International Journal of Bifurcation and Chaos, Vol. 12, No. 6 (2002) 1273–1293

NONLINEAR SIGNAL CLASSIFICATION

P. E. RAPP*,† and T. A. A. WATANABE†
Department of Pharmacology and Physiology,
Medical College of Pennsylvania–Hahnemann University, PA, USA
**Paul.E.Rapp@drexel.edu*

P. FAURE
Institut Pasteur, Paris

C. J. CELLUCCI†
Department of Physics, Ursinus College, Collegeville, PA, USA
† *The Arthur P. Noyes Research Foundation, Norristown, PA, USA*

Received July 26, 2001; Revised August 28, 2001

In this contribution, we show that the incorporation of nonlinear dynamical measures into a multivariate discrimination provides a signal classification system that is robust to additive noise. The signal library was composed of nine groups of signals. Four groups were generated computationally from deterministic systems (van der Pol, Lorenz, Rössler and Hénon). Four groups were generated computationally from different stochastic systems. The ninth group contained inter-decay interval sequences from radioactive cobalt. Two classification criteria (minimum Mahalanobis distance and maximum Bayesian likelihood) were tested. In the absence of additive noise, no errors occurred in a within-library classification. Normally distributed random numbers were added to produce signal to noise ratios of 10, 5 and 0 dB. When the minimum Mahalanobis distance was used as the classification criterion, the corresponding error rates were 2.2%, 4.4% and 20% (Expected Error Rate = 89%). When Bayesian maximum likelihood was the criterion, the error rates were 1.1%, 4.4% and 21% respectively. Using nonlinear measures an effective discrimination can be achieved in cases where spectral measures are known to fail.

Most classification errors occurred at low signal to noise ratios when a stochastic signal was misclassified into a different group of stochastic signals. When the within-library classification exercise is limited to the four groups of deterministic signals, no classification errors occurred with clean data, at SNR = 10 dB, or at SNR = 5 dB. A single classification error (Observed Error Rate = 2.5%, Expected Error Rate = 75%) occurred with both classification criteria at SNR = 0 dB.

Keywords: Discriminant analysis; nonlinear dynamics; complexity; fractal dimension; time series analysis; chaos.

*Author for correspondence.
Clinical Research Center, Building 52, Norristown State Hospital, 1001 Sterigere St. Norristown, PA 19401, USA.

1. Introduction

The classification of a signal into a library composed of several distinct groups is a classical problem in the signal processing literature. The objective can be stated simply: Given a test signal, we want to know if it is likely to be a member of previously characterized groups that comprise the library. Also, we want to know the confidence of this assignment. This problem presents itself in most areas of engineering practice. This problem is also encountered in the analysis of biological data, particularly behavioral data, and in clinical applications. For example, in the course of investigations of animal behavior one often wants to characterize the degree of similarity in the behavior of a specific animal against previously observed control data and against data obtained after the administration of drugs. Clinically, the classification problem is encountered during diagnosis. Given a patient's ECG or EEG we may want to know the probability that the signal was obtained from healthy, age-matched control subjects and the probability that the signal is a member of a library of well characterized clinical signals. In this contribution we use computationally generated signals and radioactive decay data to demonstrate that the incorporation of nonlinear dynamical measures into a multivariate discrimination can provide a signal classification system that is robust against additive noise to a signal to noise ratio of 0 dB.

In Sec. 2 the procedures used to generate the signal library are described. This is followed by a description in Sec. 3 of the five dynamical measures used in this study. The statistical measures used to construct the classification system are then described. The results are presented in Sec. 5, followed by a discussion (Sec. 6) which outlines the limitations of this investigation. The methods presented here generalize to treat other classes of problems commonly encountered in applied dynamics. These generalizations are also discussed.

2. Signal Library

Nine groups of signals were incorporated into this study:

Group 1: the periodic van der Pol equation
Group 2: the Lorenz equations
Group 3: the Rössler equations
Group 4: the Hénon system
Group 5: uniformly distributed random numbers
Group 6: filtered uniformly distributed random numbers
Group 7: normally distributed random numbers

Group 8: filtered normally distributed random numbers

Group 9: the time intervals between successive beta emissions from Co^{60}.

In the case of van der Pol, Lorenz, Rössler and Hénon data, the equations were iterated for ten thousand time steps from the initial values before data were recorded. The van der Pol, Lorenz, and Rössler equations were integrated using a sixth-order Runge–Kutta–Hutta algorithm [Lambert, 1973]. The governing equations for the van der Pol system were:

$$\frac{dx}{dt} = y + 2\left[x - \frac{x^3}{3}\right]$$

$$\frac{dy}{dt} = -x$$

the sampling interval was $h = 0.05$. The governing equations for the Lorenz system were:

$$\frac{dx}{dt} = -\sigma(x - y)$$

$$\frac{dy}{dt} = -xz + rx - y$$

$$\frac{dz}{dt} = xy - bz$$

where $\sigma = 10$, $b = 8/3$ and $r = 28$. The sampling interval was $h = 0.01$. The governing equations for the Rössler equations were:

$$\frac{dx}{dt} = -y - z$$

$$\frac{dy}{dt} = x + ay$$

$$\frac{dz}{dt} = b + xz - cz$$

where $a = 0.2$, $b = 0.4$ and $c = 5.7$. The sampling interval was $h = 0.10$.

The Hénon data sets were produced by iterating the Hénon map, $a = 1.4$ and $b = 0.3$.

$$x_{n+1} = 1 - ax_n^2 + y_n$$

$$y_{n+1} = bx_n$$

The sets of uniformly distributed random numbers were generated using a Park–Miller random number generator that incorporated a Bays–Durham shuffle [Park & Miller, 1988; Press *et al.*, 1992]. Data sets of uniformly distributed random numbers were filtered to produce an additional group. The Fourier transform of 8,192 uniformly distributed random numbers was calculated. The coefficient of the jth harmonic was multiplied by:

$$F_j = \max[0,\, 1 - kj^2] \quad k = 0.37 \times 10^{-6}$$

The corresponding inverse transform was used in subsequent calculations. Group 7 consists of normally distributed random numbers that were produced by transforming random data generated by the Park–Miller random number generator to a normal distribution [Press *et al.*, 1992]. The mean of each data set is approximately zero and the standard deviation is approximately one. An additional group of data sets (Group 8) was produced by filtering the normally distributed random numbers of Group 7. The calculations began with 16,384 normally distributed random numbers. The Fourier transform of this time series was calculated. The coefficients of the jth harmonic were multiplied by one for $j = 1, \ldots, 192$ and by zero for $j = 193, \ldots, 8{,}192$. The inverse transform was then calculated giving a time series of 16,384 elements. In order to minimize the end effects produced by using a step function filter, a final filtered data set was constructed by extracting the middle 8,192 elements of the inverse transform. These data sets of filtered random numbers differ from the previous sets (Group 6) in two ways. First, the input to the filter is uniformly distributed in the case of Group 6 and normally distributed in the case of Group 8. Additionally, the filter in Group 8 is far more severe. In the case of Group 6, the first 1,643 coefficients of 4,096 harmonics made a nonzero contribution to the inverse transform. In the case of Group 8, only 192 coefficients of 8,192 were retained. The final group of time series, Group 9, consists of the successive inter-decay intervals recorded from Co^{60}.

In future investigations, the classification system constructed here will be used with biological signals. An assessment of the system's robustness to noise is therefore essential. A distinction should be made between different types of noise. Three classes could be considered. Additive or observational noise is introduced by simply adding random numbers after the entire signal has been computed. Dynamic noise is introduced by adding random numbers to the dependent variables after each iteration of the computation. That is, suppose

