Digital Signal Processing ••• (••••) ••••••

Contents lists available at ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

# Decomposing time series into deterministic and stochastic influences: A survey

Felipe S.L.G. Duarte [a], Ricardo A. Rios [b], Eduardo R. Hruschka [c,d], Rodrigo F. de Mello [a]

[a] ICMC, University of São Paulo, São Carlos, Brazil
[b] DCC, Federal University of Bahia, Salvador, Brazil
[c] Computer Engineering and Digital Systems Department, University of São Paulo, Brazil
[d] Data Science Team, Itaú Unibanco, São Paulo, Brazil

ARTICLE INFO

*Article history:*
Available online xxxx

*Keywords:*
Data decomposition
Time series analysis
Data streams
Modeling and prediction

ABSTRACT

Temporal data produced by industrial, human, and natural phenomena typically contain deterministic and stochastic influences, being the first ideally modelled using Dynamical Systems while the second is appropriately addressed using Statistical tools. Although such influences have been widely studied as individual components, specific tools are required to support their decomposition for a proper modeling and analysis. This article addresses a comprehensive survey of the main time-series decomposition strategies and their relative performances in different application domains. The following strategies are discussed: i) Fourier Transform, ii) Wavelet transforms, iii) Moving Average, iv) Singular Spectrum Analysis, v) Lazy, vi) GHKSS, and vii) other approaches based on the Empirical Mode Decomposition method. In order to assess these strategies, we employ diverse and complementary performance measures: i) Mean Absolute Error, Mean Squared and Root Mean Squared Errors; ii) Minkowski Distances; iii) Complexity-Invariant Distance; iv) Pearson correlation; v) Mean Distance from the Diagonal Line; and vi) Mean Distance from Attractors. Each decomposition strategy is better devoted to particular scenarios, however, without any previous knowledge on data, GHKSS confirmed to work as a fair and general baseline besides its time complexity.



## 1. Introduction

Time series analysis has long been employed for the organization, modeling, and prediction of data observations in different application domains [1], as stock market investment [2,3], anomaly detection [4], climate studies [5,6], and biological modeling [7]. More recently, Machine Learning has reinforced the relevance of an analysis, since the representation of time series is essential for problems that involve data stream modeling and concept drift detection [8–10].

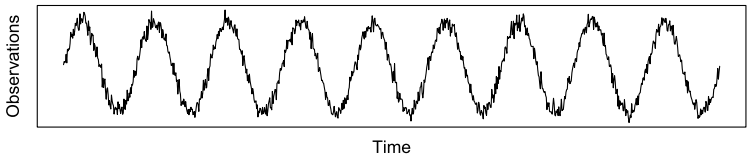
Times series analysis has its roots in two different areas of science, namely Statistics and Dynamical Systems. The former provides a wide variety of modeling strategies for time series forecasting, through the use of the so-called ARIMA models [11], which assume data observations originate from a mixture of deterministic and random variables. In Dynamical Systems, most methods are based on the search for deterministic behaviors, in which a next observation is modelled using a combination of past elements without noise [12].

Since statistical strategies are mostly indicated for the modeling of stochastic time series, whereas dynamical system tools better tackle deterministic ones, time series decomposition approaches have been proposed and studied in order to separate both components, thereby allowing each of them to be modeled according to the best set of tools [13]. Such decomposition tools have been employed under different names, as noise filtering and reduction [14], smoothing filtering [15], denoising [16], blind source separation [17], and feature extraction [18].

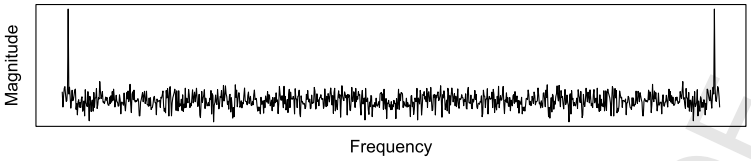
Therefore, taking into account the importance of decomposition approaches to model time series, the main contributions of this survey are two-fold: it summarizes the main time series decomposition strategies and it provides a comprehensive comparative analysis of such strategies. More precisely, we discuss: i) Fourier transforms; ii) Wavelet transforms; iii) Moving Averages; iv) Singular Spectrum Analysis; v) Lazy methods; vi) GHKSS; and vii) other approaches based on the Empirical Mode Decomposition method. In order to assess the relative merits of each method, we empirically compare them by using diverse and complementary crite-

*E-mail addresses:* fgduarte@icmc.usp.br (F.S.L.G. Duarte), ricardoar@ufba.br (R.A. Rios), hruschka@usp.br (E.R. Hruschka), mello@icmc.usp.br (R.F. de Mello).

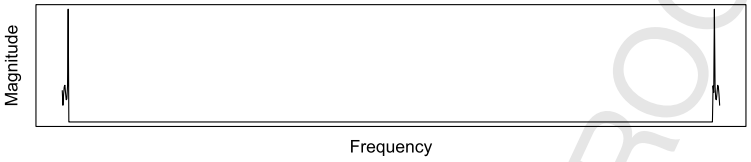
https://doi.org/10.1016/j.dsp.2019.102582
1051-2004/



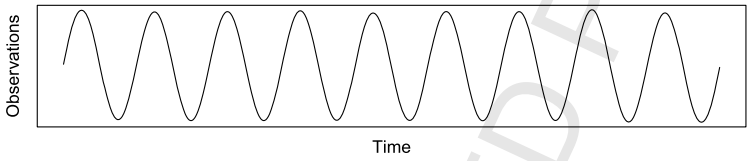

(a) Time series with a sinusoidal deterministic component added to Normal stochastic data following $\mathcal{N}(0, 0.01)$.




(b) Magnitudes of complex coefficients.




(c) Less representative complex coefficients are zeroed.




(d) Complex coefficients are taken back to the time domain. A comparison with (a) the stochastic component has been removed.

**Fig. 1.** Decomposition process using FT. In (a), a time series is composed of a sinusoidal deterministic signal added to a stochastic process following Normal distribution $\mathcal{N}(0, 0.01)$. (b) shows the magnitudes of complex coefficients. In (c), the less representative complex coefficients (at lower magnitudes) were zeroed for the removal of influences of the stochastic component from data. In (d), the series is taken back to the time domain, in which the stochastic component is no longer present.

ria: i) Mean Absolute Error, Mean Squared Error and Root Mean Squared Error; ii) Minkowski Distances; iii) Complexity-Invariant Distance; iv) Pearson correlation; v) Mean Distance from the Diagonal Line; and vi) Mean Distance from Attractors.

Experiments were run on time series data for illustrating the biases of the assessed methods, and their advantages and disadvantages for particular application scenarios. The achieved outcomes are helpful for the practitioner interested in choosing the decomposition strategy that is likely to be more suitable for a given problem in hand.

The remainder of the paper is organized as follows: Section 2 describes state-of-the-art decomposition techniques; Section 3 addresses the criteria used for the evaluation of the surveyed decomposition techniques; Section 4 reports the experiments performed and the obtained results; Finally, Section 6 is devoted to concluding remarks.

