Electrocardiogram classification using delay differential equations

Claudia Lainscsek and Terrence J. Sejnowski
Computational Neurobiology Laboratory, Howard Hughes Medical Institute, Salk Institute for Biological Studies, La Jolla, California 92037, USA and Institute for Neural Computation, University of California San Diego, La Jolla, California 92093, USA

(Received 3 February 2013; accepted 5 June 2013; published online 26 June 2013)

Time series analysis with nonlinear delay differential equations (DDEs) reveals nonlinear as well as spectral properties of the underlying dynamical system. Here, global DDE models were used to analyze 5 min data segments of electrocardiographic (ECG) recordings in order to capture distinguishing features for different heart conditions such as normal heart beat, congestive heart failure, and atrial fibrillation. The number of terms and delays in the model as well as the order of nonlinearity of the model have to be selected that are the most discriminative. The DDE model form that best separates the three classes of data was chosen by exhaustive search up to third order polynomials. Such an approach can provide deep insight into the nature of the data since linear terms of a DDE correspond to the main time-scales in the signal and the nonlinear terms in the DDE are related to nonlinear couplings between the harmonic signal parts. The DDEs were able to detect atrial fibrillation with an accuracy of 72%, congestive heart failure with an accuracy of 88%, and normal heart beat with an accuracy of 97% from 5 min of ECG, a much shorter time interval than required to achieve comparable performance with other methods. © *2013 AIP Publishing LLC*. [http://dx.doi.org/10.1063/1.4811544]

Cardiovascular diseases are the main cause of deaths worldwide and a major cost in health care. Better diagnosis methods are much needed. Here, delay differential equations (DDEs) are used to discriminate short (5 min) electrocardiography (ECG) data segments. This method does not require any preprocessing of the data, is performed on the time series themselves, is computationally fast, and could be the basis for a real time diagnostic system. DDEs reveal non-linear properties as well as spectral properties of the data. A DDE relates a differential and delay embedding in a complex manner to extract distinctive dynamical properties of the underlying dynamical system. DDE analysis is a time domain tool that combines aspects of nonlinear dynamics, Fourier analysis, and higher-order statistics.

I. INTRODUCTION

In global vector field reconstruction,[1–4] the recorded data are used to generate a model, whose dynamical behavior is equivalent to the original system. Equivalence is not required for our data analysis method. Nonetheless, the identification technique provides a global model that captures some essential features of the underlying dynamics.

The techniques introduced here are based on Delay Differential Equations (DDEs). DDEs are a generalization of ordinary differential equations (ODEs) with time delays. DDEs have to be used to describe the underlying dynamics in particular physical and biological processes. Such processes are typically characterized by a delayed reaction (see Driver[5] for a list of examples). Delays also have an important role in analyzing ECG data.[6]

Solving even the simplest linear DDE $\dot{x}(t) = \alpha x(t-\tau)$ is complicated (see, e.g., Ref. 7) and not within the scope of this paper. We do not seek DDEs that predict time series but rather global DDE models that capture distinguishing features of data for different heart conditions such as normal heart beat, congestive heart failure, and atrial fibrillation. The question whether an ECG is best modeled by a linear or non-linear process is directly related to the structure selection of the DDE: Is a linear DDE sufficient or are non-linear terms needed? How many terms, how many delays and what kind of non-linearity should be used?

DDEs can be seen as a flavor of a autoregressive (AR) model or an autoregressive moving average (ARMA) model[8–11] where the time series on the left side of the equation is replaced by the derivative. Lately delay systems have been used in the context of reservoir computing (RC).[12–15] RC is a recently introduced, bio-inspired, machine-learning paradigm for processing empirical data that is mimicking neuronal networks. A DDE in this context is the simplest nonlinear delay system with a singular node.

A motivation for DDE analysis of non-linear data comes from embedding theory in non-linear time series analysis. An embedding converts a single time series into a multidimensional object in an embedding space (Whitney,[16] Packard *et al.*,[17] Takens,[18] and Sauer *et al.*[2]). The reconstructed attractor reveals basic properties (dimension, Lyapunov spectrum, and entropy) of the true attractor of the system. It allows valuable information to be obtained about the dynamics of the system without having direct access to all the systems variables.

There are two basic embeddings: delay and derivative embeddings. For a delay embedding, the time series itself and its delayed versions are used to construct the embedding; for the derivative embedding the time series and its

successive derivatives are used. Judd and Mees[19] introduced the idea of non-uniform embeddings for time series with components of multiple time-scales. DDE analysis then relates aspects of the different embeddings: the derivative of the time-series is related to functions of non-uniformly delayed versions of that time series.

DDE data analysis can be also seen as a novel way of combining Fourier analysis and higher-order statistics in a time domain framework. The relationship between frequency analysis and analysis of frequency and/or phase couplings in the time domain is poorly understood (see, e.g., Refs. 20–24). The linear terms of a DDE correspond to the main frequencies in the signal. For n independent frequencies in a signal a linear DDE with $2n$–1 terms is needed to describe such data. The nonlinear terms in the DDE are related to nonlinear couplings between the harmonics. DDEs can also be expanded in a Yule-Walker-like way[25,26] and the DDE coefficients then can be rewritten as functions of dynamical higher-order data correlations. These dynamical higher-order data correlations are generalizations of Nth order data moment functions such as, e.g., the auto-correlation (2nd order moment) and the bi-correlation (3rd order moment). The paper is organized as follows: Sec. II shows the connection between DDEs and classical Fourier analysis and higher order statistics (HOS). In Sec. III, good classifiers for ECG data are found via DDE analysis. Section IV is the discussion.

II. DELAY DIFFERENTIAL EQUATIONS AND TIME-DOMAIN FREQUENCY ANALYSIS

DDE analysis is done in the time domain on the time-series themselves and not in the frequency domain. The DDE framework combines linear and non-linear information from the data in a complex and not easily understandable way. To gain some insight in the meaning of the different terms of a DDE, we show the correspondence of the linear terms to the main time-scales or frequencies in the signal and how the non-linear terms contain information about non-linear couplings.

A. Linear DDEs

The simplest linear DDE is

$$\dot{x} = a x_\tau \,, \tag{1}$$

where $x_\tau = x(t-\tau)$. Solving this equation in general is non-trivial (see, e.g., Ref. 7) and beyond the scope of this paper. However, looking at special solutions can lead to understanding the terms in a DDE as used here for detection/classification purposes. A special solution of Eq. (1) is (see Ref. 7)

$$x(t) = \cos(\omega t)\,; \quad a = (-1)^n \omega\,; \quad \tau = \frac{\pi\,(2n-1)}{2\omega}, \tag{2}$$

where $n \in \mathbb{N}$. The coefficient a is proportional to the frequency and the time delay τ is inversely proportional to the frequency. For a signal with a frequency f and $\omega = 2\pi f$ the delay is then

$$\tau = \frac{(2n-1)}{4f}\,. \tag{3}$$

The delay is inversely proportional to the frequency and the coefficient a is directly proportional to the frequency.