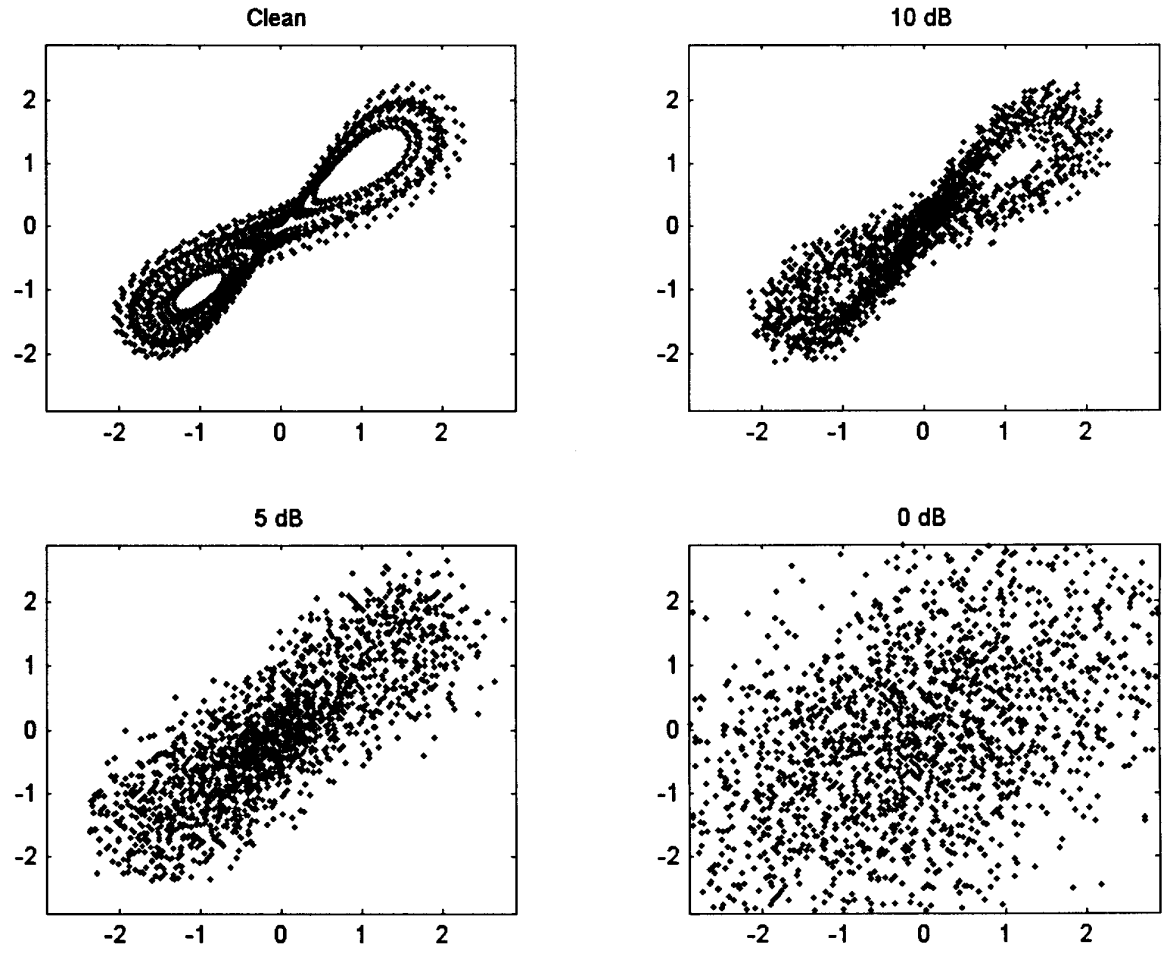

Fig. 1. The effect of additive noise on the structure of deterministic signals. Two-dimensional embeddings of Lorenz data (x_i, x_{i+5}) at four different noise levels are shown: clean data, SNR = 10 dB, SNR = 5 dB, and SNR = 0.

that x_i, y_i and z_i, the values of x, y and z at the ith time step, have been determined. In the case of dynamic noise, a random number is added to each of these values before computing the system's position at the next time point. In contrast, parametric noise is introduced into a system by replacing a parameter in the governing equations, for example, parameter σ in the Lorenz equations, with a time-dependent stochastic term $\sigma+\varepsilon_i$, where $\{\varepsilon_i\}$ is a set of random numbers satisfying some specification. In the computations presented here, we consider only observational noise, and in doing so we explicitly recognize that this form of noise is the least destructive to signal integrity. In order to test the classification system's robustness to additive noise, each data set was used to produce three additional time series. Normally distributed random numbers (mean = 0) were added to produce signals with a signal to noise ratio of 10 dB, 5 dB and 0 dB. The impact of additive noise can be assessed subjectively by examining the content of Fig. 1. At SNR = 0 dB, the signal variance is equal to the noise variance and the visually discernable geometrical structure of the data is lost.

3. Dynamical Measures

3.1. *Characteristic fractal dimension*

We begin the discussion of dynamical measures with a general observation. The number of measures that can be applied is enormous. As a matter of computational practicality, choices must be made. The choice of measures is not arbitrary. An unnecessary failure to classify data can result from inappropriate choices. The selection should reflect the known properties of the data and the objectives of the classification process. The results of the present study will be used in the analysis of animal motor behavior. This anticipated subsequent application

influenced the choice of measures used here. The value of any given measure in the classification process can be assessed quantitatively. This assessment procedure is described in the discussion section.

Five dynamical measures were used in this study: the characteristic fractal dimension, the Richardson dimension, Lempel–Ziv complexity, the Hurst exponent and the relative dispersion. The characteristic fractal dimension, CFD [Katz & George, 1985], measures the degree to which a trajectory in an arbitrary dimension departs from a straight-line path. The CFD is greater than or equal to one and has the value one in the case of a straight line. Let $\{p_i,\, i = 1, \ldots, N_{\text{Data}}\}$ be a trajectory, that is an ordered sequence of points in $\Re^m$, where $m \geq 2$. In the case of single channel data, $\{x_i,\, i = 1, \ldots,\, N_{\text{Data}}\}$, points are created in $\Re^m$ by embedding, $p_j = (x_j,\, x_{j+L}, \ldots,\, x_{j+(m-1)L})$, where $L \in I^+$ is the lag. The question of optimal choices for m and L is addressed elsewhere [Cellucci *et al.*, 2002].

Let $d_j = |p_{j+1} - p_j|,\ j = 1, \ldots,\, N_{\text{Data}} - 1$, where the Euclidean distance is used. (Note: Other distance metrics can be used. The Euclidean metric, however, carries a significant advantage. Euclidean distances are invariant under an orthogonal transformation. This is helpful when the orthogonal transformation generated by the singular value decomposition is used to reduce noise.)

Let

$$D_{\text{sum}} = \sum_{j=1}^{N_{\text{Data}}-1} d_j$$

and let

$$D_{\max} = \max |p_i - p_j|, \quad \text{for all } i,\, j \text{ pairs}$$

$D_{\max}$ is found by considering all possible i, j pairs, not just sequential, i to $i + 1$ pairs. Then

$$\text{CFD} = \frac{\log D_{\text{sum}}}{\log D_{\max}}$$

where, since it is expressed as a ratio, the base of the logarithm is immaterial.

It is necessary to specify the procedure used to apply this measure to the data sets in the test library. In the case of the van der Pol data (Group 1), 1024 two-dimensional points, $(x_i,\, y_i)$, were used in the calculations. The same procedure was used for the Lorenz data (Group 2), the Rössler data (Group 3), and the Hénon data (Group 4). The z values of the Lorenz and Rössler trajectories were not used. In the case filtered uniformly distributed random numbers (Group 6), and the filtered normally distributed random numbers (Group 8), which are univariate data sets, 1024 two-dimensional points of the form $(x_i,\, x_{i+1})$ were used in the calculations. For uniformly distributed random numbers (Group 5), normally distributed random numbers (Group 7) and cobalt disintegrations (Group 9), 1024 two-dimensional points of the form $(x_i,\, x_{i+1024})$ were used in the calculations. A different procedure is used for filtered and unfiltered random numbers because the filtered random numbers have a nonzero autocorrelation time.

3.2. *Richardson dimension*

As in the case of the characteristic fractal dimension, the Richardson dimension, D_R [Richardson, 1961; Mandelbrot, 1983; Çambel, 1993], is a measure of the degree to which a trajectory departs from a straight-line path. Additionally, it provides a quantitative characterization of how the estimate of the length of a curve changes with the precision of the measurement. This is the classical coastline of Britain problem [Mandelbrot, 1967]. D_R, like the characteristic fractal dimension, is greater than or equal to one, and has the value $D_R = 1$ in the case of a straight line.

As before, let $\{p_i,\, i = 1, \ldots,\, N_{\text{Data}}\}$ be a trajectory in $\Re^m$. Let τ denote a step length, $\tau \in I^+$. Let

$$d(\tau)_j = |p_{(j-1)\tau+1} - p_{j\tau+1}|, \quad j = 1, \ldots,\, J_{\max}$$

$$J_{\max} = \text{Int}\left(\frac{N_{\text{Data}}}{\tau}\right) - 1$$

where $\text{Int}(z)$ is the least integer greater than or equal to z. $d(\tau)_j$ is the distance between two consecutive points measured in step length τ. Let

$$D_{\text{sum}}(\tau) = \sum_{j=1}^{J_{\max}} d(\tau)_j$$

$D_{\text{sum}}(\tau)$ is the length of the curve measured in step lengths of τ. As τ gets bigger, $D_{\text{sum}}(\tau)$ decreases. Thus, if $D_{\text{sum}}(\tau)$ is plotted as a function of τ, the curve will be monotone decreasing. For many processes, it can be shown that the decreasing function of $D_{\text{sum}}(\tau)$ versus has a straight line segment called the scaling region [Mandelbrot, 1967]. Let S

denote the slope of $\log D_{\text{sum}}(\tau)$ versus $\log \tau$ in the scaling region. The Richardson dimension is defined by:

$$D_R = |S| + 1$$

where $|S|$ is the absolute value.

A complication is encountered when estimating D_R from finite, noisy data sets; namely, the lower and upper bound of the scaling region must be determined. It is possible to construct numerical procedures that search for the scaling region by finding the largest range of τ in which the derivative of $\log D_{\text{sum}}(\tau)$ versus $\log \tau$ is constant to some specified tolerance [Rapp *et al.*, 2001]. In the calculations presented here, a less demanding procedure was implemented. $D_{\text{sum}}(\tau)$ was calculated for $\tau = 1, \ldots, 10$ and the slope of the best fit straight line was used to estimate D_R. It is recognized that the resulting value is only a very approximate estimate of the true dimension. However, the procedure results in a numerically robust measure that has been shown to have empirical value in classifying dynamical systems. The embedding procedures used when calculating the characteristic fractal dimension were also used when calculating the Richardson dimension. That is, points were created in two-dimensional space by using the x and y coordinates of multivariate data (van der Pol, Lorenz, Rössler and Hénon) or by embedding univariate data in two dimensions.

3.3. *Lempel–Ziv complexity*

The term complexity has come to have a wide range of meaning in both general language and in the technical literature. Our use of the term is restricted to the quantitative examination of structure in symbol sequences. Specifically, a complexity measure is a sequence-sensitive measure of a symbol string that gives an indirect measure of the structure of the dynamical system that generated the message or provides a measure of the degree to which the observed symbol string differs from a random sequence that has the same symbol distribution. By sequence-sensitive, we mean a measure that, in contrast with distribution-determined measures, changes when the order of the symbols is changed. Several different definitions of complexity that satisfy this definition have been constructed (reviewed in [Rapp & Schmah, 1996]). It is possible to construct a taxonomy of complexity measures based on their mathematical properties [Rapp & Schmah, 2000]. In this presentation, we consider only one of the many possibilities, the Lempel–Ziv measure of complexity [Lempel & Ziv, 1976; Ziv & Lempel, 1978] which according to our taxonomic criterion is a nonprobabilistic, model-based, randomness-finding measure.