## 2. Decomposition approaches

Decomposition strategies have been designed by taking advantage of specific biases; Therefore, the quality of the practical results is problem-dependent [19,5,12]. If a decomposition approach is randomly selected with no knowledge of their biases, sub-optimal results and poor conclusions may be drawn [20]. This section reports a detailed study on the most common decomposition methods that deal with different application domains. Towards a better understanding of such techniques, we organized them into four groups namely (i) Transformation-based methods, (ii) Subspace-based methods, (iii) Phase Space-based methods, and (iv) Time Domain-based methods.

### 2.1. Transformation-based methods

This section analyzes strategies to extract stochastic and deterministic influences, which were developed on top of the Fourier and Wavelet transforms.

#### 2.1.1. Fourier transform

Fourier Transform (FT) decomposes a time series into a sum of sinusoidal functions [21], an ability that has been explored in several application domains, including the separation of deterministic and stochastic components that occur at different frequency intervals. In real-world applications, researchers employ the discrete version of FT, producing a set of complex coefficients based on the input signal [22]. The coefficients correspond to sinusoidal functions following Euler's formula, as defined in Equation (1), where $j$ is the frequency associated with the sinusoidal component.

$$f_j = \frac{1}{n}\sum_{k=0}^{n-1} x_k e^{-\frac{2\pi i}{n} jk} \qquad j = 0, \ldots, n-1 \tag{1}$$

A separation of stochastic and deterministic series typically assumes complex coefficients with greater magnitudes are associated with the deterministic component, since they should reflect the most relevant data elements. Therefore, coefficients at lower magnitudes are zeroed and removed from the reconstructed deterministic signal [23]. Next, it is subtracted from the original time series to obtain the stochastic component.

Towards an illustration of the FT decomposition, consider the time series shown in Fig. 1(a), which is a combination of a deterministic sinusoidal series and stochastic data following a Normal

distribution function $\mathcal{N}(0, 0.01)$. After the computation of complex coefficients, their magnitudes are plotted along frequencies, as shown in Fig. 1(b), where the greatest magnitude is assumed to correspond to the deterministic series. A plot is symmetrical in relation to its central point (due to the FT properties) and, therefore, shows two peaks.

After the zeroing of all coefficients whose magnitudes are less relevant (Fig. 1(c)) and application of the inverse FT transform, a result similar to Fig. 1(d) is obtained, thus confirming the stochastic component has been removed. However, the claim that the stochastic component is associated with the complex coefficients at the lowest magnitudes is not necessarily true [24]. The stochastic component is obtained after the subtraction of the deterministic signal from the original series.

FT works very well in this simple scenario. However, it faces issues when tackling non-periodic and non-stationary data. This issue was first tackled by Gabor [25], who applied FT in a short windows — also known as Short Time Fourier Transform (STFT) — of the time series and empirically showed that this approach works for signals changing their behavior over time. An important limitation remained due to the sinusoidal bias, which strongly affects the expected deterministic components produced from dynamical systems [13]. For instance, by decomposing a time series formed with the Logistic map plus some noise, FT significantly affects the expected phase space, which should correspond to a negative (concave) second-order polynomial function. In fact, a behavior changes the phase space for resembling a sinusoidal time series, clearly represented by ellipsoids [26,27]. As a result, a next series observation becomes dependent on different past values, thus producing an anomalous mathematical topology and affecting forecasting tasks. Moreover, it does not locate frequencies along time, but only informs a sum of sinusoidal components results in the original time series, a lack of information hampers the extraction of features when frequency variations are decisive for decision making.

#### 2.1.2. Wavelet transform

The Wavelet Transform (WT) [28,29] was designed to improve the results from FT and locate frequencies along time. It employs an infinite set of basis functions $\psi(.)$, referred to as wavelets, which must hold the properties defined in Equations (2) and (3). Such properties ensure the integral must be zero and the energy must be bounded to one, which implies the wavelet must oscillate around the time axis (first property) and the energy must be preserved (second property). Moreover, wavelets must be in set $\mathcal{L}^2(\mathbb{R})$ of the integrable squared functions.

$$\int_{-\infty}^{\infty} \psi(t)dt = 0 \tag{2}$$

$$\int_{-\infty}^{\infty} \psi(t)^2 dt = 1 \tag{3}$$

A finite time range $[-T, T]$ at the real line is assumed to contain the wavelet, i.e., Equation (4) must hold $0 \leq \epsilon \leq 1$, when $\epsilon$ is the function energy outside range $[-T, T]$. Therefore, a wavelet is typically seen as an increasing-decreasing function at a limited time period [30]. An energy concentration provides the frequency location in comparison to FT. As an example, let us consider a wavelet formed by the second-order derivative of Gaussian function $\psi(t) = (1 - 2t^2)e^{-t^2}$ (Fig. 2), which concentrates all its energy at range $[-5, 5]$. As a consequence, $\epsilon = \int_{-\infty}^{5} \psi(t)^2 dt + \int_{5}^{\infty} \psi(t)^2 dt = 0$.

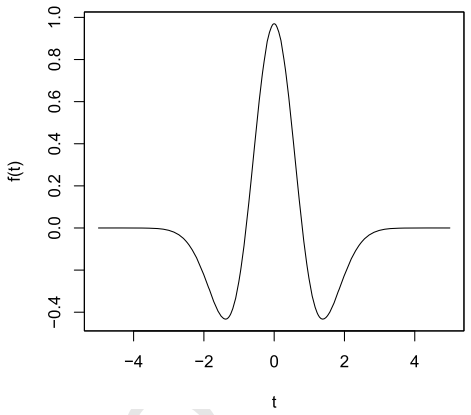



**Fig. 2.** Wavelet formed after the second-order derivative of the Gaussian function.

$$\int_{-T}^{T} \psi(t)^2 dt > 1 - \epsilon \tag{4}$$

In time series decomposition, WT is performed on different scales for the extraction of components at different frequencies. The first scale results from a full execution of WT on the time series for the production of two sub-series with half elements of the original. One half contains an approximation to the original series, while the other contains details typically associated to high frequencies. The same process can be applied for the extraction of further components at lower frequencies. The process can be compared to a filter bank that extracts features from the highest to the lowest frequencies.

The most well-known and, therefore, typically used WT is Haar [31], which simply computes sums and differences between consecutive time series observations. It can be applied along scales for studies on frequencies and their separation towards a composition of deterministic and stochastic components. Most commonly, studies refer to high frequencies as stochastic series, while the lowest ones are considered deterministic, which may not be the real situation. For example, a system can produce some series collected at 200 Hz, in which the most important frequency to compose the deterministic behavior may be from 40 to 100 Hz, while all frequencies smaller than 40 Hz may be associated with the stochastic component. This is the case in which researchers study the Gamma brain waves towards modeling healthy cognitive functions, such as memory and learning [32,33].

Although WT incorporates additional features to FT, it also distorts phase spaces by producing ellipsoids after the time series decomposition, which jeopardizes the time relationship among observations strongly influencing modeling and forecasting. The mathematical topology of the expected deterministic component is again changed, and forms what is expected from sinusoidal series (ellipsoids of a single one-dimensional void). Any undesirable change in the phase space reflects the bias imposed by those series decompositions. Consequently, techniques must provide the expected deterministic phase spaces [34].

