of the others in X^m_k's recent past. To exclude sequences that are very close in time to X^m_k, we define an interval called the "blind" preceding x_k, which may be taken to be equal to the autocorrelation time of the time series. All points inside the blind are excluded in the search for x_Q.

The quality of a set of predictions is assessed by the *prediction error*,

$$err=\frac{\frac{1}{n_p}\sum_{k=1}^{n_p}\left|x_k-(x_k)_{pred}\right|}{\sigma_x}, \qquad (16)$$

where n_p is the number of predictions made and σ_x is the standard deviation of the x's.

V.2a Vision is unpredictable

Evaluating the predictability, or more precisely the prediction error, of the EEG's of Figs. 7a and b presents the same problems as evaluating their complexity. When there are long-term trends that may be due to external causes, the process of finding nearest neighbors may be compromised.

As in the case of complexity, we used ten detrended non-overlapping 1000-point epochs that sample the entire time series and calculate their average prediction error. The results are shown in Fig. 12. The prediction errors for the eyes open condition are consistently and significantly larger than those for eyes closed. This is not unexpected since, with the eyes open, the brain is receiving considerably more external stimuli leading to changes in brain potentials that are not solely determined by the brain's previous state.

Not surprisingly, Lempel-Ziv complexity and prediction error are highly correlated—the more complex, the less predictable. Their cross-correlation coefficient for the eyes closed condition is 0.97 while that for the eyes open condition is 0.98.

V.3 Covariance complexity

An alternative approach to calculating the complexity of a multichannel signal can be obtained by estimating its covariance complexity. This is a measure of the dimensionality of the original multichannel signal. Let V^k_j be the voltage measurement of channel k at time t_j. This multichannel record can be expressed as a matrix A.

$$A=\begin{pmatrix} V^1_1 & V^2_1 & \cdots & V^Z_1 \\ V^1_2 & V^2_2 & & V^Z_2 \\ \vdots & \vdots & & \vdots \\ V^1_N & V^2_N & & V^Z_N \end{pmatrix}$$

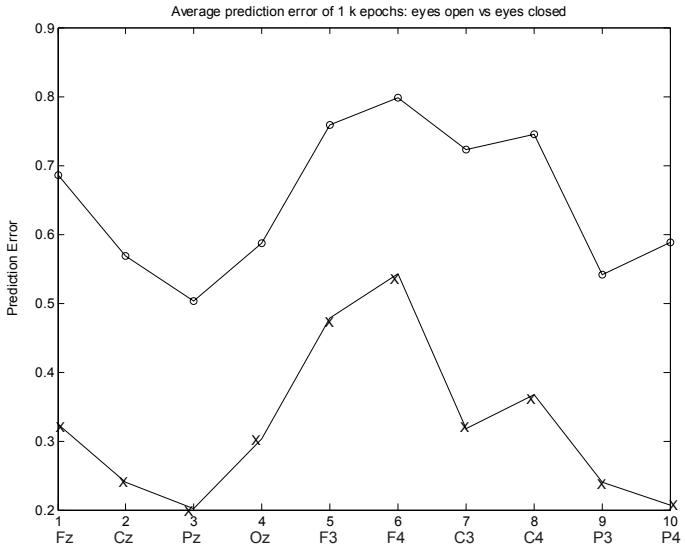

Figure 12. Average prediction error of ten 1000-point epochs at each scalp site of the EEG data shown in Figs. 7a and b. x: eyes closed, dots: eyes open.

Matrix A can be re-expressed using its singular value decomposition (see, *e.g.*, Golub and van Loan 1983).

$$A = VDU^T$$

Matrix D is important for determining the covariance complexity. Matrix D is a diagonal matrix containing the singular values, $D = \text{diag}(\lambda_1, \lambda_2, \ldots\ldots\lambda_Z)$, where we order the columns of the decomposition so that $\lambda_i \geq \lambda_{i+1}$. The fraction of total variance of the original z-channel signal in the k-th principal component is

$$\text{Var}_k = \lambda_k^2 \Big/ \sum_{j=1}^{Z}\lambda_j^2$$

In nonrandom multichannel signals, like multichannel EEGs, variance is not uniformly distributed across all of the principal components. A significant fraction can be concentrated in the first component. For the ten dimensional EEGs considered here, the fraction of variance in the first component is 0.753±0.086 in the eyes closed condition and 0.681±0.123 in the eyes open condition (Rapp, et al., 2005).