A special solution of the linear DDE

$$\dot{x} = \sum_{i=1}^{N} a_i x_{\tau_i} \tag{4}$$

is

$$x(t) = \sum_{k=1}^{2N-1} \cos(\omega_k t)\,; \quad \tau_i = \frac{\pi(2n-1)}{2\omega_j}, \tag{5}$$

where $n \in \mathbb{N}$ are arbitrary integers and all delays τ_i are related to one of the frequencies. The expressions for the coefficients a are more complicated than in Eq. (2) and each depends on all the frequencies in the signal.

Equations (4) and (5) imply that we need a DDE with $2N$–1 linear terms to describe a harmonic signal with N frequencies. If we consider Eq. (1) and a sum of three harmonics, $x(t) = \sum_{i=1}^{3} \cos(\omega_i t)$, Eq. (1) cannot be solved analytically. To estimate the value of the coefficient a we expand Eq. (1) as a Yule-Walker-like equation:[25,26] We multiply both sides of Eq. (1) with x_τ and apply the expectation operator $\langle F(t)\rangle \equiv \lim_{T\to\infty}\left(\frac{1}{T}\int_0^T F(t)\,\mathrm{d}t\right)$ and get

$$a = \frac{\langle \dot{x}\, x_\tau\rangle}{\langle x_\tau^2\rangle}\,. \tag{6}$$

The numerator in Eq. (6) looks like a "dynamical" version of the autocorrelation function $\langle x\, x_\tau\rangle$ and it can be rewritten as delay derivatives of the autocorrelation function in the case of a bounded stationary signal,

$$\begin{aligned}
\langle \dot{x} x_\tau\rangle &= \lim_{T\to\infty}\frac{1}{T}\int_0^T \dot{x} x_\tau\, dt \\
&= \lim_{T\to\infty}\frac{1}{T}[x x_\tau]_0^T - \lim_{T\to\infty}\frac{1}{T}\int_0^T x\frac{dx_\tau}{dt}\,dt \\
&= \lim_{T\to\infty}\frac{1}{T}\int_0^T x\frac{dx_\tau}{d\tau}\,d\tau \\
&= \frac{d}{d\tau}\langle x x_\tau\rangle\,.
\end{aligned} \tag{7}$$

For $x(t) = \sum_{i=1}^{3}\cos(\omega_i t)$, the expressions in Eq. (6) are

$$\begin{aligned}
\langle \dot{x}\, x_\tau\rangle &\overset{\omega_3\neq\omega_j}{=} -\frac{1}{2}\sum_{i=1}^{3}\sin(\omega_i\tau)\omega_i, \\
\langle \dot{x}\, x_\tau\rangle &\overset{\omega_3=\omega_j}{=} -\frac{1}{2}\,\omega_2\sin(\tau\omega_2) - 2\omega_1\sin(\tau\omega_1), \\
\langle \dot{x}\, x_\tau\rangle &\overset{\omega_3=\omega_j}{=} -\frac{1}{2}\,\omega_1\sin(\tau\omega_1) - 2\omega_2\sin(\tau\omega_2), \\
\langle x_\tau^2\rangle &\overset{\omega_3\neq\omega_j}{=} \frac{3}{2}, \\
\langle x_\tau^2\rangle &\overset{\omega_3=\omega_j}{=} \frac{5}{2}\,; \quad j = 1,2\,.
\end{aligned} \tag{8}$$

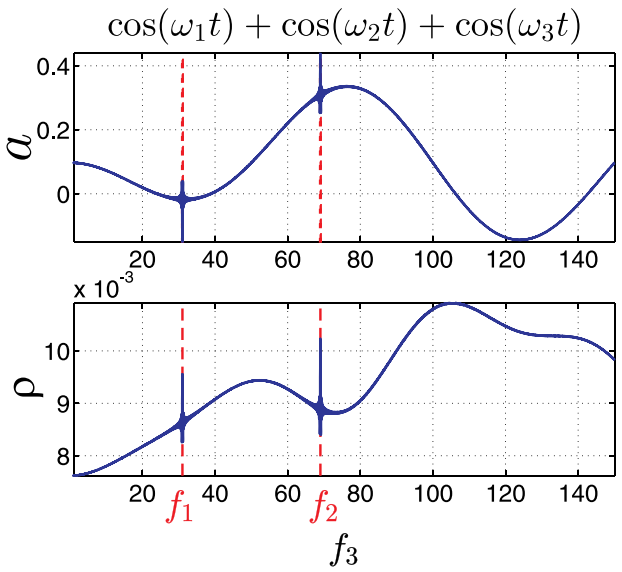

FIG. 1. Error ρ and coefficient a for the linear DDE $\dot{x} = a\,x_\tau$ with $\tau = 10\,\delta t$ vs. frequency f_3 for the signal $x(t) = \cos(\omega_1 t) + \cos(\omega_2 t) + \cos(\omega_3 t)$ with $f_1 = 31$ Hz and $f_2 = 69$ Hz ($\omega = 2\pi f$).

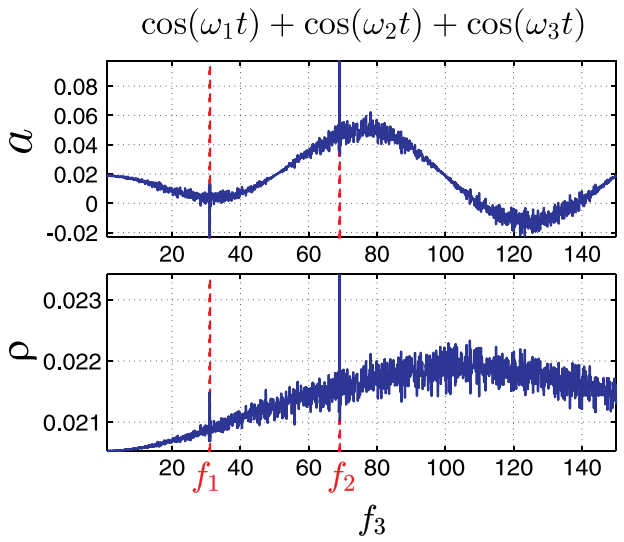

FIG. 2. Error ρ and coefficient a for the linear DDE $\dot{x} = a\,x_\tau$ with $\tau = 10\,\delta t$ vs. frequency f_3. White noise η was added to the signal: $D = \cos(\omega_1 t) + \cos(\omega_2 t) + \eta$, where the signal-to-noise ration SNR $= -10$ dB and $f_1 = 31$ Hz and $f_2 = 69$ Hz ($\omega = 2\pi f$). The coefficient a and error ρ were then plotted for $x(t) = \mathcal{D} + \cos(\omega_3 t)$ with f_3 varying between 0 and 150 Hz.