Two processes must precede the application of a complexity measure to dynamical data. First, in the case of a multivariate record, the signal must be expressed in a single channel. Second, the univariate time series of dynamical data must be reexpressed as a symbol sequence. There are several possible procedures for accomplishing each process. The simplest possibilities are used here. In the case of the van der Pol data (Group 1), the x and y values form two-dimensional points, (x_i, y_i). A time series d_j, $j = 1, \ldots, 1024$ was constructed by calculating the distance between consecutive points:

$$d_j = |(x_{j+1}, y_{j+1}) - (x_j, y_j)|$$

where the Euclidean distance is used. The time series $\{d_j\}$ was partitioned into a binary symbol sequence about its median. The complexity of this symbol sequence was then calculated. The same procedure was used for the Lorenz data (Group 2), the Rössler data (Group 3) and the Hénon data (Group 4). The z values of the Lorenz and Rössler trajectories were not used. In the case of filtered uniformly distributed random numbers (Group 6), and filtered normally distributed random numbers (Group 8), a 1024 point time series $\{d_j\}$ was calculated from the trajectory of two-dimensional points of the form (x_j, x_{j+1}). As before, the complexity was calculated after $\{d_j\}$ was partitioned into a binary symbol sequence about its median. For uniformly distributed random numbers (Group 5), normally distributed random numbers (Group 7), and cobalt disintegrations (Group 9), d_j, $j = 1, \ldots, 1024$ was calculated from successive points of the form (x_i, x_{i+1024}).

Once the original data have been represented by a symbol sequence, it is necessary to construct a quantitative measure that can characterize the complexity of this sequence. As stated previously, several candidate measures are available. In this study the Lempel–Ziv measure of complexity, C_{LZ}, has been used. A specification of C_{LZ} is preceded by the presentation of necessary definitions and notation.

Let message M be a sequence of symbols of length L_M.

$$M = \lambda_1 \lambda_2 \cdots \lambda_{L_M}$$

If, for example, $M = abacedge$, then $L_M = 8$. The index of a symbol in a message is the positive integer indicating its position. The index of symbol "c" in the preceding example is four. The vocabulary of a symbol sequence, denoted $v(M)$, is the set of its subsequences. By definition, a symbol sequence is an element of its own vocabulary. If $M = abcc$, then:

$$v(M) = \{a, b, c, ab, bc, cc, abc, bcc, abcc\}$$

A symbol string X is said to be an element of M, $X \in M$, if X is an element of the vocabulary of M, $X \in v(M)$. The concatenation of string X and Y, denoted by $X \oplus Y$, is the symbol sequence formed by appending sequence Y to X. Let Y be a symbol sequence of length L_Y. Y_π is the symbol sequence of length $L_Y - 1$ formed by deleting the last symbol from sequence Y. The notation ΣX_π denotes

$$\Sigma X_\pi = (X_1 \oplus X_2 \oplus \cdots X_J)_\pi$$

The Lempel–Ziv algorithm restates the original symbol sequence M by the concatenation of a uniquely defined series of subsequences.

$$M = X_1 \oplus X_2 \oplus \cdots X_k \oplus X_{k+1} \cdots \oplus X_N$$

The complexity, C_{LZ}, is the positive integer N which is equal to the number of subsequences required by this process. Thus, a crucial element in understanding the algorithm is an understanding of the rule used to specify the subsequences.

The Lempel–Ziv algorithm is initialized with X_1 consisting of the first symbol of M only. The length of X_1 is always equal to one. Successive subsequences, X_J, are constructed by following the previous subsequence and sequentially adding symbols to until

$$X_J \notin v\{(X_1 \oplus X_2 \oplus \cdots X_J)_\pi\}$$

at which point the generation of X_J is terminated and the construction of X_{J+1} begins. The process is most effectively presented by considering a specific example [Lempel & Ziv, 1976].

$M = 0\ \ 0\ \ 0\ \ 1\ \ 1\ \ 0\ \ 1\ \ 0$

$X_1 = 0$	
$X_2 = 0$	$X_2 \in (X_1 X_2)_\pi$
$X_2 = 00$	$X_2 \in (X_1 X_2)_\pi$
$X_2 = 001$	$X_2 \notin (X_1 X_2)_\pi$
$X_3 = 1$	$X_3 \in (X_1 X_2 X_3)_\pi$
$X_3 = 10$	$X_3 \notin (X_1 X_2 X_3)_\pi$
$X_4 = 1$	$X_4 \in (X_1 X_2 X_3 X_4)_\pi$
$X_4 = 10$	$\Sigma_J X = M$ Decomposition Complete
	$M = X_1 \oplus X_2 \oplus X_3 \oplus X_4$
	$M = 0 \oplus 001 \oplus 10 \oplus 10$

Under this definition, the complexity of symbol sequence M is 4. An important additional point should be made concerning the algorithm. By construction, the symbol sequence consisting of the first $L_J - 1$ elements of subsequence X_J is an element of the vocabulary $v(\Sigma X_\pi)$. This means that there is an index I_J such that the first $L_J - 1$ elements of X_J can be reproduced exactly by copying symbols $\lambda_{I_J} \cdots \lambda_{I_J - (L_J - 1)}$. Therefore, three values can determine X_J exactly. They are (i) L_J, its length, (ii) I_J, the starting position *earlier* in the message of the first $L_J - 1$ elements, and (iii) S_J, the identity of the last symbol of subsequence X_J. This is true irrespective of the length of X_J. Substring X_J could, in the case of a very long message, be ten thousand elements long. Nonetheless, it is fully specified by these three elements. This is the basis of Lempel–Ziv data compression algorithms.

3.4. *Hurst exponent*

The Hurst exponent [Hurst *et al.*, 1965; Çambel, 1993; Bassingthwaighte *et al.*, 1994; Elliott *et al.*, 1995; Kikkinides & Burganos, 1999] provides a quantitative measure of persistence (like is followed by like: an increase is followed by an increase, a decrease is followed by a decrease) and anti-persistence (an increase is followed by a decrease, a decrease is followed by an increase). Persistence is indicated by a value of $H > 1/2$, and anti-persistence is indicated by a value $H < 1/2$. Let $\{x_1, x_2, \ldots, X_{N_{\text{Data}}}\}$ denote a time series in $\Re^1$. Positive integer τ is specified. (In our implementation, τ is varied from $\tau = 10$ to $\tau = N_{\text{Data}}/2$ in steps of one.) The time series is partitioned into sequential, nonoverlapping subepochs

containing τ elements. The Kth subepoch is, therefore, $\{x_{(K-1)\tau+1}, \ldots, x_{K\tau}\}$.

Let $\{x'_1, x'_2, \ldots, x'_\tau\}$ denote the elements of the Kth subepoch following mean normalization within the subepoch. Let $\{S_1, S_2, \ldots, S_\tau\}$ denote the sequence generated by calculating the cumulative sum in the Kth subepoch following mean normalization.

$$S_j = \sum_{i=1}^{j} x'_i$$

Let $R(\tau)_K$ denote the range of the cumulative sum for the Kth subepoch specified by this value of τ, that is,

$$R(\tau)_K = \max |S_i - S_j|,\ i, j = 1, \ldots, \tau$$

and let σ_K denote the standard deviation of $\{x'_1, x'_2, \ldots, x'_\tau\}$ in subepoch K.

$\langle R^*(\tau)\rangle$ is the average of $R(\tau)_K/\sigma_K$ over all subepochs of length τ. It is the average value of the range of the cumulative sum normalized against the variability of the signal. For many dynamical systems, the function $\log\langle R^*(\tau)\rangle$ versus $\log \tau$ has a linear scaling region. The slope of that line is the Hurst coefficient. As previously noted, τ is varied from $\tau = 10$ to $\tau = N_{\text{Data}}/2$ in steps of one in our implementation. The estimated value of H is the slope of the best-fit straight line.

The procedures used in the complexity calculations to express multivariate data as a single channel time series were used in the calculations of the Hurst exponent. In the case of van der Pol data, Lorenz data, Rössler data and Hénon data, consecutive (x, y) points were used to calculate the distance function $\{d_j\}$. In the case of filtered uniformly distributed random numbers and filtered normally distributed random numbers, $\{d_j\}$ is calculated from two-dimensional points (x_j, x_{j+1}). For uniformly distributed random numbers, normally distributed random numbers and cobalt disintegration data, $\{d_j\}$ is calculated from (x_j, x_{j+1024}). However, in contrast with complexity calculations, the Hurst exponent is calculated with these time series of real numbers and not with their symbolic reduction.

3.5. *Relative dispersion*

The relative dispersion [Klocke *et al.*, 1995; Pedersen *et al.*, 1996; Willis *et al.*, 1997] quantifies the dependency of variance on the length of the data set. It is expressed as a dimension, D_{RD}, and in the case of a straight line $D_{RD} = 1$. Let $\{x_1, x_2, \ldots, x_{N_{\text{Data}}}\}$ denote the time series. In an iterative process, the time series is partitioned into sequential, nonoverlapping subepochs of length 2^0, $2^1, 2^2, \ldots, 2^{n_{\max}}$ where $n_{\max}$ is the greatest positive integer such that $2^{n_{\max}} \leq N_{\text{Data}}/4$.