### 2.2. Subspace-based methods

This section describes a method widely adopted in real-world scenarios and developed to perform decomposition taking into account advantages of sub-space learning. In summary, it extracts stochastic influences reducing time series dimensionality.

#### 2.2.1. Singular spectrum analysis

The Singular Spectrum Analysis (SSA) is a nonparametric strategy for the decomposition of time series into linear independent

components [35]. It assumes series are linear and stationary, and any additional noise follows a Normal distribution. However, the practice has confirmed SSA still provides fair results even in the presence of nonstationary data [36,37].

SSA is divided into two stages namely decomposition and reconstruction. The former immerses the time series into a $K$-dimensional vectorial space, which is exactly the same as a phase space using Takens' immersion theorem [38] with embedding dimension $m = K$ and time delay $\tau = 1$. For instance, let us consider time series $X = \{x(0), \ldots, x(n)\}$, where $n$ is the number of observations or time index. A new multidimensional series $Y \in \mathbb{R}^K$ is produced as follows:

$$Y_n = (Y_1, Y_2, \ldots, Y_K) = \begin{pmatrix} x(1) & x(2) & \ldots & x(K) \\ x(2) & x(3) & \ldots & x(K+1) \\ \vdots & \vdots & \ddots & \vdots \\ x(L) & x(L+1) & \ldots & x(n) \end{pmatrix},$$

where $L$ is a parameter set by the user. This forms a Hankel matrix [39], i.e., all elements $Y_{i,j}$ are equal when $i + j = c$ where $c$ as a constant. Note that $Y_{1,4} = Y_{4,1} = Y_{2,3} = Y_{3,2} = x(4)$, because $i + j = 5$. Series $Y$ is then multiplied by its transposition for the obtaining of a covariance matrix $S = YY^T$. Next, the Singular Value Decomposition (SVD) is applied on $S$, decomposing it as $S = U\Sigma V^*$, where $U$ is a changing basis, $\Sigma$ is a diagonal matrix (as the eigenvalues matrix) and $V^*$ is the conjugate transposed matrix of $V$ to return back to the original vectorial space.

SVD produces a basis in space $\mathcal{R}^L$, representing relationships among vectors in the row space of $Y$, such as those captured by Sub-space learning techniques (SL) [40]. Greater variances imply higher relevances for dimensions associated with row vectors. Therefore, the eigenvalues in $\Sigma$ are connected to the rows identified with the widest spread sub-series along the space of covariances. The separation of those most relevant eigenvalues leads to the reconstruction components of greater energy along time, which is the second stage of the SSA technique. And zeroes part of the eigenvalues to reconstruct a version of $Y$, whose components at a given energy range are maintained, while the others are removed. Therefore, energy is considered a criterion for the association of an element with either the deterministic, or the stochastic component.

SSA-based approaches face difficulties for large values of $K$, because that implies a long sub-series over time, producing a lot of variances in the space $S$ analyzed with SVD. Under a circumstance, very similar eigenvalues would be obtained. If $K$ is too small, a less representative scenario because eigenvalues would be too small and no conclusion could be drawn. Therefore, $K$ must be set in some adequate range, which certainly depends on the target problem. Parameter $L$ also imposes limitations, since few or many dimensions may compose the space formed by $S$. Few dimensions would impact in terms of removing few eigenvalues, while so many would cause the opposite issue, i.e., the removal of more elements than necessary.

SSA inherits some limitations from SVD, which assumes the space formed by $S$ is provided by a linear combination of eigenvectors. This may impact the deterministic components obtained from the reconstruction of the most relevant eigenvectors (due to the importance of eigenvalues). The phase spaces of such deterministic series may also be undesirably changed, as for FT and WT. In this case, $K$ and $L$ influence the obtained phase spaces. Therefore, detailed analyses must be performed in advance, towards setting set up these parameters.

### 2.3. Phase-space-based methods

This section describes methods that decompose time series after unfolding their relationships into high-dimensional spaces. Their main advantage is the possibility of modeling the deterministic influences present in nonlinear chaotic systems.

#### 2.3.1. Lazy

In order to reduce the additive noise present in time series, Schreiber proposed a nonlinear decomposition method, later referred to as Lazy [41], which works on top of phase spaces. The construction of a type of space relies on Takens' immersion theorem [38] which employs an embedding dimension $m$ and a time-delay separation $\tau$ to construct a time series $X = \{x_{-\infty}, \ldots, x_t, \ldots, x_{-1}, x_0\}$ as follows:

$$s_t = (x_t, x_{t-\tau}, \ldots, x_{t-(m-1)\tau}),$$

where $s_t$ are the states of an attractor $\mathcal{S}$ in an Euclidean space referred to as phase space. Such states represent the relationships/dependencies among time series observations.

The decomposition strategy is based on the fact that the noise changes the position of each state of the attractor, but preserves its topology, i.e., at the end, the shape of the attractor is preserved (see Fig. 3). The series observation relations, tendencies, and recurrences are also unfolded in this space, preserving the deterministic bias.

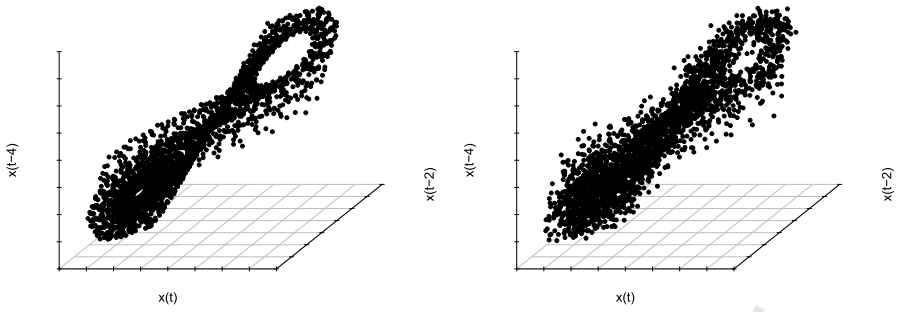
After reconstruction, Lazy considers a simple approach to separate noise from data. It moves a state $s_t$ to the average value of its nearby neighbors according to an open ball of radius $\epsilon$, producing nonlinear data modifications on time series. First, it calculates Euclidean distance matrix $\mathcal{D}$, such that each element is given by $\mathcal{D}_{ij} = d(s_i, s_j)$ for $i = 1, \ldots, n$ and $j = 1, \ldots, n$, i.e., the matrix gives the distance of each state $s_i$ to the others. The identification of the $k$ nearest neighbors of each state $s_i$ inside the open ball of radius $\epsilon$. Iteratively, the position of each state is modified according to an inverse weighted average of the position of its neighbors [42], i.e., a closer state more strongly influences the new position, rather than a distant one. The entire process is repeated until the stop criterion is reached.

All phase space states are then mapped back to the time domain for the obtaining of a signal referred to as deterministic. Its subtraction from the original time series yields a stochastic series. The main advantage of Lazy is the possibility of denoising time series by phase spaces. Instead of decomposing a time series by only considering its temporal behavior, as usually performed by the aforementioned strategies, Lazy unfolds the time series to ensure the expected space pattern is preserved in the resulting attractor after the removal of noise.