The coefficient a in Eq. (6) is a smooth function with singularities at $\omega_3 = \omega_i$, $i = 1,2$. Therefore estimating a numerically can be used as a time domain frequency detection tool: To detect the two frequencies in the signal $\mathcal{D} = \cos(\omega_1\tau) + \cos(\omega_2\tau)$ $(f_1 = 31\,\text{Hz}, f_2 = 69\,\text{Hz}, \omega = 2\pi f)$ the term $\cos(\omega_3\tau)$ was added for a range of frequencies $f_3 = \frac{\omega_3}{2\pi}$. In Fig. 1, we estimated the coefficient a numerically by a singular value decomposition (SVD) algorithm[27] and then computed the least square error $\sqrt{\sum(\dot{x} - a\,x_\tau)^2}$ for Eq. (6) for $\mathcal{D} + \cos(\omega_3\tau)$ with f_3 varying from 0 to 150 Hz. The delay was 10 δt, where $\delta t = \frac{1}{f_s}$ with a sampling rate $f_s = 1000$ Hz. The two singularities at the two frequencies $f_1 = 31$ Hz and $f_2 = 69$ Hz are clearly visible in both plots. The choice of a different delay would change the shape of curve (see Eqs. (7) and (8)). The error $\rho = |\dot{x} - \frac{\langle \dot{x}\,x_\tau\rangle}{\langle x_\tau^2\rangle}\,x_\tau|$ is nonzero since Eq. (6) is not an exact solution.

The possible advantages for the use this time-domain frequency analysis tool are that it can be applied for short time series and for sparse data: a can be estimated for the set of all points for which the derivative can be computed and the delayed point exists. Missing points can be left out.

The method is also fairly noise insensitive. In Fig. 2, we added to the signal white noise η with a signal-to-noise ratio of SNR $= -10$ dB which is more noise than signal. We then did the same numerical experiment as we did in Fig. 1: we estimated the coefficient a numerically by a singular value decomposition (SVD) algorithm and then computed the least square error for Eq. (6) for $D + \cos(\omega_3\tau)$ with f_3 varying from 0 to 150 Hz. Again, the two singularities at the two frequencies $f_1 = 31$ Hz and $f_2 = 69$ Hz are clearly visible in both plots (Fig. 2).

B. Nonlinear DDEs

In real world data, various frequency components do not always appear completely independently of one another. Those non-linear interactions of frequencies and their phases (e.g., quadratic phase coupling) cannot be detected by a power spectrum, the Fourier transform of the autocorrelation function (second-order cumulant), since phase relationships and frequency couplings of signals are lost. Such couplings are usually detected via higher order spectra or bispectral analysis.[28–35] The bispectrum or bispectral density is the Fourier transform of the third-order cumulant (bicorrelation function). Consider the signal $x(t) = A_1\cos(\omega_1 t + \varphi_1) + A_2\cos(\omega_2 t + \varphi_2)$ which is passed through a quadratic nonlinear system $h(t) = b\,x^2(t)$ where b is a non-zero constant. On the output of the system, the signal will include the harmonic components: $(2\omega_1, 2\varphi_1), (2\omega_2, 2\varphi_2), (\omega_1 + \omega_2, \varphi_1 + \varphi_2)$, and $(\omega_1 - \omega_2, \varphi_1 - \varphi_2)$. These phase relations are called quadratic phase coupling (QPC). Since we are here interested in the couplings of frequencies we will consider the special case of quadratic frequency coupling (QFC) when the phases are zero $(\varphi_1 = \varphi_2 = 0)$: For a signal $x(t) = \cos(\omega_1 t) + \cos(\omega_2 t) + \cos(\omega_3 t)$ frequency coupling occurs when ω_3 is a multiple of one of the frequencies or of the sum or difference of the two frequencies.

A simple DDE with one non-linear term,

$$\dot{x} = a\,x_{\tau_1}\,x_{\tau_2}, \tag{9}$$

cannot be solved analytically, but as shown for the linear case in Sec. I, it can be expanded as a Yule-Walker-like equation:[25,26] We multiply both sides of Eq. (9) with $x_{\tau_1}\,x_{\tau_2}$ and apply the expectation operator and get

$$a = \frac{\langle \dot{x}\,x_{\tau_1}\,x_{\tau_2}\rangle}{\langle x_{\tau_1}^2\,x_{\tau_2}^2\rangle}\,. \tag{10}$$

The numerator $\langle \dot{x}\,x_{\tau_1}\,x_{\tau_2}\rangle$ in Eq. (10) looks like a dynamical version of the bicorrelation[36] $\langle x\,x_{\tau_1}\,x_{\tau_2}\rangle$. It can be rewritten as delay derivatives of the moments in the case of a bounded stationary signal,

$$\langle \dot{x}x_{\tau_1}x_{\tau_2}\rangle = \lim_{T\to\infty}\frac{1}{T}\int_0^T \dot{x}x_{\tau_1}x_{\tau_2}dt = \lim_{T\to\infty}\frac{1}{T}[xx_{\tau_1}x_{\tau_2}]_0^T - \lim_{T\to\infty}\frac{1}{T}\int_0^T x\frac{d}{dt}(x_{\tau_1}x_{\tau_2})dt = \lim_{T\to\infty}\frac{1}{T}\int_0^T x\frac{dx_{\tau_1}}{d\tau_1}x_{\tau_2}dt$$
$$+\lim_{T\to\infty}\frac{1}{T}\int_0^T xx_{\tau_1}\frac{dx_{\tau_1}}{d\tau_2}dt = \frac{d}{d\tau_1}\langle xx_{\tau_1}x_{\tau_2}\rangle + \frac{d}{d\tau_2}\langle xx_{\tau_1}x_{\tau_2}\rangle . \tag{11}$$

For $x(t)=\cos(\omega_1 t)+\cos(\omega_2 t)+\cos(\omega_3 t)$, the bicorrelation is only non-zero when ω_3 is a multiple of one of the frequencies or of the sum or difference of the two frequencies. In these cases, the expressions for the dynamic bicorrelation are

$$\begin{aligned}\langle \dot{x}\,x_{\tau_1}\,x_{\tau_2}\rangle &\overset{\omega_3=\omega_1+\omega_2}{=} \frac{1}{4}(\omega_1(-\sin(\tau_2\omega_1\pm\tau_1\omega_2)-\sin(\tau_1\omega_1\pm\tau_2\omega_2)+\sin(\pm\tau_2\omega_2-\tau_1(\omega_1\pm\omega_2))\\ &+\sin(\pm\tau_1\omega_2-\tau_2(\omega_1\pm\omega_2)))+\omega_2(-\sin(\tau_2\omega_1\pm\tau_1\omega_2)-\sin(\tau_1\omega_1\pm\tau_2\omega_2)\\ &+\sin(\tau_2\omega_1-\tau_1(\omega_1\pm\omega_2))+\sin(\tau_1\omega_1-\tau_2(\omega_1\pm\omega_2)))),\\ \langle \dot{x}\,x_{\tau_1}\,x_{\tau_2}\rangle &\overset{\omega_3=2\omega_i}{=} \frac{1}{4}(\omega_i\sin((\tau_1-2\tau_2)\omega_i)-\omega_i\sin((2\tau_1-\tau_2)\omega_i)-2\omega_i\sin((\tau_1+\tau_2)\omega_i)),\\ \langle \dot{x}\,x_{\tau_1}\,x_{\tau_2}\rangle &\overset{\omega_3=\frac{\omega_i}{2}}{=} \frac{1}{8}\omega_i\left(\sin\left(\frac{1}{2}(\tau_1-2\tau_2)\omega_i\right)-\sin\left(\frac{1}{2}(2\tau_1-\tau_2)\omega_i\right)-2\sin\left(\frac{1}{2}(\tau_1+\tau_2)\omega_i\right)\right);\ i=1,2.\end{aligned} \tag{12}$$