The following computation is performed in the Kth subepoch of length τ. Let $\langle x(\tau)\rangle_K$ denote the mean of $\{x\}$ in the Kth subepoch. Let $\sigma(\tau)$ denote the standard deviation of the means obtained in subepochs of length τ. That is, $\sigma(\tau)$ is the standard deviation of the set $\{\langle x(\tau)\rangle_1, \langle x(\tau)\rangle_2, \ldots\}$. $\sigma'(\tau)$ is a normalization of $\sigma(\tau)$ defined by:

$$\sigma'(\tau) = \frac{\sigma(\tau)\{(\text{number of subepochs of length } \tau) - 1\}}{\{\text{number of subepochs of length } \tau\}}$$

Let $\langle x\rangle$ denote the mean of the entire time series. The relative dispersion, $RD(\tau)$, is defined by

$$RD(\tau) = \frac{\sigma'(\tau)}{\langle x\rangle}$$

The function $\log RD(\tau)$ versus $\log \tau$ is a decreasing function. Let S denote the slope of the linear scaling region. Then,

$$D_{RD} = 1 + |S|$$

The data reduction protocols used to reduce multivariate data to a single channel in the calculation of the Hurst exponent are also used in the calculation of the relative dispersion.

4. Statistical Procedures

4.1. *Calculation of* P_{SAME}

Four statistical procedures have been employed in this investigation. The first two, calculation of $P_{\text{SAME}}(G_A, G_B)$ and $P_{\text{ERROR}}(G_A, G_B)$, provide a global characterization of the signal library and the discriminatory capacity of the dynamical measures. The third and fourth procedures, calculation of $P_{\text{ABS}}(\underline{x}_{\text{Test}}|G_A)$ and $P_{\text{BAYES}}(\underline{x}_{\text{Test}}|G_A)$, provide two methods for classifying signals within the library. Thus the methods presented here only consider the problem of classification of a test signal amongst previously specified groups. A distinction should be drawn between classification and cluster analysis. Cluster analysis [Aldenderfer & Blashfield,

1984] considers a related problem: given a large collection of objects (in this case, the objects are signals) from which several measurements have been obtained, is it possible to identify distinct groups within this collection? The large literature on cluster analysis has been summarized by a National Research Council Commission [Panel on Discriminant Analysis, Classification and Clustering, 1988].

The first statistical calculation considered here examines two groups and addresses the question, what is the probability that the two groups are the same? Let us suppose that the library consists of K groups and that there are N_A, $N_B, \ldots N_K$ signals in each group. Z dynamical measures are calculated for each signal in each group. The motivating question can now be stated with greater precision. Given these signals and these dynamical measures, what is the probability that two groups in the library, denoted Group A and Group B, are the same? This probability is denoted by $P_{\text{SAME}}(G_A, G_B)$. The presentation here follows the development in [Lachenbruch, 1975, p. 10], [Flury & Riedwyl, 1988, p. 97], and the report of the National Research Council Panel on Discriminant Analysis and Clustering [1988, p. 13]. $\underline{\hat{\mu}}_A$ is the Z-dimensional vector containing the average values of the discriminating variables using members of Group A.

$$\underline{\hat{\mu}}_A = (\hat{\mu}_{A1}, \hat{\mu}_{A2}, \ldots, \hat{\mu}_{AZ})$$

$$\hat{\mu}_{Ai} = \frac{1}{N_A} \sum_{m=1}^{N_A} x_i(m)$$

where $x_i(m)$ is the mth value of discriminating variable i in Group A. $(\sigma_A^2)_{i,j}$ is element (i, j) of the Group A covariance matrix.

$$(\sigma_A^2)_{i,j} = \frac{1}{N_A - 1} \sum_{m=1}^{N_A} (x_i(m) - \hat{\mu}_{Ai})(x_j(m) - \hat{\mu}_{Aj})$$

Σ_A denotes the $Z \times Z$ matrix of elements $(\sigma_A^2)_{i,j}$. Σ_A^{-1} denotes its inverse. These variables are defined analogously for Group B. $(\sigma_{A,B}^2)_{i,j}$ is element (i, j) of the between-Group covariance matrix for Groups A and B.

$$(\sigma_{A,B}^2)_{i,j} = \frac{(N_A - 1)(\sigma_A^2)_{i,j} + (N_B - 1)(\sigma_B^2)_{i,j}}{N_A + B_B - 2}$$

$\Sigma_{A,B}$ denotes the matrix formed by these elements, and $\Sigma_{A,B}^{-1}$ is its inverse. The between-group Mahalanobis distance, $D_{A,B}^2$, is defined by

$$D_{A,B}^2 = \begin{pmatrix} \hat{\mu}_{A1} - \hat{\mu}_{B1} \\ \hat{\mu}_{A2} - \hat{\mu}_{B2} \\ \vdots \\ \hat{\mu}_{AZ} - \hat{\mu}_{BZ} \end{pmatrix}^T \Sigma_{A,B}^{-1} \begin{pmatrix} \hat{\mu}_{A1} - \hat{\mu}_{B1} \\ \hat{\mu}_{A2} - \hat{\mu}_{B2} \\ \vdots \\ \hat{\mu}_{AZ} - \hat{\mu}_{BZ} \end{pmatrix}$$

The probability that the two groups are the same is given by an F-test.

$$P_{\text{SAME}}(G_A, G_B) = I_{\frac{\upsilon_2}{\upsilon_2 + \upsilon_1 F}} \left(\frac{\upsilon_2}{2}, \frac{\upsilon_1}{2} \right)$$

where $\upsilon_1 = Z$, the number of discriminating variables, is the number of degrees of freedom in the numerator, and $\upsilon_2 = N_A + N_B - Z - 1$, is the degrees of freedom in the denominator.

$$F = \frac{N_A N_B (N_A + N_B - Z - 1) D_{A,B}^2}{(N_A + N_B)(N_A + N_B - 2) Z}$$

$I_x(a, b)$ is the incomplete β function.

$$I_x(a, b) = \frac{1}{B(a, b)} \int_0^x t^{a-1}(1 - t)^{b-1} dt$$

and $B(a, b)$ is the beta function

$$B(a, b) = \int_0^1 t^{a-1}(1 - t)^{b-1} dt$$

which is seen to be monotone decreasing with the Mahalanobis distance. From the numerator of F, it is seen that $N_A + N_B > Z + 1$ is an absolute requirement of the analysis.

The sensitivity of $P_{\text{SAME}}(G_A, G_B)$ to the choice of dynamical measures should be explicitly recognized. The failure to reject the null hypothesis does not mean that it will invariably be impossible to distinguish between the two groups. This remains an open question because it is always possible that the introduction of additional measures will result in a statistically significant discrimination.

4.2. *Calculation of* P_{ERROR}

It is sometimes supposed that if P_{SAME} is small, then the dynamical measures provide an effective means of classifying the signals between groups. This is not the case. A small value of P_{SAME} is necessary for discrimination, but it is not sufficient. If the distributions of the two groups overlap, it is possible for P_{SAME} to be small even

though there is a high error rate when signals are classified. (A numerical example will be given presently.) A more immediately pertinent measure of discrimination is $P_{\mathrm{ERROR}}(G_A, G_B)$ which is the probability of an error in pairwise discriminations between the two groups. A precise definition of $P_{\mathrm{ERROR}}(G_A, G_B)$ must be expressed in terms of a minimum Mahalanobis distance classification criterion. Suppose a test signal is selected at random from either Group A or Group B. Let $\underline{x}_{\mathrm{Test}}$ be the Z-dimensional vector of dynamical measures obtained from the test case.

$$\underline{x}_{\mathrm{Test}} = (x_{1-\mathrm{Test}}, x_{2-\mathrm{Test}}, \ldots, x_{Z-\mathrm{Test}})$$

The Mahalanobis distance between $\underline{x}_{\mathrm{Test}}$ and Group A, denoted $D^2_{\mathrm{Test},A}$, is given by:

$$D^2_{\mathrm{Test},A} = \begin{pmatrix} x_{1-\mathrm{Test}} - \hat{\mu}_{A1} \\ x_{2-\mathrm{Test}} - \hat{\mu}_{A2} \\ \vdots \\ x_{Z-\mathrm{Test}} - \hat{\mu}_{AZ} \end{pmatrix}^T \Sigma_A^{-1} \begin{pmatrix} x_{1-\mathrm{Test}} - \hat{\mu}_{A1} \\ x_{2-\mathrm{Test}} - \hat{\mu}_{A2} \\ \vdots \\ x_{Z-\mathrm{Test}} - \hat{\mu}_{AZ} \end{pmatrix}$$

where $\underline{\hat{\mu}}_A$ and Σ_A are as previously defined. $D^2_{\mathrm{Test},B}$ is defined analogously for Group B. Under this criterion $\underline{x}_{\mathrm{Test}}$ will be deemed to be a member of the group corresponding to the smaller Mahalanobis distance. $P_{\mathrm{ERROR}}(G_A, G_B)$ is the overall error rate of this pairwise discrimination. Mahalanobis [1936] (see also [Lachenbruch, 1975]) has shown that P_{ERROR} can be approximated by

$$P_{\mathrm{ERROR}}(G_A, G_B) = 1 - \Phi\left(\frac{\sqrt{D^2_{A,B}}}{2}\right) = \Phi\left(-\frac{\sqrt{D^2_{A,B}}}{2}\right)$$

As in the case of P_{SAME}, the Mahalanobis distance is calculated using the between-group covariance matrix. Φ is the cumulative normal distribution.

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2} du = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{x}{\sqrt{2}}\right)\right]$$

where erf is the error function.

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$$

A numerical example of the distinction between P_{SAME} and P_{ERROR} is given in Fig. 2.