The covariance complexity measures the distribution of variance across principal components.

$$C_{\text{cov}} = -\left\{\sum_{j=1}^{Z}\text{Var}_j \log \text{Var}_j\right\} / \log Z$$

If 100% of the signal variance is in the first component, C_{cov}=0. If variance is spread equally across all components, which happens with random signals, C_{cov}=1. Using the ten channel EEG data described here C_{cov}=0.396±0.100 in the eyes closed condition and C_{cov}=0.469±0.102 in the eyes open condition (Rapp et al. 2005).

VI. Hypothesis testing: Scientist, falsify thyself!

Initial analysis may suggest a hypothesis about the nature of a data set or of the process or the system that generated the data. If so, it is important to verify if the hypothesis is indeed correct. That is, in the spirit that the philosopher Karl Popper described as the "falsifiability" of science, one sets up a procedure that makes it possible to determine if the hypothesis is false. Statisticians call this *hypothesis testing*. It consists of the following (see, *e.g.*, Theiler et al. 1992, Rapp et al. 2001)

(1) Make a *null hypothesis* about the data,

(2) Perform a transformation of the data that preserves properties consistent with the hypothesis but destroys those that are not. The transformed data set is called a *surrogate*.

(3) Calculate a measure that is sensitive to the destroyed properties (a *discriminating statistic*) using the original data and several realizations of the surrogate, then use the results to determine the likelihood that the original data and the surrogates belong to the same class. If it is sufficiently unlikely that the original and the surrogates belong to the same class, then the null hypothesis is rejected.

VI.1 Surrogates

Each hypothesis about the nature of a data set requires its own surrogate. Here, we discuss only a few that are commonly used and illustrate their use with a small subset of the EEG data shown in Figs. 7a and b. See Theiler 1992, Kantz and Schreiber 1997, Rapp 2001 for more details and other surrogates.

VI.1a Random shuffle

If the null hypothesis is that the data are independent and identically distributed random numbers (*iid*)-called *white noise* by physicists and engineers—we create the appropriate surrogates by merely randomly shuffling the data. This destroys all information that depends on the time sequence of the data. Fig. 13a shows a 1500-point segment of EEG from location Cz of the eyes closed data of Fig. 7a. A random shuffle surrogate is shown in Fig. 13b. The two are clearly different. In fact, visual inspection of the top panel is enough to establish that it is not white noise.

VI.1b Small shuffle

As shown by Figs 13a and b, a random shuffle is a pretty drastic operation. The only things it leaves unchanged are identically distributed random numbers (iid). Many interesting data, such as the examples we have considered so far, are usually not iid.

A less drastic shuffle, due to Nakamura and Small 2005 and which they call the "Small shuffle," keeps the long-term trends intact, but randomizes local fluctuations. This is achieved by perturbing the time order of small segments of the data, but leaving the over-all structure unchanged. A number drawn from a normal distribution with unit variance is added to each of the indexes of the data. Since the additions are normally distributed, small changes are more likely than big ones. The perturbed indexes are sorted, but since the perturbations are typically small, the sorted list will differ only locally from the original. The indexes of the sorted list are then used to index the original time series to obtain the surrogate.

Figure 13c shows a Small shuffle surrogate of the data in Fig. 13a. Fig. 13c clearly preserves the long-term trends. Its local fluctuations are random, but this is something that needs some analysis to establish.

VI.1c Random phase surrogates

White noise has a power spectrum that undergoes random oscillations about a more or less level value. There are so-called *colored noises*, random time series whose power spectra are more structured than white