and for the denominator of Eq. (10)

$$\begin{aligned}x_{\tau_1}^2\,x_{\tau_2}^2 &\overset{\omega_3\neq\omega_i}{=} \frac{1}{8}(\cos(2(\tau_1-\tau_2)\omega_1)+4\cos((\tau_1-\tau_2)(\omega_1-\omega_2))+\cos(2(\tau_1-\tau_2)\omega_2)+4\cos((\tau_1-\tau_2)(\omega_1+\omega_2))\\ &+4\cos((\tau_1-\tau_2)(\omega_1-\omega_3))+4\cos((\tau_1-\tau_2)(\omega_2-\omega_3))+\cos(2(\tau_1-\tau_2)\omega_3)+4\cos((\tau_1-\tau_2)(\omega_1+\omega_3))\\ &+4\cos((\tau_1-\tau_2)(\omega_2+\omega_3))+18),\\ x_{\tau_1}^2\,x_{\tau_2}^2 &\overset{\omega_3=\omega_1}{=} 2\cos(2(\tau_1-\tau_2)\omega_1)+2\cos((\tau_1-\tau_2)(\omega_1-\omega_2))+\frac{1}{8}\cos(2(\tau_1-\tau_2)\omega_2)+2\cos((\tau_1-\tau_2)(\omega_1+\omega_2))+\frac{25}{4},\\ x_{\tau_1}^2\,x_{\tau_2}^2 &\overset{\omega_3=\omega_2}{=} \frac{1}{8}\cos(2(\tau_1-\tau_2)\omega_1)+2\cos((\tau_1-\tau_2)(\omega_1-\omega_2))+2\cos(2(\tau_1-\tau_2)\omega_2)+2\cos((\tau_1-\tau_2)(\omega_1+\omega_2))+\frac{25}{4}.\end{aligned} \tag{13}$$

The coefficient a in Eq. (10) is only non-zero for QFC frequencies. We therefore can use this equation to detect non-linear couplings in the time domain in the same way as we used the linear DDE Eq. (6) to detect frequencies.

Fig. 3 shows how non-linear terms can detect frequency couplings. For $x(t)=\cos(\omega_1 t)+\cos(\omega_2 t)+\cos(\omega_3 t)$ ($\omega=2\pi f$) where the frequencies f_1 and f_2 were 31 Hz and 69 Hz, respectively, and f_3 was varied from 0 to 150 Hz. For each f_3 we estimated the coefficient a numerically by a singular value decomposition (SVD) algorithm[27] and then computed the least square error $\sqrt{\sum(\dot{x}-a\,x_{\tau_1}\,x_{\tau_2})^2}$ for Eq. (9). The nonlinear coefficient a (top plot in Fig. 3) is only non-zero when there is QFC and is zero otherwise. The least square error (bottom plot in Fig. 3) shows a slope with spikes when there is QFC as well as at the 2 frequencies f_1 and f_2.

Equation (9) can therefore be used to detect QFC. Since the coefficient a in Eq. (10) is not an exact solution of Eq. (9), the error (lower plot in Fig. 3)) is $\rho=|\dot{x}-a\,x_{\tau_1}\,x_{\tau_2}|$. For all values of f_3 that are different from any coupling cases (sum or multiples of the two other frequencies) it should be $\rho=\dot{x}$ since $a=0$ for all non-coupling values of f_3. Numerically a will never be exactly zero, but a small value. Therefore, the error is $\rho=|\dot{x}-\frac{\langle \dot{x}\,x_{\tau_1}\,x_{\tau_2}\rangle}{\langle x_{\tau_1}^2\,x_{\tau_2}^2\rangle}\,x_{\tau_1}\,x_{\tau_2}|$. $\langle x_{\tau_1}^2\,x_{\tau_2}^2\rangle$ has a spike when f_3 is equal to f_1 and f_2 and $\langle \dot{x}\,x_{\tau_1}\,x_{\tau_2}\rangle$ has spikes for QFC cases. Therefore the error (lower plot in Fig. 3)) shows bumps for QFC cases and the frequencies. The delays $\tau_{1,2}$ only change the shape of the error function.

In this section, we wanted to show the connection of DDEs to spectral analysis: Delays connected to linear terms in the DDE relate to frequencies (see Eqs. (6)–(8) and Fig. 1) and delays connected to non-linear terms in the DDE relate to couplings between those frequencies (see Eqs. (10), (12), (13), and Fig. 3). Here, we are not aiming to interpret these plots quantitatively. Different delays would only change the shape of the curves, but not the fact that there are spikes for the frequencies in the linear case and spikes for the frequency couplings in the non-linear case. Therefore, any arbitrary choice of delays will give the same qualitative behavior.

III. DDE ANALYSIS OF ECG DATA

In Sec. II, we showed how linear terms of a DDE relate to the time scales or frequencies in the signal and the non-linear terms relate to non-linear couplings of those time

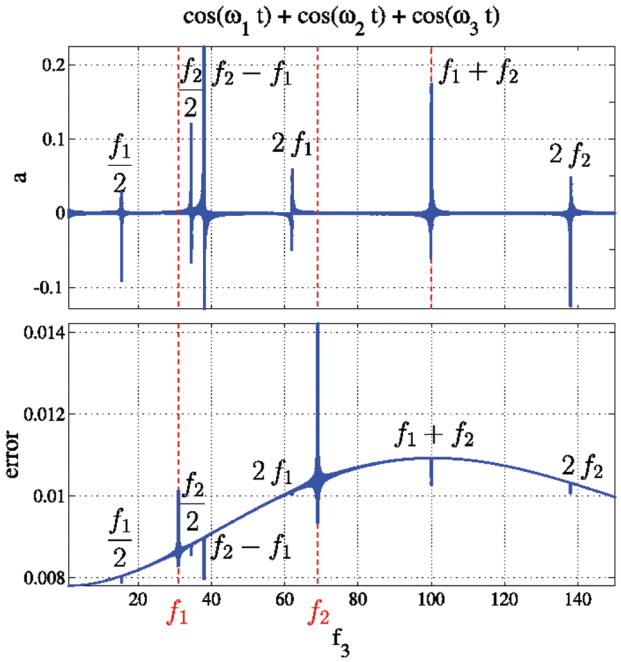

FIG. 3. Error and coefficient a for the nonlinear DDE $\dot{x} = a\, x_{\tau_1}\, x_{\tau_2}$ with $\tau_1 = 5\,\delta t$ and $\tau_2 = 11\,\delta t$ vs. frequency f_3 for the signal $x(t) = \cos(\omega_1 t) + \cos(\omega_2 t) + \cos(\omega_3 t)$ with $f_1 = 31$ Hz and $f_2 = 69$ Hz ($\omega = 2\pi f$).

scales or frequencies. Any DDE can be expanded in a Yule-Walker-like way and the coefficients will then be combinations of dynamical data correlations.