#### 2.3.2. Locally projective nonlinear decomposition methods

Locally projective nonlinear decomposition methods [43] assume data are produced by low-dimensional dynamical systems added to random or high-dimensional noise processes. A time series can be reconstructed in some over-embedded phase space, i.e., where embedding dimension $m$ is greater than necessary. Therefore, the deterministic data can lie on a low-dimensional manifold contained in such high-dimensional space, which motivates studies on the noise projection onto manifolds [44,45]. Those orthogonal projections depend on the metric used, since the Euclidean is the mostly used one, but not necessarily the best for all scenarios. The following three different local projection algorithms have been commonly used: GHKSS, PROJECT, and NOISE.[1]

[1] All of them are available with the Tisean software [46].




(a) (b)

**Fig. 3.** In (a), the attractor of Lorenz system with shape that reminds to the Infinity symbol. In (b), the same attractor but now with noise added. The addition of noise changes the position of each state, but preservers the topology of the attractor.

GHKSS and PROJECT depend on the full dataset allowing its use with data streams, while NOISE operates in an online fashion. GHKSS assumes a strictly linear manifold over which noise is white and has the same probability in all directions. A process performed makes noise equal to zero by summing them up. If the manifold is curved, a previous stage is locally applied to linearize it. PROJECT uses a different strategy which builds up local coordinate systems in regions of the manifold to project states according to the average of nearest neighbors.

Modifications in the previous methods have been proposed to enable the online noise reduction [47] to tackle the nonlinear projective filtering of data streams. In such scenario, past time series observations are considered to adapt next observations in order to separate the deterministic from the stochastic component. NOISE is a representative algorithm that considers a neighborhood search strategy in phase space simply based on past observations.

### 2.4. Time domain-based methods

The aforementioned methods are characterized by an important step that transforms the time series before decomposing their observations. This section details a set of methods that directly extract stochastic and deterministic components along the time domain.

#### 2.4.1. Moving average

Statistical techniques have been widely adopted to decompose time series into stochastic and deterministic components [48,49]. In general, such unsupervised techniques apply linear filters to smooth noisy influences contained in time series. One of the most used linear filters is the moving average (MA) model [50], which smooth observations by averaging them according to an order parameter that represents seasonal periods.

Given the known behavior and results of ARIMA models, a common questioning refers to its use in the context of series decomposition. However, those models represent the phenomenon itself and noisy influences operating on it. In this sense, we assume the phenomenon produces deterministic influences, while the noise is associated to the stochastic behavior. Since the Auto-Correlation term (AR) is typically employed to model the phenomenon, Integration (I) is used to remove trends, and the Moving Average (MA) associated with noise. From this, we observe the need of only using a modeling strategy capable of dealing with the time series component mostly connected to the stochastic influences, which is typically the noise. If the other modeling terms (specifically AR and I) are used, the dynamical system attractor suffers influences and looses its typical shape related to the physical phenomenon of interest [13].

$$x_t = \varepsilon_{t-1} + \theta_1 \cdot \varepsilon_{t-1} + \theta_2 \cdot \varepsilon_{t-2} + \cdots + \theta_q \cdot \varepsilon_{t-q} \tag{5}$$

To be used as a filter, an MA model can be simplified by Equation (6), which considers only the average of past observations and disregard the influence of noisy components. The equation also promotes the smoothening of the stochastic component, leaving a trend-cycle component. Once this component is considered deterministic, by subtracting it from a time series with additive noise, it is possible to find out the stochastic one.

$$y_t = \frac{1}{q}\left(\sum_{i=0}^{q/2} x_{t-i} + \sum_{i=1}^{q/2} x_{t+i}\right) \tag{6}$$

MA is also referred to as a convolution filter [51], and in order to exemplify its use, let us consider a time series generated by a sinusoidal function with an angular frequency $\omega = \pi$. Some noise produced by a Normal probability distribution with mean and standard deviation of 0 and 0.2, respectively, is added the resulting noisy series corresponds to the red line in Figs. 4 (a) and (b). The black lines correspond to the time series obtained after removal of noise setting as $q = 1$ and $q = 100$ in Equation (6), respectively. As shown in the figure, the greater the value of $q$, the smoother the time series. However, the resulting series has fewer $q$ observations than the original one.

Autoregressive Models are also used as filters and have been commonly referred to as recursive operators [11]. A type of technique was specially designed to smooth linear series [51].

#### 2.4.2. EMD methods

The Empirical Mode Decomposition (EMD) method was designed to convert a time series into a set of monocomponents, called Intrinsic Mode Functions (IMFs), for supporting the study of instantaneous frequencies and amplitudes by the Hilbert Spectrum Analysis (HSA) [27]. Its main advantage is it decomposes a time series regardless of its linearity, stationarity and/or stochasticity.

The key step to perform EMD is the sifting process, summarized in Fig. 5, which initially analyzes a time series $x(t)$ to identify local maxima (filled dots) and minima (empty dots) along time (Step 1 – Fig. 5(a)). All maxima compose an upper envelope $l(t)$ by the cubic spline method (Step 2 – Fig. 5(b)). The same procedure is applied, so that all minima form a lower envelope $u(t)$ (Step 3 – Fig. 5(c)). The envelopes are then used to compute the mean envelope (Equation (7)), illustrated in Fig. 5(d) (Step 4).

$$m(t) = \frac{l(t) + u(t)}{2} \tag{7}$$

Later, $m(t)$ is removed from the original time series $x(t)$, which produces the first monocomponent candidate $h_{1,1}(t) = x(t) - m(t)$, in which the first subindex corresponds to the IMF identifier and the second refers to the candidate identifier. $h_{1,1}(t)$ is then used in place of the original time series and the sifting process is repeated

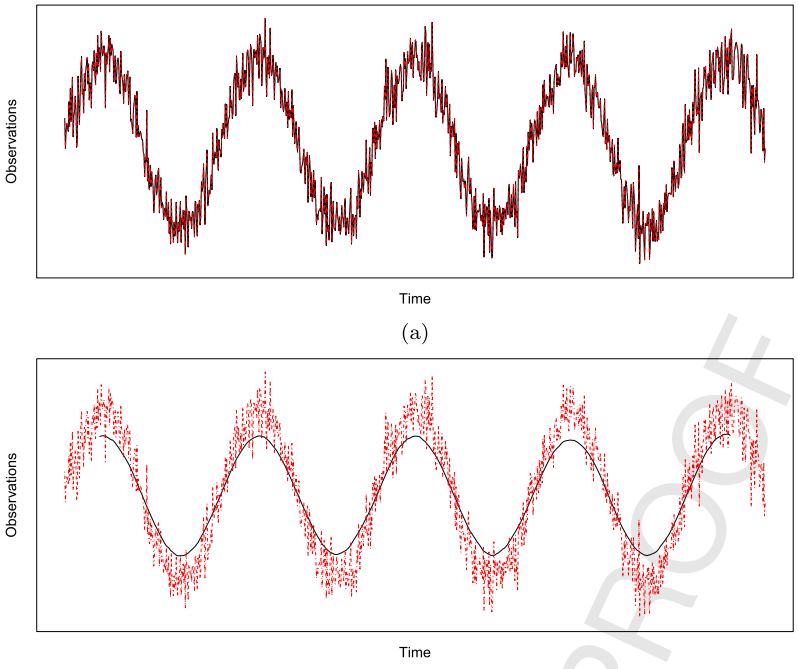



**Fig. 4.** MA filter applied to denoise a sinusoidal time series with trend. Figures (a) and (b) were produced using $q = 1$ and $q = 100$ (Equation (6)), respectively.

until the candidate satisfies the IMF monocomponent definition, which must agree with one of the following stop criteria: i) both number of extrema and number of zero-crossings must either be equal, or differ by one at most; or ii) for every $t$, $m(t)$ is zero. After obtaining the first candidate satisfying the IMF definition, the first IMF is obtained according to $h_1(t) = h_{1,k}(t)$, assuming $k$ candidates were produced until reaching a stop criterion.