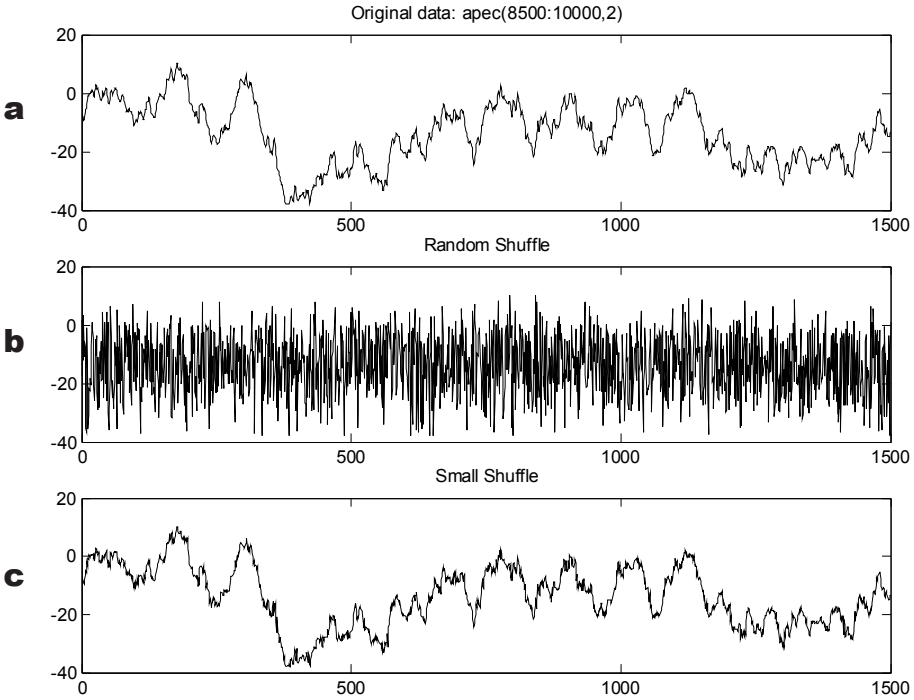

Figure 13. a, A 1500-point segment of the EEG data from scalp site Cz shown in Fig. 7a; **b**, Random shuffle; and **c**, Small shuffle surrogates of the data shown in Fig. 13a.

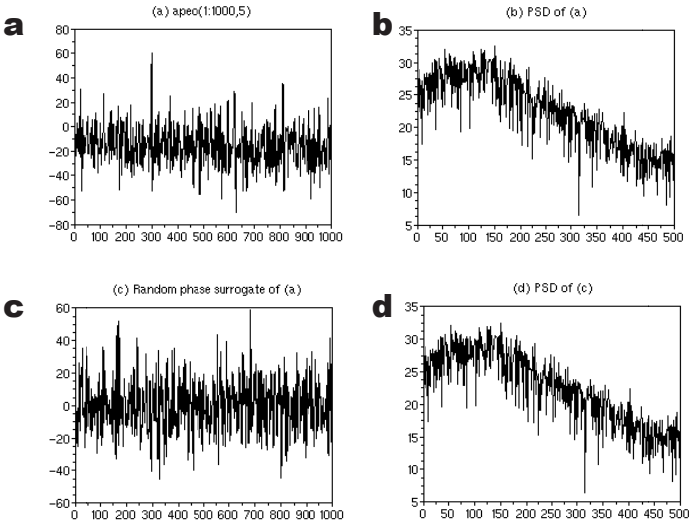

Figure 14. a, A 1000-point segment of the EEG recorded at site F3 in the "eyes open" condition (Fig. 7b); **b**, PSD of Fig. 14a; **c**, A random phase surrogate of Fig. 14a; **d**, PSD of Fig. 14c.

noise. These time series can be generated by subjecting random numbers to linear filters. To test the hypothesis that a time series is linearly filtered noise (also called linearly correlated noise), we need surrogate data that have the same power spectra as the original time series, but which are otherwise random. These are called *random-phase surrogates*.

The power spectrum does not depend on the phases of the Fourier coefficients. This provides a way for generating surrogates that have the same power spectra as an original data set, but which are otherwise random. It goes as follows:

1. Calculate the Fourier Transform of the original signal—*i.e.*, generate its Fourier coefficients,
2. Replace the phases of the coefficients by random angles.
3. Calculate the inverse Fourier Transform of the coefficients with random phases to get a random phase surrogate.

The above operations leave the magnitudes of the Fourier co-efficients unchanged, so the power spectrum of the random phase surrogate is the same as that of the original, but since the phases have been randomized, the surrogate is now random. In particular, since the surrogate is linearly correlated noise, any nonlinear correlations of the original data are destroyed.

It is seen that the calculation of the original signal's Fourier transform and its inverse are the critical steps in the construction of a random phase surrogate. It has been shown that seemingly small-scale numerical errors in the calculation of the Fourier transform can result in a false-positive rejection of the surrogate null hypothesis. That is, the surrogate calculations can suggest the presence of a deterministic nonlinear structure in a signal constructed by linearly filtering random numbers. Procedures that can reduce the possibility of these failures have been identified (Rapp et al. 2001).