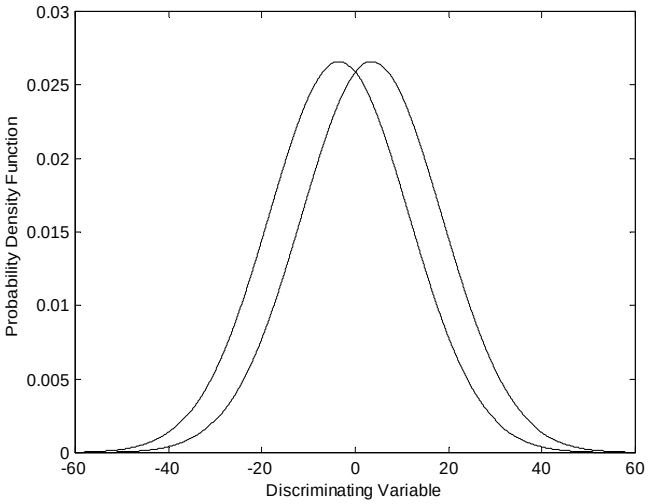

Fig. 2. Two normal distributions in which P_{SAME} is low ($P_{\mathrm{SAME}} = 3.2 \times 10^{-13}$) and P_{ERROR} is high ($P_{\mathrm{ERROR}} = 0.32$). $\hat{\mu}_A = -3.5$, $\hat{\mu}_B = 3.5$, $\sigma_A = \sigma_B = 15$, and $N_A = N_B = 500$.

Consider the case where a single discriminating variable ($Z = 1$) is used in a pairwise discrimination between Group A and Group B. The mean value of the discriminating variable in Group A is $\hat{\mu}_A$, and its variance is σ_A^2. $\hat{\mu}_B$ and σ_B^2 are defined analogously. For a one-dimensional discrimination, the between-group covariance is a single element.

$$\sigma^2_{A,B} = \frac{(N_A - 1)\sigma_A^2 + (N_B - 1)\sigma_B^2}{N_A + N_B - 2}$$

As shown in [Flury & Riedwyl, 1988], the corresponding Mahalanobis distance for the $Z = 1$ case is given by

$$D^2_{A,B} = \frac{(\hat{\mu}_A - \hat{\mu}_B)^2}{\sigma^2_{A,B}}$$

This expression for $D^2_{A,B}$ can be used in the previously presented expressions for P_{SAME} and P_{ERROR}. It is possible to produce cases in which P_{SAME} is small while P_{ERROR} approaches the maximum possible error rate in a pairwise discrimination of $P_{\mathrm{ERROR}} = 0.5$. In Fig. 2 $P_{\mathrm{SAME}} = 3.2 \times 10^{-13}$, but $P_{\mathrm{ERROR}} = 0.32$.

4.3. *Calculation of a priori membership probabilities*

The minimum Mahalanobis distance classification criterion presented in the definition of the

pairwise P_{ERROR} generalizes to an arbitrary number of groups. In a library of K groups, the test signal is classified into Group J if

$$D^2_{\text{Test,J}} = \min(D^2_{\text{Test,I}},\, I = 1,\, 2, \ldots,\, K)$$

However, it is also important to estimate the confidence of a classification. One way of doing this is to calculate the probability that $\underline{x}_{\text{Test}}$ is an element of the distribution formed by members of Group A. Informally, we refer to this as the probability that $\underline{x}_{\text{Test}}$ is a member of Group A. This probability is denoted by $P_{\text{ABS}}(\underline{x}_{\text{Test}}|G_A)$, where 'ABS' in P_{ABS} indicates that this is the absolute or *a priori* probability. A procedure for estimating P_{ABS} is given in [Flury & Riedwyl, 1988, p. 136]. The probability $P_{\text{ABS}}(\underline{x}_{\text{Test}}|G_A)$ is given by:

$$P_{\text{ABS}}(\underline{x}_{\text{Test}}|G_A) = I_{\frac{v_2}{v_2+v_1F}}\left(\frac{v_2}{2},\, \frac{v_1}{2}\right)$$

where $v_1 = Z$ is the number of discriminating variables. $v_2 = N_A - Z$, where N_A is the number of members in Group A and F is given by

$$F = \frac{N_A(N_A - Z)}{(N_A^2 - 1)Z} D^2_{\text{Test},A}$$

As before, $I_x(a,\, b)$ is the incomplete beta function, and $B(a,\, b)$ is the beta function. It is seen that this is equivalent to $P_{\text{SAME}}(G_A,\, G_B)$ for the special case where $N_B = 1$ and $\underline{\hat{\mu}}_B = \underline{x}_{\text{Test}}$.

4.4. *Calculation of Bayesian membership probabilities*

An alternative classification procedure can be derived from Bayes' theorem [McLachlan, 1992, p. 53]. The group-specific density estimate at $\underline{x}_{\text{Test}}$ from Group A is

$$f_A(\underline{x}_{\text{Test}}) = \frac{1}{(2\pi)^{Z/2}(\det \Sigma_A)^{1/2}} \exp\left\{-\frac{1}{2}\begin{pmatrix} x_{1-\text{Test}} - \hat{\mu}_{A1} \\ x_{2-\text{Test}} - \hat{\mu}_{A2} \\ \vdots \\ x_{Z-\text{Test}} - \hat{\mu}_{AZ} \end{pmatrix}^T \Sigma_A^{-1} \begin{pmatrix} x_{1-\text{Test}} - \hat{\mu}_{A1} \\ x_{2-\text{Test}} - \hat{\mu}_{A2} \\ \vdots \\ x_{Z-\text{Test}} - \hat{\mu}_{AZ} \end{pmatrix}\right\}$$

Using Bayes' theorem, the posterior probability of $\underline{x}_{\text{Test}}$ belonging to Group A is

$$P_{\text{BAYES}}(\underline{x}_{\text{Test}}|G_A) = \frac{P'_A f_A(\underline{x}_{\text{Test}})}{\sum_{M=1}^{K} P'_M f_M(\underline{x}_{\text{Test}})}$$

where P'_M is the prior probability of membership in Group M. $\underline{x}_{\text{Test}}$ is deemed to be a member of the group corresponding to the maximum value of P_{BAYES}.

When calculating $P_{\text{SAME}}(G_A, G_B)$ and $P_{\text{ERROR}}(G_A, G_B)$ we used the Mahalanobis distance calculated with $\Sigma_{A,B}$, the between-group covariance matrix. Also, P_{ABS} and P_{BAYES} were calculated with $D^2_{\text{Test},A}$ which was computed using Σ_A, the Group A covariance matrix. An alternative practice particularly favored in the earlier literature replaces $\Sigma_{A,B}$ and Σ_A with Σ_{Pool}, the pooled covariance matrix computed using all K groups

$$\sigma^2_{\text{Pool}} = \frac{(N_A - 1)\sigma_A^2 + (N_B - 1)\sigma_B^2 + \cdots + (N_K - 1)\sigma_K^2}{N_A + N_B + \cdots + N_K - K}$$

When compared against computations performed with Σ_{Pool}, we found a lower incidence of classification error with our test signals using Σ_A computed separately for each group. All of the results reported in the Appendix were obtained by separately computed $\Sigma_{A,B}$ and Σ_A matrices. An examination of the expressions for $P_{\text{ABS}}(\underline{x}_{\text{Test}}|G_A)$ and $P_{\text{BAYES}}(\underline{x}_{\text{Test}}|G_A)$ indicates that if Σ_A is replaced by Σ_{Pool}, then the same classifications are obtained with minimum Mahalanobis distance and with maximum Bayesian likelihood criteria. In the computations using our signal library, the results are usually the same. Operationally, we compute both classifications and use disagreements between the methods and the accompanying low membership probabilities as a warning of uncertainty in the classification.

5. Results

5.1. *Classification errors, including stochastic signals*

A detailed presentation of the results is given in the

Table 1. Within-Library Classification Errors — Nine Groups, including Stochastic Signals.

	Error rate expected	Error rate minimum Mahalanobis distance	Error rate maximum Bayesian likelihood
Clean Data	89%	0%	0%
SNR = 10 dB	89%	2.2%	1.1%
SNR = 5 dB	89%	4.4%	4.4%
SNR = 0 dB	89%	20%	21%

Appendix. A summary is provided here in which results obtained with each of the four noise levels (clean, SNR = 10 dB, SNR = 5 dB, SNR = 0 dB) are examined sequentially. Since there are nine groups in this study, the expected classification error rate based on random assignment is 89%. The observed within-library classification error rates at each noise level are presented in Table 1.

In the case of clean data there is a statistically significant separation between all groups. The average value of P_{SAME} is 0.174×10^{-6}. The largest single contribution is $P_{\mathrm{SAME}} = 0.6 \times 10^{-5}$. The probabilities of errors in pairwise discriminations were also low. The average value of P_{ERROR} is 0.155×10^{-3} and the largest single value is 0.6×10^{-2}. No classification errors were made using either minimum Mahalanobis distance or maximum Bayesian likelihood as the classifying criterion when clean time series were examined. Also, the confidence of the classification is high. In the case of correct classifications, P_{ABS} is 0.797 and the average value of P_{BAYES} is 0.999. Detailed results are given in the Appendix. As shown in A.3 and A.4, these measures were able to discriminate successfully between uniformly and normally distributed random numbers. This discrimination cannot be achieved with spectral measures.

The introduction of additive noise to SNR = 10 dB results in two classification errors when the minimum Mahalanobis distance is the classification criterion (Error Rate = 2.2%) and one classification error (Error Rate = 1.1%) when the maximum Bayesian likelihood is used to classify the signals. At this noise level, the average value of P_{SAME} is 0.718×10^{-4} with a maximum value of 0.3×10^{-2}. The average predicted error rate in pairwise discriminations is $\langle P_{\mathrm{ERROR}} \rangle = 0.263 \times 10^{-2}$ with a maximum value of 0.0742.

When the noise level is increased to SNR = 5 dB, the average value of P_{SAME} is 0.0006, with a maximum value of 0.0154. This maximum value is obtained when filtered, uniformly distributed random numbers are compared against filtered, normally distributed random numbers. One would expect this distinction to be obscured by additive normally distributed noise. The same pattern is observed when the values of P_{ERROR} are examined. At this noise level, $\langle P_{\mathrm{ERROR}} \rangle = 0.0130$, and again the maximum value ($P_{\mathrm{ERROR}} = 0.1229$) is obtained when filtered, uniformly distributed noise is compared with filtered, normally distributed noise. At SNR = 5 dB, there is a 4.4% classification error rate (again, we note that the expectation rate is 89%) when the minimum Mahalanobis distance criterion is used. The observed error rate is also 4.4% when the maximum Bayesian likelihood is used to classify signals.