This first IMF is then removed from the data in the $x(t) - h_1(t)$ form, and the resulting series is again analyzed by the whole process, thus producing IMFs until reaching a final stop criterion, i.e., when the last IMF becomes a monotonic function, from which no further IMF can be extracted from. The last component is referred to as final residue $r(t)$. In summary, according to EMD, a time series $x(t)$ is composed of a set of IMF monocomponents plus a residue, as shown in Equation (8).

$$x(t) = \sum_{n=1}^{N} h_n(t) + r(t) \tag{8}$$

Due to the sifting process, EMD acts as a dyadic filter bank, extracting IMFs with different frequency bandwidths [52]. Based on this assumption, Rios et al. [53] used the Nyquist-Shannon sampling theorem [54] to prove the maximum frequency of the next extracted IMF would be lower than the previous one, i.e., IMFs are obtained from high to low frequencies. This proof has confirmed the possibility of separating IMFs ($h_n(t)$) into two classes: one with higher-frequency components and another containing lower-frequency ones. As typically considered in the signal processing area, the first class corresponds to stochastic influences, whereas the second represents deterministic ones (which does not necessarily hold for any target problem). A proof was the basis for the development of EMD-RP and EMD-MI, two new approaches for the decomposition of time series into stochastic and deterministic components.

EMD-RP combines the Empirical Mode Decomposition method (EMD) with Recurrence Plots (RP) [19]. RP analyzes a time series by reconstructing it into a multidimensional space, also referred to as phase space or time-delay coordinate space, which maps the relationships among observations [55,56]. After a reconstruction, the relationships are organized into a two-dimensional binary matrix, called Recurrence Matrix, as defined by Equation (9), where $\varepsilon$ is a distance threshold, $||\cdot||$ is a norm that calculates the distance between observations, and $\Theta(\cdot)$ is a Heaviside step function, defined in Equation (10) [56].

$$R_{i,j} = \Theta(\varepsilon - ||\vec{x}_i - \vec{x}_j||) \tag{9}$$

$$\Theta(\alpha) = \begin{cases} 0, & \alpha < 0 \\ 1, & \alpha \geq 0 \end{cases} \tag{10}$$

The Recurrence Matrix is plotted in different colors according to the binary entries [56], e.g., black dots when $R_{i,j} = 1$ and white ones when $R_{i,j} = 0$. The visual structures provide information about the time series under study such as i) isolated points mean system states are rarely repeated, i.e., the time series is highly stochastic, and ii) diagonal lines correspond to a persistent behavior, i.e., the determinism rate of a time series is directly associated with the number and length of diagonal lines. Aiming at quantifying the information present in such structures, several measurements, called Recurrence Quantification Analysis (RQA), were developed by Zbilut and Webber Jr. [57,58], such as DET to estimate the determinism rate of time series. DET is defined in Equation (11), where $P(\varepsilon, l)$ is the frequency of diagonal lines of length $l$ in the RP structures calculated by Equation (12).

$$DET = \frac{\sum_{l=l_{min}}^{N} lP(\varepsilon, l)}{\sum_{i=1}^{N}\sum_{j=1}^{N} R_{i,j}, \forall i \neq j} \tag{11}$$

$$P(\varepsilon, l) = \sum_{i,j=1}^{N} (1 - R_{i-1,j-1(\varepsilon)})(1 - R_{i+l,j+l(\varepsilon)}) \prod_{k=0}^{l-1} R_{i+k,j+k(\varepsilon)} \tag{12}$$

Since the determinism rate provided by RQA, EMD-RP starts applying EMD on the time series, thus producing a set of IMFs plus