Figure 14 illustrates the properties of the random phase surrogate. (a) shows the first 1000 points of the eyes open EEG from scalp site F3 in Fig. 7b; (b) is the PSD of (a). (c) is a random phase surrogate of (a), (d) is the PSD of (c). Note that (a) and (c) are different, but that (b) and (d) are identical.

In the case of multivariate data, if the phases of the Fourier co-efficients corresponding to the same frequency are replaced by the same random angles for all variates, then not only are the PSD's of the individual variates preserved, so are their cross-correlations (Pritchard and Theiler 1994). This makes multivariate random phase surrogates constructed in this manner appropriate for testing the null hypothesis that the correlations of the different data streams are linear.

VI.1d Gaussian-scaled surrogates

Suppose that linearly correlated noise, x, is measured using an instrument that outputs

$$y = h(x) \tag{17}$$

The function, h is called an *observation function* or *measurement function*. If h is linear, then y is also linearly correlated noise. Otherwise, y becomes another kind of noise that must be modeled by a different kind of surrogate. A commonly used case is that for which h is *static, monotonic,* and *nonlinear*, in which case one gets *gaussian-scaled surrogates*.

Surrogates of this type are obtained by another controlled shuffle. These surrogates have exactly the same statistics as the original but their spectra are not the same. They are obtained as follows:

1. Given a time series, $\{x_1, x_2, \ldots, x_N\}$, generate a set of gaussian distributed random numbers, $\{y_1, y_2, \ldots, y_N\}$ and then arrange the y's so that they are in the same rank order as the x's.
2. Generate $\{y_1^s, y_2^s, \ldots, y_N^s\}$, a random phase surrogate of the y's,
3. Rearrange the x's so that they have the same rank order as the y's. The resulting set, call it $\{x'\}$, is a *gaussian-scaled surrogate* of $\{x\}$.

Since this is a controlled shuffle of the x's, the surrogate has the same distribution function, hence the same statistics as the original. But not the same power spectrum. For surrogates that preserve both the power spectrum and the distribution, see Schreiber and Schmitz 1996.

VI.2 Z-scores, Probabilities and Monte Carlo

Having constructed a surrogate for testing a particular null hypothesis, it then becomes necessary to calculate the probability that the original data and the surrogates belong to the same population.

Let us denote a discrimitating statistic by M, let M_{orig} be its value for the original time series and let $M_{\text{surr},1}, M_{\text{surr},2}, \ldots, M_{\text{surr},N}$ be values characterizing N implementations of the surrogate. We denote by $\langle M_{\text{surr}} \rangle$ and σ_{surr}, respectively, the mean and standard deviation of the $M_{\text{surr},i}$'s. A commonly used measure of the separation between M_{orig} and $\langle M_{\text{surr}} \rangle$ is the *z-score*,

$$Z = \frac{M_{orig} - \langle M_{surr} \rangle}{\sigma_{surr}}. \tag{18}$$

Physicists have a more descriptive term for this. They call it the "number of sigmas" that the original differs from the surrogates. If the values of M for the surrogates are normally distributed, for a z-score of Z, the null hypothesis can be rejected with a confidence level of