When the variance of added noise is equal to the variance of the original signal (SNR = 0 dB), the classification error rate is approximately one quarter of the expectation rate of 89%. When the minimum Mahalanobis distance criterion is applied, the error rate is 20%. When the Bayesian maximum likelihood is the assignment criterion, the error rate is 21%. The average value of P_{SAME} is 0.0381, and the average value of P_{ERROR} is 0.0520. In operational terms, however, the system's performance is better than these numbers alone would suggest. A disagreement between a minimum Mahalanobis distance classification and a Bayesian likelihood classification provides a warning of an uncertain classification. Also, for any given test case, low values of P_{ABS} and P_{BAYES} provide an indication of reduced confidence. The average values of P_{ABS} and P_{BAYES} in the case of correct classifications are shown in Table 1. The case by case probabilities are given in the Appendix.

5.2. *Classification errors, deterministic signals only*

It should be noted that most classification errors occur when stochastic signals that have been obscured by high levels of additive noise are misclassified into an incorrect stochastic signal group. For example, at SNR = 0 dB, four errors (out of ten test cases) occur when a time series obtained by adding normally distributed noise to uniformly distributed random numbers is classified as belonging to the group consisting of normally distributed

Table 2. Within-Library Classification Errors — Four Groups, Deterministic Signals only.

	Error rate expected	Error rate minimum Mahalanobis distance	Error rate maximum Bayesian likelihood
Clean Data	75%	0%	0%
SNR = 10 dB	75%	0%	0%
SNR = 5 dB	75%	0%	0%
SNR = 0 dB	75%	2.5%	2.5%

random time series. This is hardly surprising since at SNR = 0 dB the test case was composed by adding equal elements of each signal category. At high noise levels a more valid measure of the system's robustness is obtained when only deterministic signals are considered. Table 2 shows the results obtained from a four-group study consisting of the van der Pol oscillator and the Lorenz, Rössler and Hénon attractors.

When the classification is limited to four groups of deterministic systems, no errors occurred with clean data, at SNR = 10 dB or at SNR = 5 dB. A single error occurred at SNR = 0 dB. The error rates at negative signal to noise ratios in studies that have more than ten cases in each group is deferred to a subsequent investigation.

6. Discussion

The results in Tables 1 and 2 suggest that the classification system shows a capacity to discriminate between dynamical systems, and that it is robust to additive noise. The present investigation is a preliminary study. Future research will address a number of issues. First the sensitivity to the number of elements in each group should be investigated systematically. The number of cases in a group determines the accuracy of the within-group average value estimates. These numbers appear in the equation specifying the Mahalanobis distance. Additionally, and perhaps more importantly, the number of cases in each group has an impact on the determination of the covariance matrix and its inverse. Limited calculations suggest that if the number of members in a group is low, the inverse of the within-group covariance matrix cannot be calculated meaningfully.

The introduction of additional dynamical measures should be investigated. The identification of appropriate measures is not a question with a single answer. Dynamical measures useful in one classification problem may be nondisclosing in another. For example, the measures that can successfully discriminate between classes of waveform data may not be particularly helpful when interevent interval data are to be classified. It should be recalled that by definition the performance obtained in the nine- group and four-group classification studies presented here is the worst case for this signal libraray. If the introduction of an additional measure does not improve performance, it does not have to be incorporated into the system. The question: "Is a given dynamical measure helpful?" can be resolved mathematically. Flury and Riedwyl [1988] described the method of backward elimination. The contribution of each dynamical measure can be determined by first calculating the partial F-value associated with each variable and then eliminating those variables with small partial F-values.

As indicated in the introduction, signal classification has been applied in medical diagnosis, for example the differential diagnosis of neuropathies and myopathies [Dengler *et al.*, 1986; Rapp *et al.*, 1993] and in clinical monitoring [Fan *et al.*, 1994; Storella *et al.*, 1996]. Psychiatric diagnosis based on EEGs and ERPs (the time-locked electroencephalographic response to punctate stimuli) has received significant effort and has generated a large literature. The work of John, Prichep and their colleagues is particularly noteworthy [Prichep *et al.*, 1993; John *et al.*, 1994a, 1994b; Hughes & John, 1999]. While some investigators [Donchin *et al.*, 1986] have suggested that the diagnostic utility of EEGs and ERPs is limited by the lack of disorder-specific signal abnormalities, others have argued that a sufficiently large data base and an appropriate choice of quantitative measures can result in a high degree of diagnostic specificity and sensitivity [Hughes & John, 1999]. It has been proposed [Rapp, 1999; Rapp & Schmah, 2000] that independently of the procedure's value in diagnosis, this strategy may have value as a within-patient longitudinal measure in which P_{Normal}, the probability that the observed signal is an element of the age and gender-matched healthy control group, is determined at regular intervals and used to follow the course of the disease and the response to treatment. Prediction of treatment outcome is a pressing problem in all areas of medicine and most particularly in psychiatry. Early classification between groups of treatment responders and nonresponders would

be valuable. Examples of outcome prediction based on electroencephalographic measures include the use of acute challenge ERPs as a prognostic of response/nonresponse to methylphenidate treatment of attention deficit-hyperactivity disorder [Young *et al.*, 1995] and the analysis of EEGs and the prediction of treatment outcome for cocaine dependent males [Prichep *et al.*, 1999].

Nonclinical applications of signal classification systems based on nonlinear measures of dynamical behavior to biological data are also being implemented. The system constructed and validated here will be used in the analysis of fish swimming behavior. Digitized trajectories of freely swimming fish will be obtained in uncontaminated control water. Additionally, trajectories obtained using water contaminated by known concentrations of specified toxins will be collected and analyzed. The operational question then becomes: given a test trajectory obtained in water of unknown purity, what is the probability that it was obtained in clean water and what are the probabilities that the trajectory was observed in water contaminated by toxins previously cataloged in the trajectory library.

The system presented here provides a dynamical and statistical methodology that can be applied to a broader range of problems. To an approximation, applied dynamical analysis addresses four problems: signal classification, detection of change, prediction and optimal control of dynamical behavior. This contribution considers only the first problem but generalizations to the remaining three problems can be constructed on this foundation. Detection of change is the central problem of smart alarm technology and is particularly important in medical applications. As suggested by Azen and Afifi [1972] (for additional comments see [Lachenbruch, 1975]) detection of change can be addressed by treating the Mahalanobis distance as a function of time. The distance of the signal's immediate past from preceding observations or from a standardized baseline can be calculated. For example, an attempt could be made to use this procedure to detect seizure onset and termination based on electroencephalographic and/or cardiovascular variables. The problem of detecting change is closely related to the problem of prediction. In some systems it may be possible to use the time-dependent Mahalanobis distance to predict changes in dynamical behavior. Prediction based on the time-dependent Mahalanobis distance has a significant advantage over methods based on a single dynamical measure, (the usual practice in the seizure prediction literature), because this method provides a systematic statistical procedure for optimally combining several complementary dynamical measures and for combining different classes of observed variables (for example, a combination of electroencephalographic and cardiovascular variables in a single discrimination).

The fourth application of dynamical analysis is system control. In an earlier contribution [Rapp *et al.*, 1988], we addressed the treatment of what Mackey and Glass [1977] have termed dynamical diseases. These are disorders that result from parameter-dependent bifurcations in physiological control systems. An intervention to a dynamical disease is constructed by altering accessible system parameters, usually with medication, until acceptable dynamical behavior is restored. A response is optimal if this restoration is achieved in a minimum time with a minimum intervention (the smallest pharmacological impact). In [Rapp *et al.*, 1988], it was suggested that the parameter dependence of a single dynamical measure (in that paper Hurwitz polynomials or Floquet coefficients were suggested) could be used as the objective function in an optimal control problem. It is now suggested that it could be possible to introduce multiple dynamical measures into this process by incorporating the Mahalanobis distance into the objective function.

Acknowledgments

The ideas implemented in this contribution were developed in discussions with Tanya Schmah. She provided essential intellectual leadership during the initial stages of this project. This work was sponsored in part by the Defense Advanced Projects Agency (DARPA) Defense Sciences Office under the auspices of Dr. Alan Rudolph through the Space and Naval Warfare Systems Center, San Diego Contract No. N66001-00-C-8012. We would like to acknowledge support from the U.S. Department of Education Award No. H235J000001 to the Krasnow Institute, from the Janssen Research Foundation, and from the Bristol-Myers Squibb Pharmaceutical Research Institute. We would also like to acknowledge the encouragement of R. C. Josiassen, Director and E. Kohegyi, Associate Director of the Arthur P. Noyes Research Foundation at Norristown State Hospital. We would like to acknowledge valuable discussions with D. Faber, H. Korn, T. Nakamura and H. Neumeister.

References

Aldenderfer, M. S. & Blashfield, R. K. [1984] *Cluster Analysis* (Sage Publications, Newbury Park, CA).

Azen, S. P. & Afifi, A. A. [1972] "Two models for assessing prognosis on the basis of successive observations," *Math. Biosci.* **14**, 169–176.

Bassignthwaighte, J. B., Liebovitch, L. S. & West, B. J. [1994] *Fractal Physiology* (Oxford University Press, NY).

Çambel, A. B. [1993] *Applied Chaos Theory. A Paradigm for Complexity* (Academic Press, San Diego, CA).

Cellucci, C. J., Albano, A. M., Rapp, P. E. & Krystal, A. D. [2002] "A comparative study of embedding methods," in preparation.