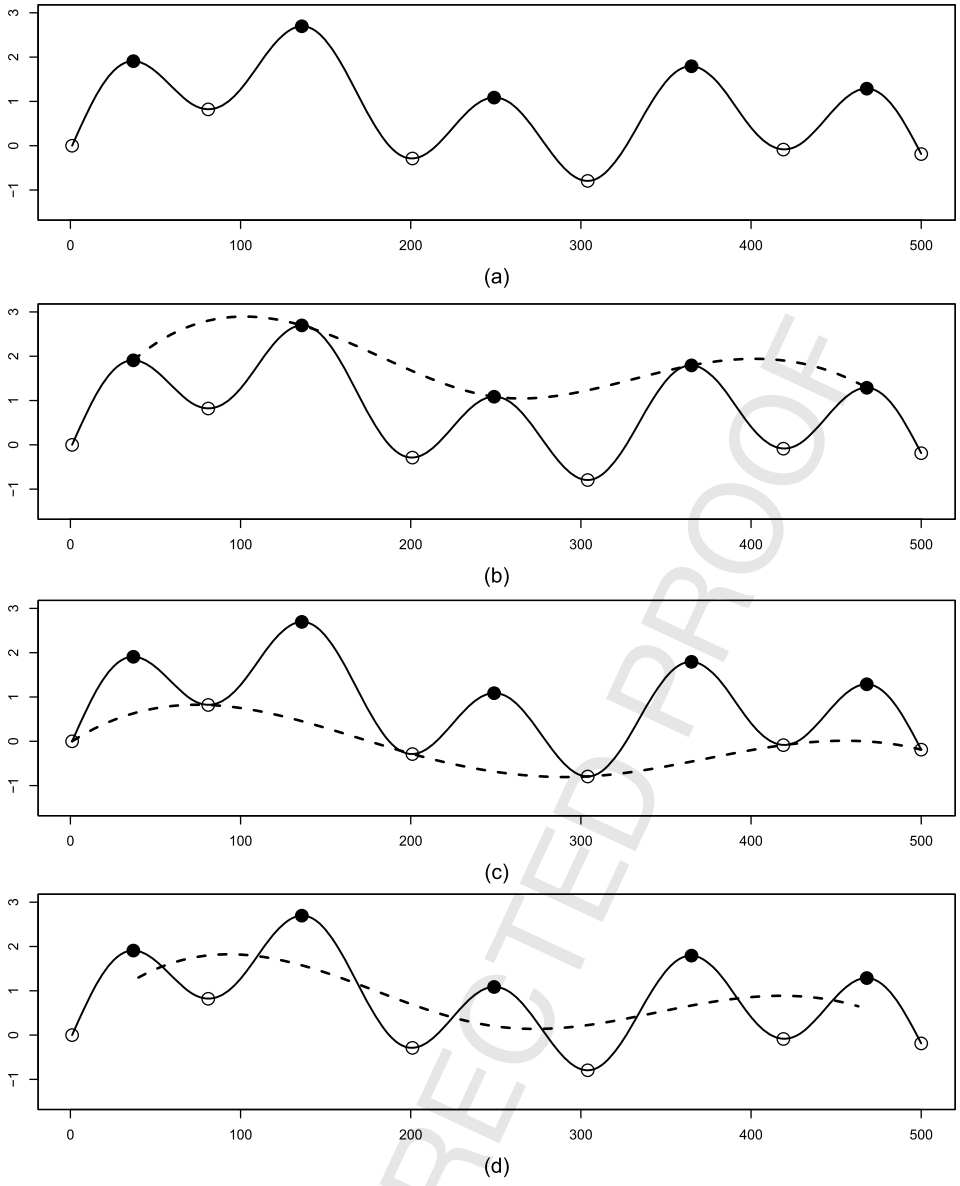



**Fig. 5.** Original time series is represented by solid lines. In (a), maxima and minima observations are selected. In (b, c), the cubic spline method is applied to the maxima and minima, thus connecting all observations and producing the upper and lower envelopes, respectively. In (d), the mean envelope is obtained by calculation of the mean values along the upper and lower envelopes.

a residue. The determinism level for each IMF is evaluated by RQA and then compared against a lower bound that defines whether an IMF is deterministic. If the level is higher than a threshold, the IMF is considered deterministic, otherwise it is stochastic. Finally, all stochastic IMFs are summed up to form the stochastic component. Similarly, the deterministic component is obtained by the addition of all deterministic IMFs in conjunction with the series residue ($r(t)$). Although the residue is not considered an IMF, its influence on the time series is relevant, since it is related to the trend component. Consequently, by adding all IMFs, disregarding the residue, the original time series can be detrended, i.e., one can obtain a nonstationary version of it.

All stages involved in EMD-RP are summarized in Fig. 6. At first we have the series from Fig. 4, which was produced after adding a white noise process with zero mean and standard deviation equals to 0.2 to a sinusoidal function. From that, we apply EMD, producing 4 IMFs as shown in Figs. 6(a,c,e,g). The Recurrence Plots for those IMFs are presented in Figs. 6(b,d,f,h), respectively. Note that the IMFs from 1 to 3 contain isolated points, which are associated to stochastic behavior. However, the fourth is closer to the expected sinusoidal function, presenting a strong recurrent (deterministic) pattern. By applying the DET measure (Equation (11)) on this last IMF (Fig. 6(h)), we obtain a result greater than 0.95,[2] suggesting determinism.

EMD-MI also decomposes time series into stochastic and deterministic components. However, it combines EMD with the Mutual Information (MI) method to automatize the process [53,26] which supports the estimation of shared information between variables. The MI method employed by EMD-MI was proposed by Darbellay and Vajda [59]. It is based on the partitioning of data space into a finite number of non-overlapping rectangular cells for the mapping of variable dependencies. Instead of using fixed-width bins, this version of MI continuously partitions the data space until reaching a conditional independence among cells. Therefore, the mutual in-

[2] An empirical analysis was already performed as presented in [19], confirming a threshold equals to or greater than 0.95 is enough to suggest determinism.

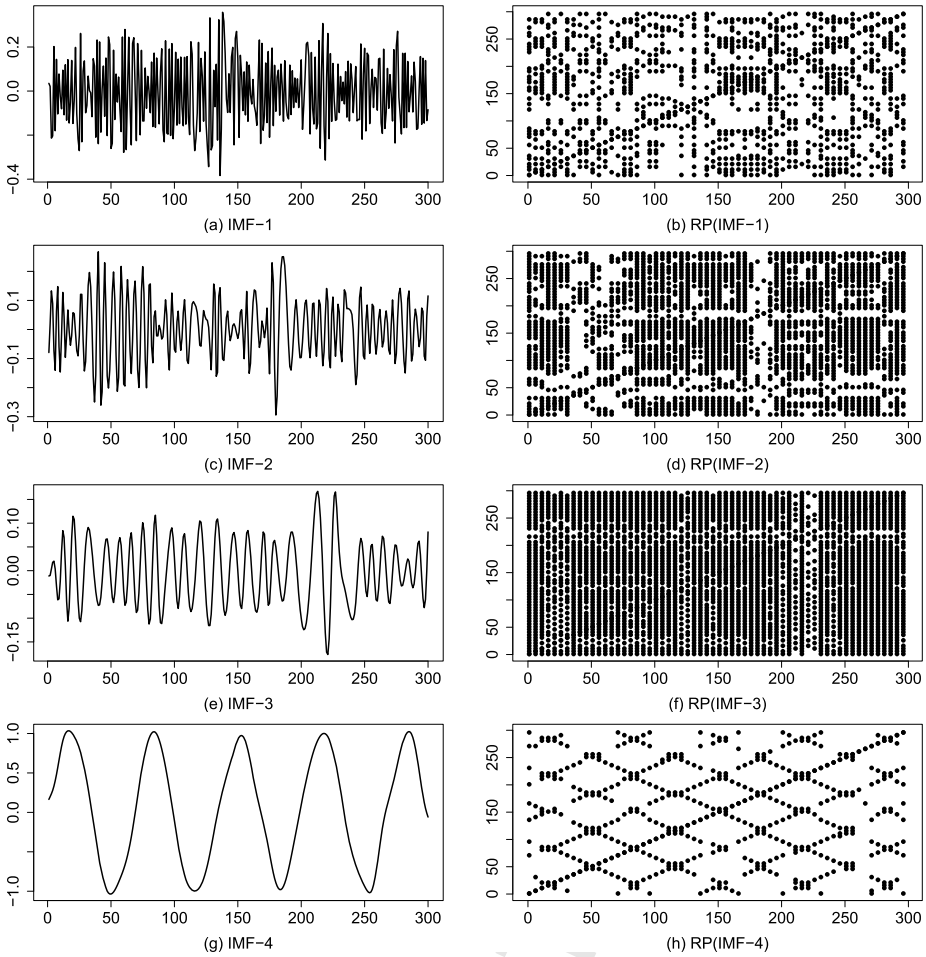



**Fig. 6.** (a), (c), (e), and (g) show the IMFs extracted after the application of EMD on the noisy time series from Fig. 4(a). (b), (d), (f), and (h) display the RP structures for each IMF. The determinism rate increases as new IMFs are extracted.

formation is computed on top of such partitions and provides an estimation closer to its theoretical value.