$$c = \frac{1}{2}\left(1 + erf\left(|Z| / \sqrt{2}\right)\right), \tag{19}$$

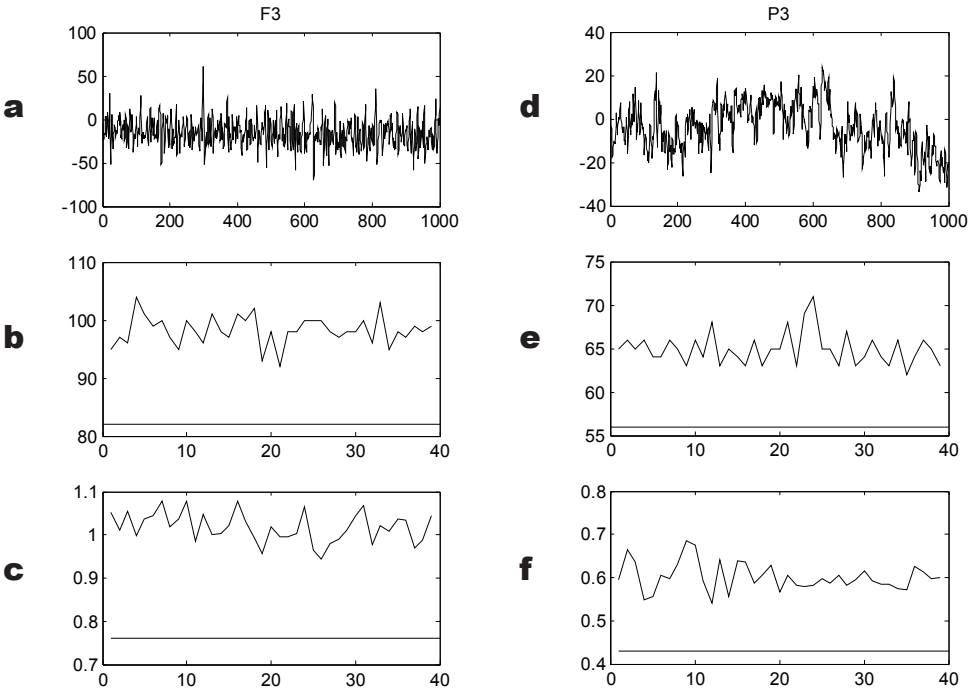

Figure 15. **a**, 1000-point EEG segment from F3 with eyes open; **b**, Lempel-Ziv complexity of 39 Small shuffle surrogates of Fig. 15a; **c**, Prediction errors of 39 Small shuffle surrogates of Fig. 15a; **d**, a 1000-point EEG segment from P3 with eyes open; **e**, Lempel-Ziv complexity of 39 Small shuffle surrogates of Fig. 15d; **f**, Prediction errors of pf 39 Small shuffle surrogates of Fig. 15d. In the two bottom rows, the level lines are the values obtained using the original data.

where erf(.) is the error function. In this case, with a Z-score of 2 the null hypothesis can be rejected at a 97% confidence level.

If the M_{surr}'s are not distributed normally, a *Monte Carlo probability* is usually used to estimate the likelihood that the null hypothesis is not true. For a so-called two-sided test, arrange the M_{surr}'s and M_{orig} in rank order. If, in this list, M_o is the n_{MC}^{th} largest or smallest, then the null hypothesis is *rejected* with a confidence level (1-s)×100 % where s is defined by (Theiler and Prichard 1997)

$$s = \frac{2n_{MC}}{N+1}. \qquad (20)$$

VI.2a EEGs tested

1. Single channels

We use 1000-point segments of the EEG data from sites F3 and P3 in the eyes open condition (Fig. 7b) to test a few hypotheses about the nature of the signals from individual scalp sites and about their relationships with each other. We chose these two sites as examples specifically because they have the largest value of net transfer entropy.

We start with the null hypothesis that local fluctuations of the EEG are random. If this is true, then a Small shuffle should not affect the values of such measures as Lempel-Ziv complexity and Prediction error, so we use these as our discriminating statistics.

The top row of Fig. 15 shows the data from F3 and P3 (eyes open). The middle and bottom rows are, respectively, values of the Lempel-Ziv complexity and of the Prediction error for 39 realizations of the Small shuffle for each of the data files. The level lines in these graphs indicate M_{orig} for each case. According to Eq (20), the null hypothesis can be rejected at a better than 95% Monte Carlo confidence level. Assuming that the Complexity and Prediction errors of the surrogates are normally distributed, Eq. (19) implies that the null hypothesis can be rejected almost with certainty.

We get a different story when testing the hypothesis that the EEG data are linearly correlated noises, using random phase surrogates and again using Lempel-Ziv complexity and prediction error as discriminating statistics.

The top panel of Fig. 16 shows the values of Lempel-Ziv Complexity (left) and Prediction error (right) for 39 random phase surrogates of the F3 data shown in Fig. 15a. In both cases, the Monte Carlo confidence level for rejecting the null hypothesis is 5%. The confidence level for rejecting the null hypothesis, assuming normal distribution of the surrogate discriminating statistics are 0.57 for Complexity and 0.63 for Prediction error. The null hypothesis cannot be rejected.

In the case of P3, however, the results are ambiguous. Using complexity as the discriminating statistic, the null hypothesis can be rejected at a Monte Carlo confidence level of 30% or a gaussian confidence level of 50%. On the other hand, using prediction error as the discriminating statistic, the null hypothesis can be rejected at a Monte Carlo confidence level of 95% and at a gaussian confidence level of almost certainty.