A DDE can also be interpreted as generic non-uniform embedding. Such a generic non-uniform embedding (the DDE model) unfolds timescales for the linear terms and couplings between frequencies for the non-linear terms as a spectrogram unfolds frequencies and higher order statistics (HOS) unfolds the couplings between frequencies. DDEs and spectral analysis are connected as we showed in the previous section. They are just operating in two different domains, the time domain and the spectral domain.

A DDE is a nonlinear extension of a non-uniform embedding with linear and/or nonlinear functions of the timeseries. Therefore, it can be tailored to the dynamics. HOS would have to be extended to a network of all possible HOS moments combined with linear spectral analysis to do the same. Therefore, a DDE is the simpler approach.

The structure or model form of the DDE and the delays for classification of heart data were in this section selected by an exhaustive search.

A. Data

We analyzed 24 h data from 15 young healthy persons in normal sinus rhythm (NSR) (ECG sample frequency: 128 Hz) of 15 congestive heart failure (CHF) patients (ECG sample frequency: 250 Hz) as well as of 15 subjects suffering from atrial fibrillation (AF) (ECG sample frequency: 128 Hz) selected from the Physionet database.[37] Table I lists the files. The first five subjects of each group were used for the CHAOS Controversial Topics in Nonlinear Dynamics "Is the Normal Heart Rate Chaotic?" (http://physionet.org/challenge/chaos). The other ten subjects from each group are randomly selected records from the same databases.

TABLE I. ECG data used. The three conditions are normal sinus rhythm (NSR), congestive heart failure (CHF), and atrial fibrillation (AF). The data were downloaded from the PHYSIONET database.[37]

	NSR	CHF	AF
subj.	http://physionet.org/physiobank/database/		
#	nsrdb/	chfdb/	ltafdb/
1	16265m.dat	chf01m.dat	11m.dat
2	16272m.dat	chf03m.dat	12m.dat
3	16786m.dat	chf07m.dat	15m.dat
4	16795m.dat	chf08m.dat	17m.dat
5	19090m.dat	chf12m.dat	18m.dat
6	16273m.dat	chf02m.dat	03m.dat
7	16420m.dat	chf04m.dat	05m.dat
8	16483m.dat	chf05m.dat	06m.dat
9	16539m.dat	chf06m.dat	07m.dat
10	16773m.dat	chf09m.dat	08m.dat
11	17052m.dat	chf10m.dat	10m.dat
12	17453m.dat	chf11m.dat	00m.dat
13	18177m.dat	chf13m.dat	01m.dat
14	18184m.dat	chf14m.dat	13m.dat
15	19088m.dat	chf15m.dat	16m.dat

B. Supervised structure selection

Typically, a nonlinear delay differential equation has the form

$$\begin{aligned}\dot{x} = f(a_i, x_{\tau_j}) = {} & a_1 x_{\tau_1} + a_2 x_{\tau_2} + a_3 x_{\tau_3} + \cdots + a_{i-1} x_{\tau_n} \\ & + a_i x_{\tau_1}^2 + a_{i+1} x_{\tau_1} x_{\tau_2} + a_{i+2} x_{\tau_1} x_{\tau_3} + \cdots \\ & + a_{j-1} x_{\tau_n}^2 + a_j x_{\tau_1}^3 + a_{j+1} x_{\tau_1} 2x_{\tau_2} + \cdots \\ & \vdots \\ & \cdots + a_l x_{\tau_n}^m, \end{aligned} \tag{14}$$

where $x = x(t)$ and $x_{\tau_j} = x(t - \tau_j)$. The DDE Eq. (14) has n delays, l monomials with coefficients $a_1, a_2, ..., a_l$, and a degree m of nonlinearity. By a k-term DDE, we mean a DDE with k monomials selected from the right-hand side of Eq. (14). Although quite flexible, as for any global modeling technique, there is a significant gain in accuracy by carefully selecting the structure of the model.[38–40] By structure selection or model learning, we mean retaining only those monomials that make the most significant contribution to the data dynamics. An equally important task is to select the right time-delays, since they are directly related to the primary time-scales and non-linear couplings between them of the dynamics under study.

Lainscsek *et al.*[40] used a genetic algorithm to find a single DDE model for the classification of Parkinson movement data. Here, we want to do an exhaustive search of models and delays and find the models and delays that best separate classes of data. To do so we look at all possible polynomial DDE models

TABLE II. One- and two-term models. An "x" denotes that the term $a_\star$ is nonzero. The different types of models are: "L"—linear, "S"— symmetric, "1"—single delay DDE. All other DDEs are non-linear and have two non-interchangeable delays. To save space in the table x_{τ_i} was written in the short form x_i.

model #	a_1 x_1	a_2 x_2	a_3 x_1^2	a_4 x_1x_2	a_5 x_2^2	a_6 x_1^3	a_7 $x_1^2x_2$	a_8 $x_1x_2^2$	a_9 x_2^3	model type
1	x									1,L
2			x							1
3				x						S
4						x				1
5							x			
6	x	x								S,L
7	x		x							1
8	x			x						
9	x					x				1
10	x						x			
11		x	x							
12		x		x						
13		x				x				
14		x					x			
15			x	x						
16			x		x					S
17			x			x				1
18			x				x			
19				x		x				
20				x			x			
21						x	x			
22						x			x	S
23							x	x		S

$$\dot{x} = a_1 x_{\tau_1} + a_2 x_{\tau_2} + a_3 x_{\tau_1}^2 + a_4 x_{\tau_1} x_{\tau_2} + a_5 x_{\tau_2}^2 + a_6 x_{\tau_1}^3 + a_7 x_{\tau_1}^2 x_{\tau_2} + a_8 x_{\tau_1} x_{\tau_2}^2 + a_9 x_{\tau_2}^3, \quad (15)$$

with some of the a_i equal to zero. Only models with up to three terms were considered. If the analysis did not give satisfactory results we added additional delays, increased the order of non-linearity and/or used DDEs with more than three terms. There were 5 one-term models, 18 two-term models, and 32 three-term models.

Tables II and III list all these models. Note that e.g. the DDE models $\dot{x} = a_1 x_{\tau_1} + a_2 x_{\tau_1} x_{\tau_2}$ and $\dot{x} = a_1 x_{\tau_2} + a_2 x_{\tau_1} x_{\tau_2}$ are the same with exchanged delays τ_1 and τ_2. Therefore, only the first of these two models were used. All such redundant DDE models were omitted. There were only two linear DDEs (model 1 and 5) while all others are non-linear. Seven of the DDEs had only one delay (models 1, 2, 4, 7, 9, 17, and 30) and nine models were symmetric (models 3, 6, 16, 22, 23, 25, 43, 52, 53) with two interchangeable delays.