EMD-MI follows two strategies: i) it computes the mutual information between consecutive pairs of IMFs to find a cutoff point that hopefully separates stochastic from deterministic IMFs in terms of the MI average value and ii) it deals with time series which supports no detection of different temporal patterns from IMFs and clearly suggests a cutoff point. In such situation, Darbellay-Vajda's MI (DV) is applied on the frequency domain of Fourier complex coefficients. Fourier Transform $\mathcal{F}(\cdot)$ is performed in every extracted IMF and, since the residue contains only the series trend, it is expected to be deterministic. A set of complex coefficients $C_n(t)$ is obtained by (13) for every IMF $h_n(t)$.

$$C_n(t) = \mathcal{F}(h_n(t)) \tag{13}$$

Therefore, the output produced by $\mathcal{F}(h_n(t))$ is a set of complex coefficients $C_n(t) = \{c_{n,1}, c_{n,2}, \ldots, c_{n,k}, \ldots, c_{n,T}\}$, such that $c_{n,k}$ is calculated by the Equation (14).

$$c_{n,k} = \sum_{t=1}^{T} h_n(t) \cdot e^{-i2\pi \frac{k}{T} t}, \quad \forall k \in \{1, 2, \ldots, T\} \tag{14}$$

$$\theta_n(t) = \arctan\left(\frac{\Im(C_n(t)))}{\Re(C_n(t)))}\right), \quad \forall n \in \{1, 2, \ldots, N\} \tag{15}$$

The phase components are found after the application of the arc-tangent function on the ratio between the imaginary $\Im(\cdot)$ and real $\Re(\cdot)$ parts of complex coefficients (see Equation (15)). Similarly to the first strategy, after the computation of MI on every pair of consecutive IMF phases, the resulting values are stored in a vector. By comparing two stochastic IMFs, the obtained Mutual Information is close to zero when their coupling is weak. On the other hand, the comparison between deterministic IMFs tends to provide greater MI values. By considering the mean value ($\zeta$) for vector **v**, we use the abrupt change of MI values to define the cutoff point and separate the stochastic components (first set of IMFs) from the deterministic (second set of IMFs) ones.

EMD-MI might be considered nonparametric, as the cutoff point is directly estimated from data. However, when no time constraint is provided for the analyses of series, we suggest the application of the EMD-RP approach — since its cutoff point usually provides better results to support the time series decomposition steps.

The stages of EMD-MI are illustrated in Fig. 7, which considers the same noisy time series from Fig. 4, as one can notice by observing the extracted IMFs presented in Figs. 7(a,c,e,g) whose phase components are shown in Figs. 7(b,d,f,h). After the calculation of the Mutual Information between consecutive phase components, their average value is computed for finding the cutoff point, which separates stochastic from deterministic IMFs. In this scenario, it is possible to notice how phases change as new IMFs with deterministic behaviors are extracted from the original time series.

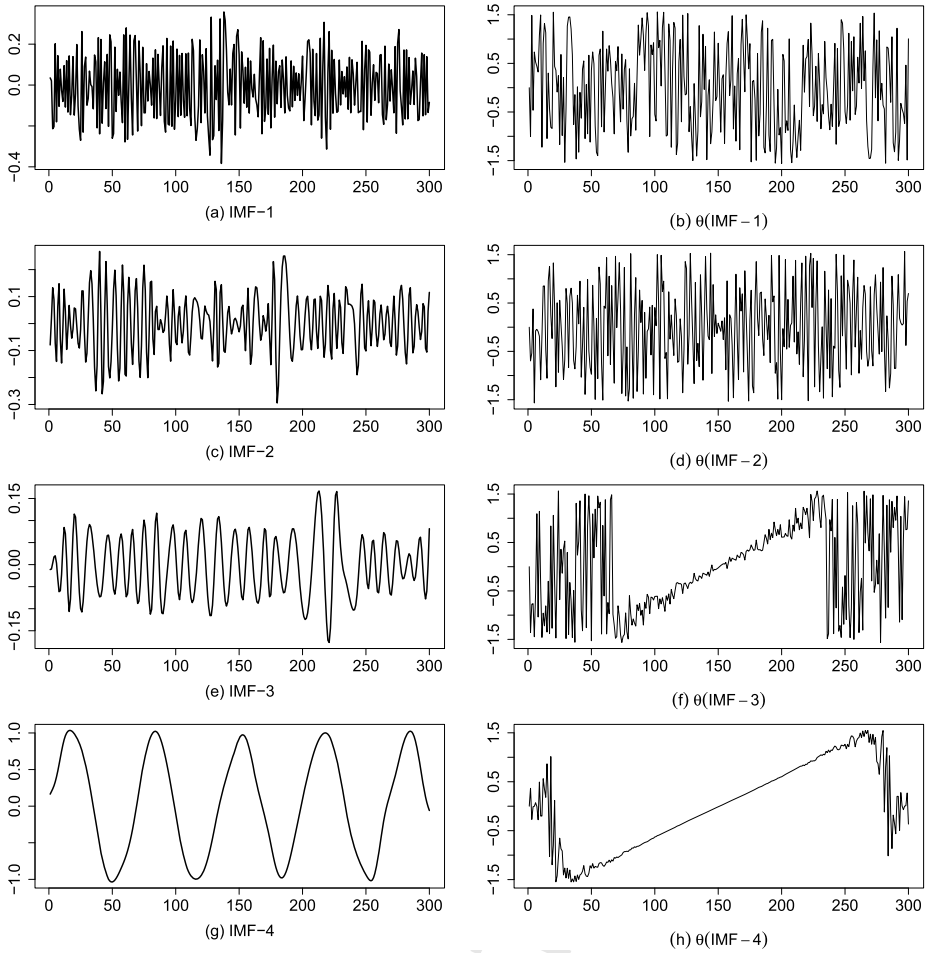



**Fig. 7.** (a), (c), (e), and (g) show the IMFs extracted after the application of EMD on the noisy time series displayed in Fig. 4(a). The phase components for each IMF are shown in plots (b), (d), (f), and (h).

## 3. Evaluation measures

This section presents a review of measures commonly adopted for comparisons of time series similarities. In the literature, such measures are widely considered to evaluate outcomes produced from time series predictors, quantifying how distant the predicted values are from the expected ones. In our study, they were used for assessing the capability of the decomposition approaches discussed in previous sections.

### 3.1. Mean absolute, mean squared, and root mean squared errors

A simple way to compare two values is by calculating their difference. This notion can be generalized to time series. All time-corresponding observations of a given time series $t^1$ are subtracted from a second time series $t^2$, and the summation of their individual differences provide a summary of the results. However, the simple sum of positive and negative errors would make them compensate for each other, thus yielding misleading results — as illustrated in Fig. 8, where a sinusoidal series (represented by empty dots) is fitted by a constant series (black filled dots). Observe in this particular scenario, the overall error is (misleadingly) zero.

As widely known, one can correct such biased estimates can be corrected by the sum of all absolute (or squared) errors, thus avoiding the canceling effect caused by opposite error signs. More precisely, the approach leads to two popular measures, namely Mean Absolute Error (MAE) in Equation (16), and Mean Squared Error (MSE), in Equation (17), where $N$ is the series length. A more intuitive measure based on MSE is the (squared) Root of MSE (RMSE for short) in Equation (18).

$$\mathrm{MAE}\left(t^1, t^2\right) = \frac{1}{N}\sum_{i=1}^{N}\left|t_i^1 - t_i^2\right| \tag{16}$$

$$\mathrm{MSE}\left(t^1, t^2\right) = \frac{1}{N}\sum_{i=1}^{N}\left(t_i^1 - t_i^2\right)^2 \tag{17}$$

$$\mathrm{RMSE}\left(t^1, t^2\right) = \sqrt{\mathrm{MSE}\left(t^1, t^2\right)} \tag{18}$$

Note the results for MAE and MSE are within the $[0, \infty)$ interval, in which zero means the compared series are exactly the same. MSE gives a relatively great relevance to large errors, and is more sensitive to outliers. Apart from such widely known measures, others less popular, but still useful for our study, are addressed in what follows.