The foregoing is corroborated by results obtained using gaussian-scaled surrogates. If F3 is indistinguishable from linearly correlated noise, then it should be possible to distinguish it from linearly correlated noise that has been passed through a static, monotonic, nonlinear function. Indeed, the results shown on the upper panels of Fig. 17 indicate that, with complexity as a discriminating statistic (upper left), the hypothesis that F3 is this kind of nonlinearly filtered noise can be rejected at a Monte Carlo confidence level of 90%, or a gaussian confidence level of 93%. With prediction error as a discriminating statistic (upper right), the corresponding confidence levels for rejecting the null hypothesis are 90% and 97% respectively.

In the case of P3, the ambiguity resulting from the use of random phase surrogates persists when gaussian-scaled surrogates are used, as shown by the lower panels of Fig. 17. With complexity as a discriminat-

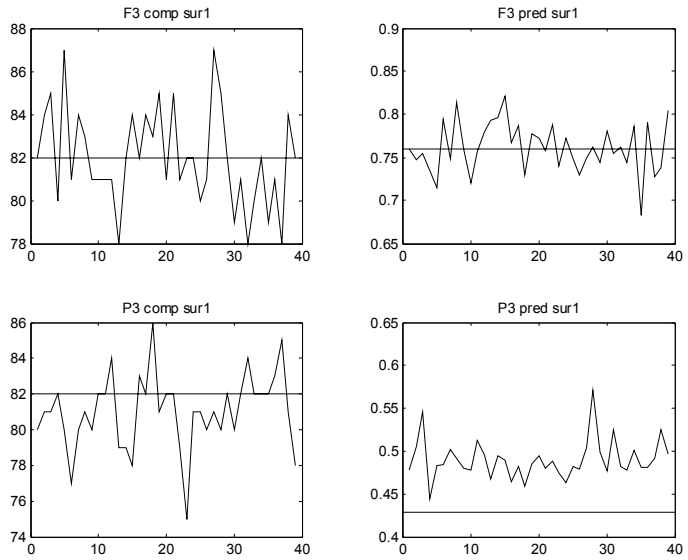

Figure 16. Lempel-Ziv complexity (left) and prediction error (right) of 39 random phase surrogates of the F3 (top) and P3 (bottom) data shown in Fig. 15. The level lines indicate the values for the original data.

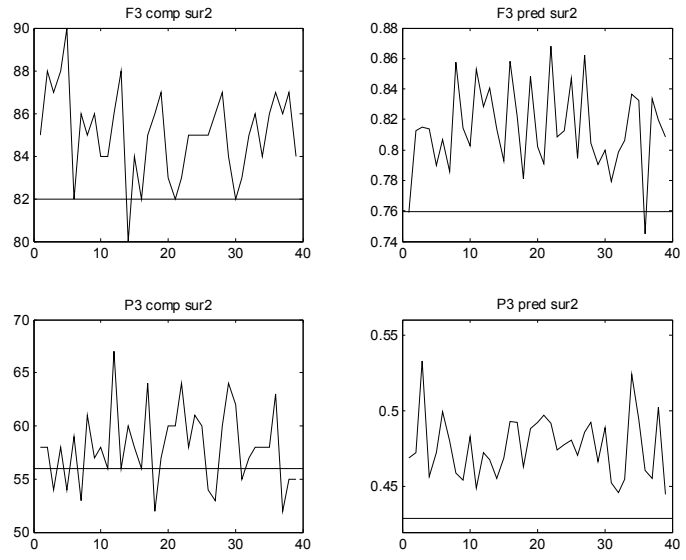

Figure 17. Lempel-Ziv complexity (left) and prediction error (right) of 39 gaussian-scaled surrogates of the F3 (top) and P3 (bottom) data shown in Fig. 15. The level lines indicate the values for the original data.

ing statistic (lower left), the Monte Carlo confidence level for rejecting the null hypothesis is 50%, the gaussian confidence level is 72%. On the other hand, with prediction error as a discriminating statistic (lower right), the null hypothesis can be rejected with almost certainty.

The above should serve as reminders that the results of hypothesis testing are functions of both the surrogate and the discriminating statistic (see, *e.g.*, Theiler and Rapp 1996).