C. Data analysis

The data were analyzed without filtering and no artifacts were removed from the data. The downloaded NSR and AF data were sampled at 128 Hz, but the CHF data were sampled at 250 Hz. To use the same DDE with the same delays for all data, the NSR and AF data were up-sampled using the MATLAB function resample[41] with the default options. Throughout this paper, we use 5 min non-overlapping data windows for our analysis. Each window was re-normalized to zero mean and unit variance to be able to compare data of different origin.

TABLE III. Three term models. An "x" denotes that the term $a_\star$ is nonzero. The different types of models are: "L"—linear, "S" —symmetric, "1"—single delay DDE. All other DDEs are non-linear and have two non-interchangeable delays. To save space in the table x_{τ_i} was written in the short form x_i.

model #	a_1 x_1	a_2 x_2	a_3 x_1^2	a_4 x_1x_2	a_5 x_2^2	a_6 x_1^3	a_7 $x_1^2x_2$	a_8 $x_1x_2^2$	a_9 x_2^3	model type
24	x	x	x							
25	x	x		x						S
26	x	x				x				
27	x	x					x			
28	x		x	x						
29	x		x		x					
30	x		x			x				1
31	x		x				x			
32	x			x		x				
33	x			x			x			
34	x					x	x			
35	x					x			x	
36	x						x	x		
37		x	x	x						
38		x	x			x				
39		x	x				x			
40		x		x		x				
41		x		x			x			
42		x				x	x			
43			x	x	x					S
44			x	x		x				
45			x	x			x			
46			x		x	x				
47			x		x		x			
48			x			x	x			
49			x			x			x	
50			x				x	x		
51				x		x	x			
52				x		x			x	S
53				x			x	x		S
54						x	x	x		
55						x	x		x	

For the model selection task, we have to choose a classifier, select training data, select a classification tool, and do some cross-validation to take the small number of subjects into account. In this manuscript, we chose seven different classifiers and tested the performance of each separately. Those classifiers were: (1) NSR vs. AF vs. CHF, (2) NSR vs. AF, (3) NSR vs. CHF, (4) AF vs. CHF, (5) NSR vs. (AF and CHF), (6) AF vs. (NSR and CHF), and (7) CHF vs. (NSR and AF). As training data we selected one 5 min data window every 20 min (e.g., for a 20 h recording of one subject 60 5 min data windows were used). We used a repeated random sub-sampling validation[42] where we trained on 10 subjects of each group and tested on the remaining 5 subjects of each group. This was repeated 300 times with each subject equally often used as training and testing subject. As classifier we used singular value decomposition (SVD).[27] As measure of performance we use Cohen's kappa κ[43–46] which can be computed directly from the confusion matrix.[47] A confusion matrix (also known as matching matrix, contingency table, or error matrix) is a specific table layout that allows visualization of the performance. Each column of the matrix

TABLE IV. Best DDE models selected for the seven classification tasks. The kappa values κ reported here were computed directly from the confusion matrices.[47] The area under the ROC curve A' can only be computed for the 6 binary classifiers. A' was computed from the ROC curves in Fig. 5. The units of the delays are time steps $\delta\tau = \frac{1}{f_s}$, where f_s is the sampling frequency.

delays		DDE model	confusion matrix	κ	A'
14	1	$\dot{x} = a_1x_1^2 + a_2x_2^2 + a_3x_1^3$		0.83	

actual \ predicted	NSR	AF	CHF
NSR	3445	398	26
AF	116	3813	193
CHF	0	600	2975

delays		DDE model	confusion matrix	κ	A'
12	18	$\dot{x} = a_1x_1 + a_2x_1x_2 + a_3x_1^3$		0.96	0.973

act. \ predicted	NSR	AF
NSR	3741	128
AF	41	4081

delays		DDE model	confusion matrix	κ	A'
15	8	$\dot{x} = a_1x_2 + a_2x_1^2 + a_3x_1^2x_2$		0.99	0.996

act. \ predicted	NSR	CHF
NSR	3841	28
CHF	3	3572

delays		DDE model	confusion matrix	κ	A'
5	49	$\dot{x} = a_1x_1x_2 + a_2x_1^3 + a_3x_1^2x_2$		0.85	0.959

act. \ predicted	AF	CHF
AF	3775	347
CHF	229	3346

delays		DDE model	confusion matrix	κ	A'
12	18	$\dot{x} = a_1x_1 + a_2x_1x_2 + a_3x_1^3$		0.97	0.981

act. \ predicted	NSR	AF,CHF
NSR	3722	147
AF,CHF	32	7665

delays		DDE model	confusion matrix	κ	A'
16	3	$\dot{x} = a_1x_2 + a_2x_1^2 + a_3x_1^3$		0.72	0.928

act. \ predicted	AF	NSR,CHF
AF	3349	773
NSR,CHF	735	6709

delays		DDE model	confusion matrix	κ	A'
1	5	$\dot{x} = a_1x_1 + a_2x_2 + a_3x_1x_2$		0.88	0.937

act. \ predicted	CHF	NSR,AF
CHF	3435	140
NSR,AF	464	7527

represents the instances in a predicted class, while each row represents the instances in an actual class.

The random sub-sampling validation gives 300 values for Cohen's kappa for each model and each set of delays. To choose the best model we searched the highest minimum of the mean of the 300 values for each classifier. The best models and delays for the seven classification tasks are listed in Table IV.

To test the performance of the classifiers in Table IV, we computed for each of the 7 models the 300 SVD-weights from the training data and then took the mean of those weights. These weights were then applied to the whole data set. Fig. 4 shows the computed features for all data for all seven classification tasks. The corresponding Cohen kappa values as well as the areas under the ROC curves are listed in Table IV. The separating hyperplanes between the conditions were selected by SVD. In Fig. 4, the distance d from these hyperplanes are shown. NSR is best separated from the 2 diseases ($\kappa = 0.96 - 0.99$). All other classifiers were also quite good.