Dengler, R., Gillespie, J., Argenta, M., Elek, J., Wolf, W. & Struppler, A. [1989] "The impact of paired motor discharges on tremor," *Electromyogr. Clin. Neurophysiol.* **29**, 113–117.

Donchin, E., Miller, G. A. & Farwell, L. A. [1986] "The endogenous components of the event-related potential — A diagnostic tool?" *Prog. Brain Res.* **70**, 87–102.

Elliott, F. W., Majda, A. J. & Horntrop, D. J. [1995] "Hierarchical Monte Carlo methods for fractal random fields," *J. Stat. Phys.* **81**, 717–736.

Fan, S. Z., Cheng, Y. J. & Lin, C. C. [1994] "Heart rate variability: A useful noninvasive tool in anesthesia," *Acata Anaesthesiol. Sinica.* **32**, 51–56.

Flury, B. & Riedwyl, H. [1988] *Multivariate Statistics. A Practical Approach* (Chapman and Hall, London).

Hughes, J. R. & John, E. R. [1999] "Conventional and quantitative electroencephalography in psychiatry," *J. Neuropsychiatry Clin. Neurosci.* **11**, 190–208.

Hurst, H. E., Blank, R. P. & Simaika, Y. M. [1965] *Long Term Storage*: *An Experimental Study* (Constable, London).

John, E. R., Prichep, L. S., Alper, K. R., Mas, F. G., Cancro, R., Easton, P. & Sverdlov, L. [1994a] "Quantitative electrophysiological characteristics and subtyping of schizophrenia," *Biol. Psychiatry* **36**, 801–826.

John, E. R., Prichep, L. S. & Easton, P. [1994b] "Standardized varimax descriptors of event related potentials: Evaluation of psychiatric patients," *Psychiatry Res.* **55**, 13–40.

Katz, M. J. & George, E. B. [1985] "Fractals and the analysis of growth paths," *Bull Math. Biol.* **47**, 273–286.

Kikkinides, E. S. & Burganos, V. N. [1999] "Structural and flow properties of binary media generated by fractional Brownian motion models," *Phys. Rev.* **E59**, 7185–7194.

Klocke, R. A., Schunemann, H. J. & Grant, B. J. [1995] "Distribution of pulmonary capillary transit times," *Am J. Respir. Crit. Care Med.* **152**, 2014–2020.

Lachenbruch, P. A. [1975] *Discriminant Analysis* (Hafner Press, NY).

Lambert, J. D. [1973] *Computational Methods in Ordinary Differential Equations* (Wiley, NY).

Lempel, A. & Ziv, J. [1976] "On the complexity of finite sequences," *IEEE Trans. Inform. Th.* **IT-22**, 75–81.

Mackey, M. C. & Glass, L. [1977] "Oscillation and chaos in physiological control systems," *Science Wash.* **197**, 287–289.

Mandelbrot, B. B. [1967] "How long is the coast of Britain? Statistical self-similarity and fractional dimension," *Science* **156**, 636–638.

McLachlan, G. J. [1992] *Discriminant Analysis and Statistical Pattern Recognition* (Wiley, NY).

Panel on Discriminant Analysis, Classification and Clustering. National Research Council [1988] *Discriminant Analysis and Clustering* (National Academy Press, WA).

Park, S. K. & Miller, K. W. [1988] "Random number generators. A good one is hard to find," *Commun. ACM.* **31**, 1192–1201.

Pedersen, T. S., Michelsen, P. K. & Rasmussen, J. J. [1996] "Lyapunov exponents and particle dispersion in drift wave turbulence," *Phys. Plasmas.* **38**, 2939–2950.

Press, W. H., Flannery, B. P., Teukolsky, S. A. & Vetterling, W. T. [1992] *Numerical Recipes. The Art of Scientific Computing* (Cambridge University Press, Cambridge).

Prichep, L. S., Mas, F., Hollander, E., Liebowitz, M., John, E. R., Almas, M., DeCaria, C. M. & Levine, R. H. [1993] "Quantitative electroencephalographic subtyping of obsessive-compulsive disorder," *Psychiatry Res.* **50**, 25–32.

Prichep, L. S., Alper, K. R., Kowalik, S. C. Vaysblat, L. S., Merkin, H. A., Tom, M., John, E. R. & Rosenthal, M. S. [1999] "Prediction of treatment outcome in cocaine dependent males using quantitative EEG," *Drug and Alcohol Dependence* **54**, 35–43.

Rapp, P. E., Latta, R. A. & Mees, A. I. [1988] "Parameter dependent transitions and the optimal control of dynamical diseases," *Bull. Math. Biol.* **50**, 227–253.

Rapp, P. E. & Schmah, T. I. [1996] "Complexity measures in molecular psychiatry," *Molecular Psychiatry* **1**, 408–416.

Rapp, P. E. [1999] "Nonlinear dynamics and the investigation of the human central nervous system," *Nonlinear Dynamics and Brain Functioning*, eds. Pradhan, N., Rapp, P. E. & Sreenivasan, R. (NovaScience, Commack, NY), pp. 1–5.

Rapp, P. E. & Schmah, T. I. [2000] "Dynamical analysis in clinical practice," *Chaos in Brain*? eds. Lehnertz, K., Arnhold, J., Grassberger, P. & Elger, C. E. (World Scientific, Singapore), pp. 52–65.

Rapp, P. E., Cellucci, C. J., Watanabe, T. A. A., Albano, A. M. & Schmah, T. I. [2001] "Surrogate data pathologies and the false-positive rejection of the

null hypothesis," *Int. J. Bifurcation and Chaos* **11**(4), 983–997.

Richardson, L. F. [1961] "The problem of contiguity," *Gen. Syst. Yearbook* **6**, 139–187.

Willis, D. M., Stevens, P. R. & Crothers, S. R. [1997] "Statistics of the largest geomagnetic storms per solar cycle (1844–1993)," *Ann. Geophys.* **15**, 719–728.

Young, E. S., Perror, P., Price, G. W. & Sadler, T. [1995] "Acute challenge ERP as a prognostic of stimulant therapy outcome in attention-deficit hyperactivity disorder," *Biol. Psychiatry* **37**, 23–33.

Ziv, J. & Lempel, A. [1978] "Compression of individual sequences via variable-rate coding," *IEEE Trans. Inform. Th.* **IT-24**, 530–536.

Appendix

A.1. *Probability that two groups are the same, zero noise case, average* $P_{\mathrm{SAME}} = 0.174 \times 10^{-6}$

	VDP	Lor	Rös	Hénon	Uni	F(Uni)	Nor	F(Nor)	Co
VDP		$< 10^{-8}$	$< 10^{-8}$	$< 10^{-8}$	$< 10^{-8}$	$< 10^{-8}$	$< 10^{-8}$	$< 10^{-8}$	$< 10^{-8}$
Lor			$< 10^{-8}$	$< 10^{-8}$	$< 10^{-8}$	$< 10^{-8}$	$< 10^{-8}$	$< 10^{-8}$	$< 10^{-8}$
Rös				$< 10^{-8}$	$< 10^{-8}$	$< 10^{-8}$	$< 10^{-8}$	$< 10^{-8}$	$< 10^{-8}$
Hénon					$< 10^{-8}$	$< 10^{-8}$	$< 10^{-8}$	$< 10^{-8}$	$< 10^{-8}$
Uni						$< 10^{-8}$	$< 10^{-8}$	$< 10^{-8}$	$< 10^{-8}$
F(Uni)							$< 10^{-8}$	$< 10^{-8}$	$< 10^{-8}$
Nor								$< 10^{-8}$	0.6×10^{-5}
F(Nor)									$< 10^{-8}$
Co									

A.2. *Probability of errors in a pairwise discrimination, zero noise case, average* $P_{\mathrm{ERROR}} = 0.155 \times 10^{-3}$

	VDP	Lor	Rös	Hénon	Uni	F(Uni)	Nor	F(Nor)	Co
VDP		$< 10^{-8}$	$< 10^{-8}$	$< 10^{-8}$	$< 10^{-8}$	$< 10^{-8}$	$< 10^{-8}$	$< 10^{-8}$	$< 10^{-8}$
Lor			$< 10^{-8}$	$< 10^{-8}$	$< 10^{-8}$	$< 10^{-8}$	$< 10^{-8}$	$< 10^{-8}$	$< 10^{-8}$
Rös				$< 10^{-8}$	$< 10^{-8}$	$< 10^{-8}$	$< 10^{-8}$	$< 10^{-8}$	$< 10^{-8}$
Hénon					$< 10^{-8}$	$< 10^{-8}$	$< 10^{-8}$	$< 10^{-8}$	$< 10^{-8}$
Uni						$< 10^{-8}$	$< 10^{-8}$	$< 10^{-8}$	$< 10^{-8}$
F(Uni)							$< 10^{-8}$	$< 10^{-8}$	$< 10^{-8}$
Nor								$< 10^{-8}$	0.6×10^{-2}
F(Nor)									$< 10^{-8}$
Co									

A.3. *Classification matrix, minimum Mahalanobis distance criterion, zero noise case, percent incorrect* = 0%

	VDP	Lor	Rös	Hénon	Uni	F(Uni)	Nor	F(Nor)	Co
VDP	10								
Lor		10							
Rös			10						
Hénon				10					
Uni					10				
F(Uni)						10			
Nor							10		
F(Nor)								10	
Co									10

A.4. *Classification matrix, maximum Bayesian likelihood criterion, zero noise case, percent incorrect* = 0%

	VDP	Lor	Rös	Hénon	Uni	F(Uni)	Nor	F(Nor)	Co
VDP	10								
Lor		10							
Rös			10						
Hénon				10					
Uni					10				
F(Uni)						10			
Nor							10		
F(Nor)								10	
Co									10