2. Correlations and information transfer

We now test the hypothesis that the correlations between F3 and P3 data *are* linear by using bivariate random phase surrogates with Mutual Information and Transfer Entropy as discriminating statistics. By construction, the linear cross-correlation of F3 and P3 is identical to those of their bivariate random phase surrogates.

Figure 18 shows values of Mutual Information (left) and of Transfer Entropy (right) for 39 bivariate random phase surrogates of the F3 and P3 data of Fig. 12. Using these values in Eq. (20) says that with Mutual Information as the discriminating statistic, the null hypothesis can be rejected with a Monte Carlo confidence level of 85% or a gaussian confidence level of 90%. Using Transfer Entropy as a discriminating statistic, the null hypothesis cannot be rejected with similarly high confidence levels: 72 % for Monte Carlo and 55% for gaussian. The mutual information of F3 and P3, and the information transfer between them cannot be accounted for solely by linear processes.

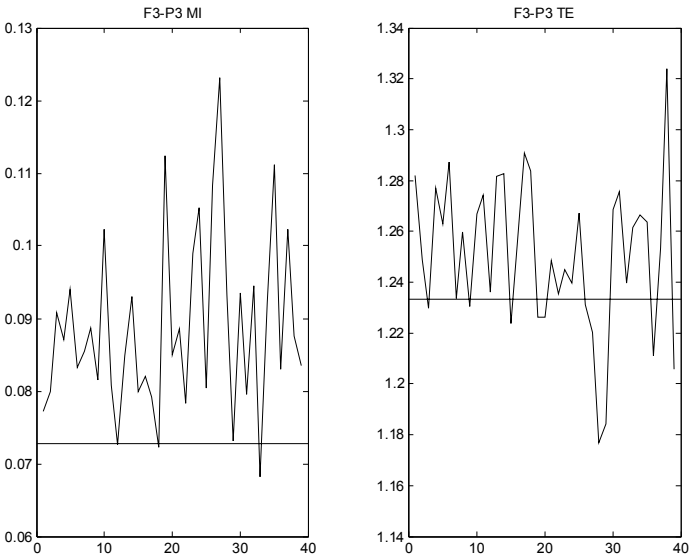

Figure 18. Mutual Information (left) Transfer Entropy (right) of 39 bivariate random phase surrogates of the F3 and P3 data shown in Fig. 15a. The level lines indicate the values for the original data.

VII. Concluding remarks

It has been said that identifying a phenomenon as non-linear is similar to identifying an animal as a non-elephant. There are vastly more of one kind than of the other. In the same spirit, one may say that if it's static, it's probably dead. In the analysis of biological data, linear and static measures by themselves are likely to miss a large part of the story. On the other hand, there is a lot of important and relatively easily calculable information obtainable from linear and static measures. It is prudent, always, to let the data suggest what analytical tools are needed to unfold their story.

Acknowledgements. Parts of this work were presented at the 2008 meetings of the Samahang Pisika ng Visayas at Mindanao and of the Samahang Pisika ng Pilipinas. AMA's attendance at these meetings was made possible by the Balik Scientist Program of the Department of Science and Technology, Republic of the Philippines. This research was supported in part by the Traumatic Injury Research Program of the Uniformed Services University (Bethesda, MD, USA.) The opinions and assertions contained herein are the private ones of the authors and are not to be construed as official or reflecting the views of the U. S. Department of Defense.

References

Abarbanel, HDI. Analysis of Observed Chaotic Data. New York: Springer-Verlag, 1996.

Albano AM, Brodfuehrer PD, Tapyrik L, Sunder S. Linear and nonlinear properties of prestimulus ventral cord signals distinguish swimming response of the leech to intracellular stimulation. Int J Bifurcation and Chaos 2006; 16:45-155.

Bastiaans MJ, Geilen MCW. On the discrete Gabor transform and the discrete Zak transform. Signal Process 1996; 49: 151-166.

Brodfuehrer PD, McCormick K, Tapyrik L, Albano AM, Graybeal C. Activation of two forms of locomotion by a previously identified trigger interneuron for swimming in the medicinal leech. Invert Neurosci 2008; 8:31-39.

Cellucci CJ, Albano AM, Rapp PE. Comparative Study of Embedding Methods. Phys. Rev. E 2003; 67: 66210.

Cellucci CJ, Albano AM, Rapp PE. Statistical validation of mutual information calculations: Comparison of alternative numerical algorithms. Phys Rev E 2005; 71: 066208.