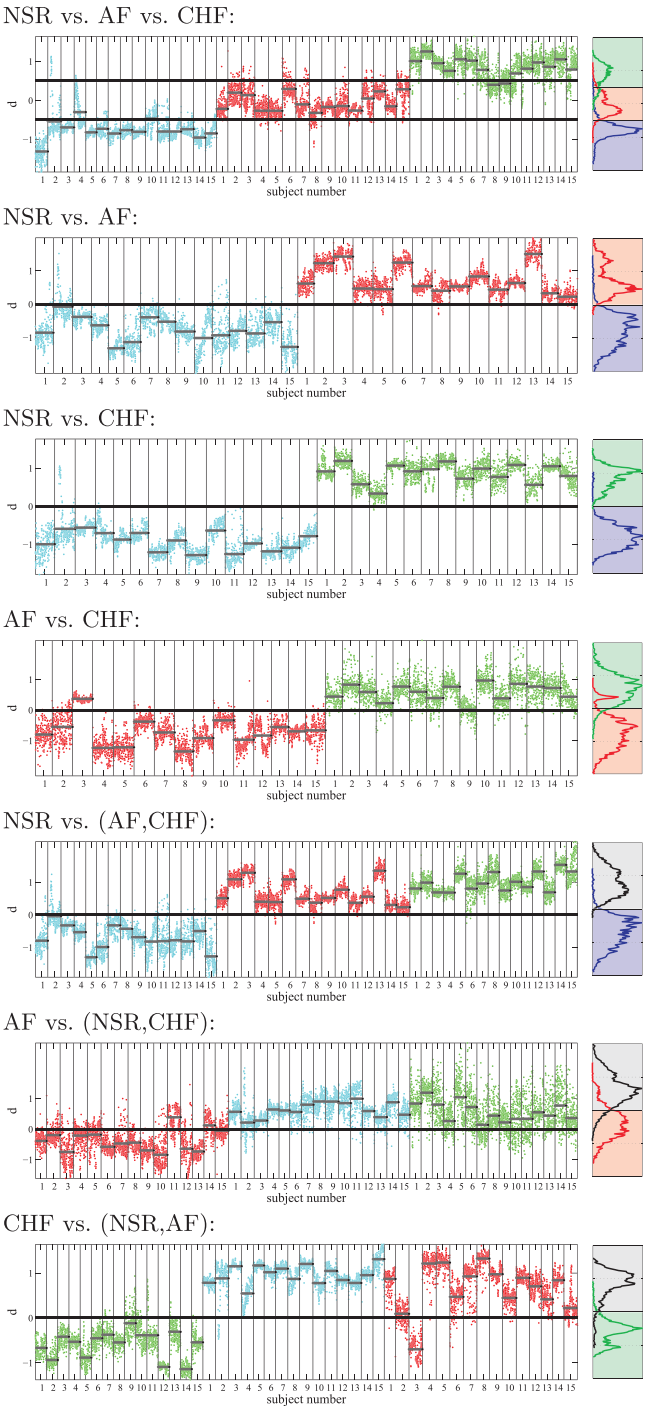

FIG. 4. Distances d from the separating hyperplanes for all 5 min data windows for all 15 subjects for the three conditions (left plots) for the best DDE models reported in Tab. IV. The mean value for each subject is shown as a black line. The histograms of these plots are shown on the right column. Blue refers to NSR, red to AF, and green to CHF. The black horizontal lines indicate the separating lines between the conditions selected by SVD. The vertical lines separate the subjects.

The model quality was also assessed by computing the area under the receiver operating characteristic (ROC) curves A' (see Fig. 5).

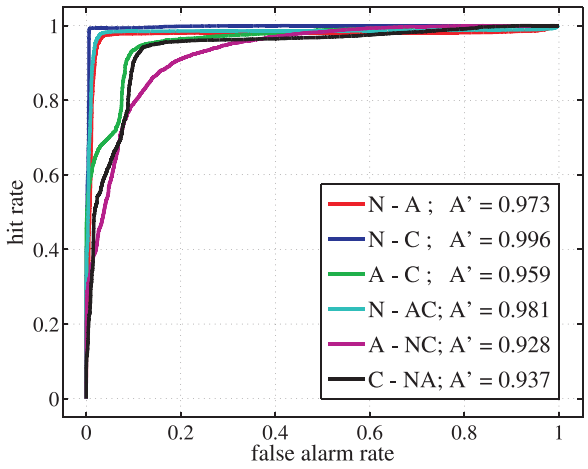

FIG. 5. Receiver operating characteristic (ROC curves) and A' (area under the ROC curve) for the classifiers NSR vs. AF, NSR vs. CHF, AF vs. CHF, NSR vs. (AF, CHF), AF vs. (NSR,CHF), and CHF vs. (NSR,AF). The abbreviations N, A, and C in the legend correspond to NSR, AF, and CHF.

A receiver operating characteristic (ROC), or simply ROC curve,[48–51] is a graphical plot which illustrates the performance of a binary classifier system as its discrimination threshold is varied. It is created by plotting the fraction of true positives out of the positives vs. the fraction of false positives out of the negatives at various threshold settings. An A' above 0.5 indicates classification performance above chance. In Fig. 5, all curves for the classifiers NSR vs. AF, NSR vs. CHF, AF vs. CHF, NSR vs. (AF,CHF), AF vs. (NSR,CHF), and CHF vs. (NSR,AF) are shown and A' is for all binary classifiers above 0.92, which is excellent. Normal heart rate is better distinguishable from the two diseases: A' is above 0.97 for NSR vs. AF, NSR vs. CHF, and NSR vs. (AF,CHF).

SVD is a simple classification tool. Performance could be easily improved by using a more sophisticated classifier such as support vector machine. Here, we wanted to emphasize on the results from the DDE analysis rather than best performance, which is already excellent.

IV. DISCUSSION

We analyzed 24 h ECG data from healthy subjects and patients with either atrial fibrillation or congestive heart failure downloaded from the PHYSIONET database.[37] These data were analyzed using delay differential equations (DDEs). First, we made a connection between DDEs in the time domain and spectral analysis tools such as Fourier analysis and higher-order statistics. We then used the outputs of DDE models with a maximum of three terms to build good classifiers for the three conditions. For 5 min data windows of the ECG data from 2 electrodes we were able to separate the three heart conditions with high accuracy.

In other studies using the same dataset, separation of the three heart conditions was only achieved using all of the data for each subject;[52–54] in comparison, we only needed 5 min of data from each condition. In our analysis, we needed non-linear terms in the DDE models to separate the three heart conditions. Non-linear methods were also better at distinguishing the data classes in Refs. 52 and 53. We found that a purely non-linear model with three non-linear terms was needed for the classification of NSR vs. AF vs. CHF, but a model with two linear and only one nonlinear term was needed for distinguishing CHF from NSR and AF in Refs. 54 and 55. This is consistent with other studies showing that CHF is a more regular heart condition.

We have thus shown that the three heart conditions are dynamically different (different DDE models are selected for the different classifiers), that the three heart conditions have different characteristic time-scales (different delays are selected for the different classifiers), and that non-linear models are needed for classification. In other studies, the same dataset was used to determine whether the heart rate is more chaotic in normal subjects, but this remains an open question.

Our analysis shows that a DDE can be considered a generic nonuniform embedding: "non-uniform" because the delays should reflect the dominant time-scales of the dynamical system and "generic" because it is a combination of a delay and a derivative embedding. The models combine functions of delayed versions of the signal with the derivative of the signal. A DDE model can unfold dynamical structures that are relevant for a single time series and the underlying dynamical system, which may be unknown as in the ECG.

ACKNOWLEDGMENTS

This work was supported by the Howard Hughes Medical Institute, NIH (grant NS040522), and the Swartz Foundation. C.L. also would like to thank Manuel Hernandez and Jonathan Weyhenmeyer for valuable discussions.