A.5. *Probability that two groups are the same, SNR* = 10 dB, *average* $P_{\mathrm{SAME}} = 0.718 \times 10^{-4}$

	VDP	Lor	Rös	Hénon	Uni	F(Uni)	Nor	F(Nor)	Co
VDP		0.5×10^{-7}	$< 10^{-8}$	$< 10^{-8}$	$< 10^{-8}$	$< 10^{-8}$	$< 10^{-8}$	$< 10^{-7}$	$< 10^{-8}$
Lor			$< 10^{-8}$	$< 10^{-8}$	$< 10^{-8}$	$< 10^{-8}$	$< 10^{-8}$	0.3×10^{-2}	$< 10^{-8}$
Rös				$< 10^{-8}$	$< 10^{-8}$	$< 10^{-8}$	$< 10^{-8}$	$< 10^{-8}$	$< 10^{-8}$
Hénon					0.6×10^{-7}	$< 10^{-8}$	$< 10^{-7}$	$< 10^{-8}$	0.3×10^{-7}
Uni						$< 10^{-8}$	0.2×10^{-5}	$< 10^{-8}$	$< 10^{-8}$
F(Uni)							$< 10^{-8}$	0.6×10^{-7}	$< 10^{-8}$
Nor								$< 10^{-8}$	0.5×10^{-4}
F(Nor)									$< 10^{-8}$
Co									

A.6. *Probability of errors in a pairwise discrimination,* $SNR = 10\,\mathrm{dB}$, *average* $P_{\mathrm{ERROR}} = 0.263 \times 10^{-2}$

	VDP	Lor	Rös	Hénon	Uni	F(Uni)	Nor	F(Nor)	Co
VDP		0.9×10^{-4}	$< 10^{-8}$	$< 10^{-8}$	$< 10^{-8}$	0.5×10^{-5}	$< 10^{-8}$	0.0002	$< 10^{-8}$
Lor			$< 10^{-8}$	$< 10^{-8}$	$< 10^{-8}$	0.7×10^{-5}	$< 10^{-8}$	0.0742	$< 10^{-8}$
Rös				$< 10^{-8}$	$< 10^{-8}$	$< 10^{-8}$	$< 10^{-8}$	$< 10^{-8}$	$< 10^{-8}$
Hénon					10^{-4}	$< 10^{-8}$	0.0002	$< 10^{-8}$	0.6×10^{-4}
Uni						$< 10^{-8}$	0.0030	$< 10^{-8}$	10^{-6}
F(Uni)							$< 10^{-8}$	10^{-4}	$< 10^{-8}$
Nor								$< 10^{-8}$	0.0166
F(Nor)									$< 10^{-8}$
Co									

A.7. *Classification matrix, minimum Mahalanobis distance criterion,* $SNR = 10\,\mathrm{dB}$, *percent incorrect* $= 2.2\%$

	VDP	Lor	Rös	Hénon	Uni	F(Uni)	Nor	F(Nor)	Co
VDP	10								
Lor		10							
Rös			10						
Hénon				10					
Uni					10				
F(Uni)						10			
Nor							9		1
F(Nor)		1						9	
Co									10

A.8. *Classification matrix, maximum Bayesian likelihood criterion,* $SNR = 10\,\mathrm{dB}$, *percent incorrect* $= 1.1\%$

	VDP	Lor	Rös	Hénon	Uni	F(Uni)	Nor	F(Nor)	Co
VDP	10								
Lor		10							
Rös			10						
Hénon				10					
Uni					10				
F(Uni)						10			
Nor							10		
F(Nor)		1						9	
Co									10

A.9. *Probability that two groups are the same, SNR = 5 dB, average* $P_{\mathrm{SAME}} = 0.0006$

	VDP	Lor	Rös	Hénon	Uni	F(Uni)	Nor	F(Nor)	Co
VDP		0.5×10^{-5}	$< 10^{-8}$	$< 10^{-8}$	$< 10^{-8}$	0.0032	$< 10^{-8}$	0.4×10^{-5}	$< 10^{-8}$
Lor			$< 10^{-8}$	$< 10^{-8}$	$< 10^{-8}$	0.0016	$< 10^{-8}$	0.0007	$< 10^{-8}$
Rös				$< 10^{-8}$	$< 10^{-8}$	$< 10^{-8}$	$< 10^{-8}$	$< 10^{-8}$	$< 10^{-8}$
Hénon					10^{-6}	$< 10^{-8}$	0.0007	$< 10^{-8}$	0.2×10^{-4}
Uni						$< 10^{-8}$	0.0006	$< 10^{-8}$	0.8×10^{-6}
F(Uni)							$< 10^{-8}$	0.0154	$< 10^{-8}$
Nor								$< 10^{-8}$	0.0003
F(Nor)									$< 10^{-8}$
Co									

A.10. *Probability of errors in a pairwise discrimination, SNR = 5 dB, average* $P_{\mathrm{ERROR}} = 0.0130$

	VDP	Lor	Rös	Hénon	Uni	F(Uni)	Nor	F(Nor)	Co
VDP		0.0049	$< 10^{-8}$	$< 10^{-8}$	$< 10^{-8}$	0.0796	$< 10^{-8}$	0.0043	$< 10^{-8}$
Lor			$< 10^{-8}$	$< 10^{-8}$	$< 10^{-8}$	0.0641	$< 10^{-8}$	0.0494	$< 10^{-8}$
Rös				$< 10^{-8}$	$< 10^{-8}$	$< 10^{-8}$	$< 10^{-8}$	$< 10^{-8}$	$< 10^{-8}$
Hénon					0.0019	$< 10^{-8}$	0.0485	$< 10^{-8}$	0.0110
Uni						$< 10^{-8}$	0.0459	$< 10^{-8}$	0.0013
F(Uni)							$< 10^{-8}$	0.1229	3×10^{-6}
Nor								$< 10^{-8}$	0.0339
F(Nor)									0.9×10^{-5}
Co									

A.11. *Classification matrix, minimum Mahalanobis distance criterion, SNR = 5 dB, percent incorrect = 4.4%*

	VDP	Lor	Rös	Hénon	Uni	F(Uni)	Nor	F(Nor)	Co
VDP	10								
Lor		9						1	
Rös			10						
Hénon				10					
Uni					10				
F(Uni)		1				8		1	
Nor				1			9		
F(Nor)								10	
Co									10

A.12. *Classification matrix, maximum Bayesian likelihood criterion,* $SNR = 5\,\mathrm{dB}$, *percent incorrect* $= 4.4\%$

	VDP	Lor	Rös	Hénon	Uni	F(Uni)	Nor	F(Nor)	Co
VDP	10								
Lor		9						1	
Rös			10						
Hénon				10					
Uni					10				
F(Uni)						9		1	
Nor				1	1		8		
F(Nor)								10	
Co									10

A.13. *Probability that two groups are the same,* $SNR = 0\,\mathrm{dB}$, *average* $P_{\mathrm{SAME}} = 0.0381$

	VDP	Lor	Rös	Hénon	Uni	F(Uni)	Nor	F(Nor)	Co
VDP		0.0447	0.2×10^{-7}	0.4×10^{-6}	0.3×10^{-6}	0.0047	0.9×10^{-5}	0.1917	10^{-6}
Lor			$< 10^{-8}$	10^{-7}	10^{-8}	0.0774	0.6×10^{-6}	0.1960	0.3×10^{-5}
Rös				$< 10^{-8}$	$< 10^{-8}$	0.2×10^{-7}	$< 10^{-8}$	$< 10^{-8}$	$< 10^{-8}$
Hénon					0.0009	$< 10^{-8}$	0.0183	0.7×10^{-7}	0.0125
Uni						$< 10^{-8}$	0.7698	0.8×10^{-7}	0.0022
F(Uni)							0.7×10^{-7}	0.0544	$< 10^{-8}$
Nor								0.8×10^{-6}	0.0391
F(Nor)									0.5×10^{-6}
Co									

A.14. *Probability of errors in a pairwise discrimination,* $SNR = 0\,\mathrm{dB}$, *average* $P_{\mathrm{ERROR}} = 0.0520$

	VDP	Lor	Rös	Hénon	Uni	F(Uni)	Nor	F(Nor)	Co
VDP		0.0890	0.3×10^{-4}	0.0007	0.0006	0.0889	0.0070	0.2277	0.0018
Lor			0.8×10^{-5}	0.0003	0.2×10^{-4}	0.1830	0.0011	0.2289	0.0032
Rös				$< 10^{-8}$	$< 10^{-8}$	0.3×10^{-4}	0.3×10^{-7}	0.2×10^{-4}	$< 10^{-8}$
Hénon					0.0534	0.2×10^{-5}	0.1284	0.0001	0.1163
Uni						0.6×10^{-5}	0.3439	0.0002	0.0710
F(Uni)							0.0001	0.1681	0.5×10^{-5}
Nor								0.0014	0.1552
F(Nor)									0.0010
Co									

A.15. *Classification matrix, minimum Mahalanobis distance criterion,* $SNR = 0$ dB, *percent incorrect* $= 20\%$

	VDP	Lor	Rös	Hénon	Uni	F(Uni)	Nor	F(Nor)	Co
VDP	9							1	
Lor		9				1			
Rös			10						
Hénon				10					
Uni					6		4		
F(Uni)		2				7		1	
Nor				1	1		7		1
F(Nor)	3	1						6	
Co				1			1		8

A.16. *Classification matrix, maximum Bayesian likelihood criterion,* $SNR = 0$ dB, *percent incorrect* $= 21\%$

	VDP	Lor	Rös	Hénon	Uni	F(Uni)	Nor	F(Nor)	Co
VDP	9							1	
Lor		6				3		1	
Rös			10						
Hénon				9	1				
Uni					6		4		
F(Uni)		1				9			
Nor					1		8		1
F(Nor)	1	1				2		6	
Co				1			1		8