Fraser AM, Swinney HL. Independent coordinates for strange attractors from mutual information. Phys Rev A 1986; 33:1134-1140.

Gabor D. Theory of communication. J. IEE (London) 1946; 93: 429-457.

Golub GH, van Loan CF. Matrix Computations. Baltimore: Johns Hopkins University Press, 1983.

Hlavácková-Schindlera K, Paluš M, Vejmelka M, Bhattacharya J. Causality detection based on information-theoretic approaches in time series analysis. Phys Rep 2007; 441: 1-46.

Kantz H, Schreiber T. Nonlinear time series analysis. Cambridge UK: Cambridge University Press, 1997.

Lempel A, Ziv J. On the complexity of finite sequences. IEEE Trans. Information Theory 1976; 22: 75-81.

Mars NJI, Lopes da Silva FH In: Gevins AS, Remond A. eds. Methods of Analysis of Brain Electrical and Magnetic Signals: EEG Handbook. Amsterdam: Elsevier Science, 1987.

Nakamura T, Small M. Testing for dynamics in the irregular fluctuations of financial data. Physica A 2006; 366: 377-3c86.

Press WH, Teukolsky SA, Vetterling WT, Flannery BP. Numerical Recipes in C++. New York: Cambridge University Press, 2002.

Pritchard D, Theiler J. Generating surrogate data for time series with several simultaneously measured variables. Phys Rev Lett 1994; 73: 951-954.

Rapp PE, Schmah, T. Complexity Measures in Molecular Psychiatry. Molec Psych 1996; 1: 408-416.

Rapp PE, Cellucci CJ, Korslund KE, Watanabe TAA, Jimenez-Montaño MA. An effective normalization of complexity measurements for epoch length and sample frequency. Phys Rev E 2001; 64: 016209

Rapp PE, Cellucci CJ, Watanabe TAA, Albano AM, Schmah TI. Surrogate data pathologies and the false-positive rejection of the null hypothesis. Int J Bifurcations Chaos 2001; 11: 983-997.

Rapp PE, Cellucci CJ, Watanabe TAA, Albano AM. Quantitative characterization of the complexity of multichannel human EEGs. Int J Bifurcations and Chaos 2005; 15: 1737-1744.

Schreiber T, Schmitz A. Improved surrogate data for nonlinearity tests. Phys Rev Lett 1996; 77:635-638.

Schreiber T. Measuring Information Transfer. Phys Rev Lett 2000; 85: 461-464.

Shannon C, Weaver W. The Mathematical Theory of Communication, Urbana, IL: University of Illinois Press, 1949

Stansberry KB. Shapiro SA, Hill MA, McNitt PM, Meyer MD, Vinik AI. Impaired peripheral vasomotion in diabetes. Diabetes Care 1996; 19: 715-721.

Theiler J, Eubank S, Longtin A, Galdrikian B, Farmer JD. Testing for nonlinearity in time series: The method of surrogate data. Physica D 1992 ; 58: 77-94.

Theiler J, and Rapp PE. Re-examination of evidence for low-dimensional nonlinear structure in the human electroencephalogram. Electroencephalography and Clinical Neurophysiol 1996; 98: 213-222.

Theiler J, Prichard D. In Cutler CD, Kaplan DT, eds. Nonlinear Dynamics and Time Series: Building a Bridge between the Natural and Statistical Sciences. Fields Institute Communications 1997; 11: 99-.

Tigno XT, Selaru IK, Angeloni SV, Hansen BC. Is microvascular flow rate related to ghrelin, leptin and adiponectin levels?. Clin Hemorheol Microcirc 2003; 29: 409-416.

Tigno XT, Ding SY, Hansen BC. Paradoxical increase in dermal microvascular flow in pre-diabetes associated with elevated levels of CRP. Clin Hemorheol Microcirc 2006; 34: 273-282 .

Tigno XT, Hansen BC, Albano AM. Vasomotion spectra and principal component of pooled measures predict diabetes in monkeys. Int. J. Bifurcation and Chaos 2009; to appear.

Watanabe TAA, Cellucci, CJ, Koheygi E, Bashore TR, Josiassen RC, Greenbaun NN, Rapp PE. The algorithmic complexity of multichannel EEG is sensitive to changes in behavior. Psychophysiology 2003; 40 77-97.