[1]G. Gouesbet, Phys. Rev. A **43**, 5321 (1991).
[2]T. Sauer, J. A. Yorke, and M. Casdagli, J. Stat. Phys. **65**, 579 (1991).
[3]B. Bezruchko and D. Smirnov, Phys. Rev. E **63**, 016207 (2000).
[4]C. Lainscsek, Phys. Rev. E **84**, 046205 (2011).
[5]R. Driver, *Ordinary and Delay Differential Equations*, Applied Mathematical Sciences (Springer-Verlag, 1977), Vol. 20.
[6]G. Clifford, F. Azuaje, and P. McSharry, *Advanced Methods and Tools for ECG Data Analysis* (Artech House, 2006).
[7]C. E. Falbo, in *Joint meeting of the Northern and Southern California sections of the MAA* (San Luis Obispo, CA, 1995).
[8]P. Whittle, *Hypothesis Testing in Time Series Analysis* (Almquist and Wicksell, 1951).
[9]P. Whittle, *Prediction and Regulation* (English Universities Press, 1963).
[10]P. Whittle, *Prediction and Regulation by Linear Least-Square Methods* (University of Minnesota Press, 1983).
[11]G. Box and G. Jenkins, *Time Series Analysis: Forecasting and Control* (Holden-Day, 1971).
[12]L. Appeltant, M. Soriano, G. Van der Sande, J. Danckaert, S. Massar, J. Dambre, B. Schrauwen, C. R. Mirasso, and I. Fischer, Nature Commun. **2**, 1 (2011).
[13]L. Larger, M. Soriano, D. Brunner, L. Appeltant, J. Gutierrez, L. Pesquera, C. R. Mirasso, and I. Fischer, Opt. Express **20**(3), 3241 (2012).
[14]Y. Paquot, F. Duport, A. Smerieri, J. Dambre, B. Schrauwen, M. Haelterman, and S. Massar, Sci. Rep. **2**, 287 (2012).
[15]R. Martinenghi, S. Rybalko, M. Jacquot, Y. Chembo, and L. Larger, Phys. Rev. Lett. **108**, 244101 (2012).
[16]Whitney, Ann. Math. **37**, 645 (1936).
[17]N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw, Phys. Rev. Lett. **45**, 712 (1980).

[18]F. Takens, *Dynamical Systems and Turbulence, Warwick 1980*, Lecture Notes in Mathematics, Vol. 898, edited by D. A. Rand and L.-S. Young (Springer Berlin/Heidelberg, 1981), pp. 366–381.
[19]K. Judd and A. Mees, Physica D **120**, 273 (1998).
[20]B. Hjorth, Electroencephalogr. Clin. Neurophysiol. **29**, 306 (1970).
[21]Y. Chan and R. Langford, IEEE Trans. Acoust., Speech, Signal Process. **30**, 689 (1982).
[22]M. Raghuveer and C. Nikias, IEEE Trans. Acoust., Speech, Signal Process. **33**, 1213 (1985).
[23]M. R. Raghuveer and C. L. Nikias, Signal Proces. **10**, 35 (1986).
[24]L. Stankovic, IEEE Trans Signal Process. **42**, 225 (1994).
[25]B. Boashash, *Higher-Order Statistical Signal Processing* (John Wiley & Sons, 1995).
[26]J. Kadtke and M. Kremliovsky, Phys. Lett. A **260**(3–4), 203 (1999).
[27]W. Press, B. Flannery, S. Teukolsky, and W. Vetterling, *Numerical Recipes in C* (Cambridge University Press, 1990).
[28]J. Tukey, in *The Collected Works of John W. Tukey*, edited by D. Brillinger (Wadsworth, Belmont, 1953), Vol. 1, pp. 165–184.
[29]A. Kolmogorov and Y. Rozanov, Theor. Probab. Appl. **5**, 204 (1960).
[30]V. Leonov and A. Shiryeav, Theor. Prob. Appl. **4**, 319 (1959).
[31]M. Rosenblatt and J. Van Ness, Ann. Math. Stat. **36**, 1120 (1965).
[32]D. Brillinger and M. Rosenblatt, in *Spectral Analysis of Time Series*, edited by B. Harris (Wiley, New York, 1967), pp. 153–188.
[33]A. Swami, see http://www.mathworks.com/matlabcentral/fileexchange/3013 for HOSA—higher order spectral analysis toolbox (2003).
[34]J. Mendel, in *Proc. IEEE* (1991), Vol. 79, pp. 278–305.
[35]J. Fackrell and S. McLaughlin, in *Proceedings of the IEE Colloquium on Higher Order Statistics* (1995), p. 9.
[36]C. Nikias and M. Raghuver, in *Proc. IEEE* (1987), Vol. 75, pp. 869–891.
[37]A. Goldberger, L. Amaral, L. Glass, J. Hausdorff, P. Ivanov, R. Mark, J. Mietus, G. Moody, C. Peng, and H. Stanley, Circulation **101**, E215 (2000).
[38]L. A. Aguirre and S. A. Billings, Int. J. Control **62**(3), 569 (1995).
[39]C. Lainscsek, C. Letellier, and I. Gorodnitsky, Phys. Lett. A **314**, 409 (2003).
[40]C. Lainscsek, P. Rowat, L. Schettino, D. Lee, D. Song, C. Letellier, and H. Poizner, Chaos **22**, 013119 (2012).
[41]See http://www.mathworks.com/help/toolbox/ident/ref/resample.html for resampling of time-domain data by decimation or interpolation.
[42]R. Kohavi, in *Proceedings of the 14th international joint conference on Artificial intelligence - Volume 2* (Morgan Kaufmann Publishers Inc., 1995), IJCAI'95, pp. 1137–1143.
[43]G. Cardillo, see http://www.mathworks.com/matlabcentral/fileexchange/15365-cohens-kappa/content/kappa.m (2009) for computation of the Cohen's kappa coefficient.
[44]W. Scott, Public Opin. Q. **19**, 321 (1955).
[45]J. Cohen, Educ. Psychol. Meas. **20**, 37 (1960).
[46]J. Fleiss and J. Cohen, Educ. Psychol. Meas. **33**, 613 (1973).
[47]R. Kohavi and F. Provost, Mach. Learn. **30**, 271 (1998).
[48]D. van Meter and D. Middleton, Transactions of the IRE Professional Group on Information Theory **4**, 119 (1954).
[49]W. W. Peterson, T. Birdsall, and W. Fox, Transactions of the IRE Professional Group on Information Theory **4**, 171 (1954).
[50]W. J. Tanner and J. Swets, Psychol. Rev. **61**, 401 (1954).
[51]C. Metz, Radiol. Phys. Technol. **1**, 2 (2008).
[52]N. Wessel, M. Riedl, and J. Kurths, Chaos **19**, 028508 (2009).
[53]U. Freitas, E. Roulin, J.-F. Muir, and C. Letellier, Chaos **19**, 028505 (2009).
[54]J. Alvarez-Ramirez, E. Rodriguez, and J. C. Echeverria, Chaos **19**, 028502 (2009).
[55]J. Hu, J. Gao, and W.-W. Tung, Chaos **19**, 028506 (2009).