### 3.2. Minkowski distance

The Minkowski distance is computed in Equation (19), where $t^1 = \{t_1^1, t_2^1, \ldots, t_N^1\}$ and $t^2 = \{t_1^2, t_2^2, \ldots, t_N^2\}$ are time series with $N$ observations and its computational complexity is $O(N)$.

$$D\left(t^1, t^2\right) = \left(\sum_{i=1}^{N}|t_i^1 - t_i^2|^p\right)^{1/p} \tag{19}$$

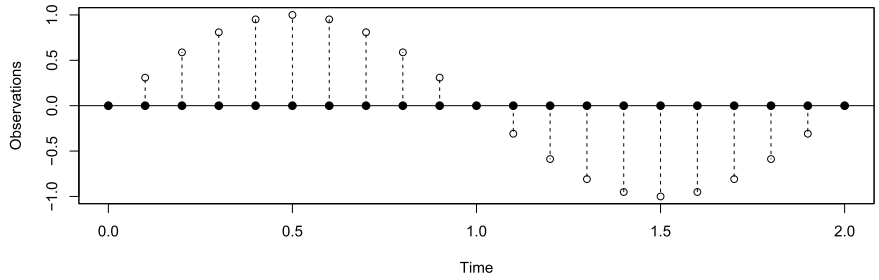



**Fig. 8.** Dotted lines represent the error from the difference of a sinusoidal series (empty dots) and a fitted constant series (black dots). The sum of the (positive and negative) errors is zero, which yields misleading results — in this case, the overall sum of the errors is zero.

This metric provides different interpretations depending on the chosen value of $p$. For example, by setting $p = 1$, it is referred to as the Manhattan distance, which sums up the absolute differences between observations. Similarly, the Euclidean distance is obtained when $p = 2$. Finally, by setting $p = \infty$, the infinite norm, which computes the largest differences among all pairs of observations, is provided.

At this point, it is helpful to recall that, to be considered a metric, a distance measure $D(\cdot,\cdot)$ must satisfy the following properties: i) (symmetry) $D(t^1, t^2) = D(t^2, t^1)$; ii) (positiveness) $D(t^1, t^2) \geq 0$; iii) (identity) $D(t^1, t^2) = 0 \iff t^1 = t^2$; and iv) (triangle inequality) $D(t^1, t^2) \leq D(t^1, t^3) + D(t^2, t^3)$.

### 3.3. Complexity-invariant distance

The Complexity-Invariant Distance (CID) was proposed for the analysis of the general time series behavior by looking for different complexities, instead of just comparing their observations [60]. In summary, the complexity invariance is obtained by a correction factor, $CF(\cdot)$ in Equation (21), applied on existing distance metrics $D(\cdot)$ in Equation (20). In Equation (21), the estimated complexity $CE(\cdot)$ for a time series $t$ is defined in Equation (22). Similarly to Minkowski, the time complexity of CID's $O(n)$. Regarding the metric definition, CID inherits the lower-bounding and triangular inequality properties of its distance function [60]. Thus, it is a metric when using a distance function holding metric space properties.

$$CID(t^1, t^2) = D(t^1, t^2) * CF(t^1, t^2) \tag{20}$$

$$CF(t^1, t^2) = \frac{max(CE(t^1), CE(t^2))}{min(CE(t^1), CE(t^2))} \tag{21}$$

$$CE(t) = \sqrt{\sum_{i=1}^{n-1} (t_i - t_{i+1})^2} \tag{22}$$

### 3.4. Pearson correlation

Time series can also be compared by taking into account correlation-based distances. In this sense, there is a commonly used measure based on Pearson's correlation factor in Equation (23) [61]. This measure captures the linear correlation of time series, and provides a value within the $[-1, 1]$ interval. If it approaches to $-1$, the two series have a negative linear correlation. If it approaches 1, the series contain a straight linear correlation. Finally, no linear correlation is established between series when is zero. The temporal and computational complexities for the traditional Pearson measure are $O(n)$. The Pearson correlation is not considered a metric, since it does not hold the triangle inequality restriction.

$$corr(t^1, t^2) = \frac{\sum_{i=1}^{N} (t_i^1 - \overline{t^1})(t_i^2 - \overline{t^2})}{\sqrt{\sum_{i=1}^{N} (t_i^1 - \overline{t^1})^2} \sqrt{\sum_{i=1}^{N} (t_i^2 - \overline{t^2})^2}} \tag{23}$$

### 3.5. Dynamic time warping

The main drawback faced by the previous measures is the dependence on the perfect alignment between time series, i.e., a small displacement among their observations affects the final result. To solve a problem, Dynamic Time Warping (DTW) searches for the best alignment between series before computing the distance [62].

Towards a better understanding of DTW, let $t^1$ and $t^2$ be two time series containing $M$ and $N$ observations, respectively. This measure starts computing a matrix $C$ whose cells contain the cumulative costs, as in Equation (24), where $D(\cdot,\cdot)$ is the distance between observations $i$ and $j$, such that $\{1 \leq i \leq M, 1 \leq j \leq N\}$, considering the first observation was collected at time instant 1.

$$C[x_i, y_j] = D(t_i^1, t_j^2) + min(dist(t_{i-1}^1, t_j^2), dist(t_i^1, t_{j-1}^2), dist(t_{i-1}^1, t_{j-1}^2)) \tag{24}$$

Next, DTW builds up the warping path by considering the reverse path, respecting a set of step size conditions [63], starting from cell $C[M, N]$ up to $C[1, 1]$ by adding cell values with the lowest cumulative cost. The warping path represents the perfect alignment between time series $t^1$ and $t^2$. The last cell in the path is the distance, whose value is normalized — divided by $M + N$. The DTW algorithm based on dynamic programming shows computational complexity of $O(MN)$. Note DTW is not a metric, since it does not satisfy the triangle inequality.

### 3.6. Mean distance from the diagonal line

Most of the measurements commonly adopted for the calculation of the distance between time series assess only the individual differences among pairs of observations without considering the global behavior that may exhibit similar trends and/or seasonalities. In order to address this issue, Rios and Mello [19] designed a new measurement, called Mean Distance from the Diagonal Line (MDDL), which is based on the Dynamic Time Warping (DTW) [62].

According to MDDL, DTW is initially computed on two time series for producing a matrix containing a warping path, whose first and last elements are connected through a straight diagonal line. The MDDL measurement is obtained by the computation of the absolute area between the warping path and the diagonal line [19]. MDDL scores vary in the $\left[0, \int_a^b |f(x)|dx\right]$ interval, such that $f(x)$ is the MDDL function used in the computation of the warping path, $a$ and $b$ are the first and last elements in the path, and $dx$ corresponds to differentials computed on the diagonal line. MDDL scores vary from zero, representing a perfect match between series, to the maximum area formed by the warping path and the diagonal line. In the worst case scenario, when the predicted and expected observations are completely different, MDDL

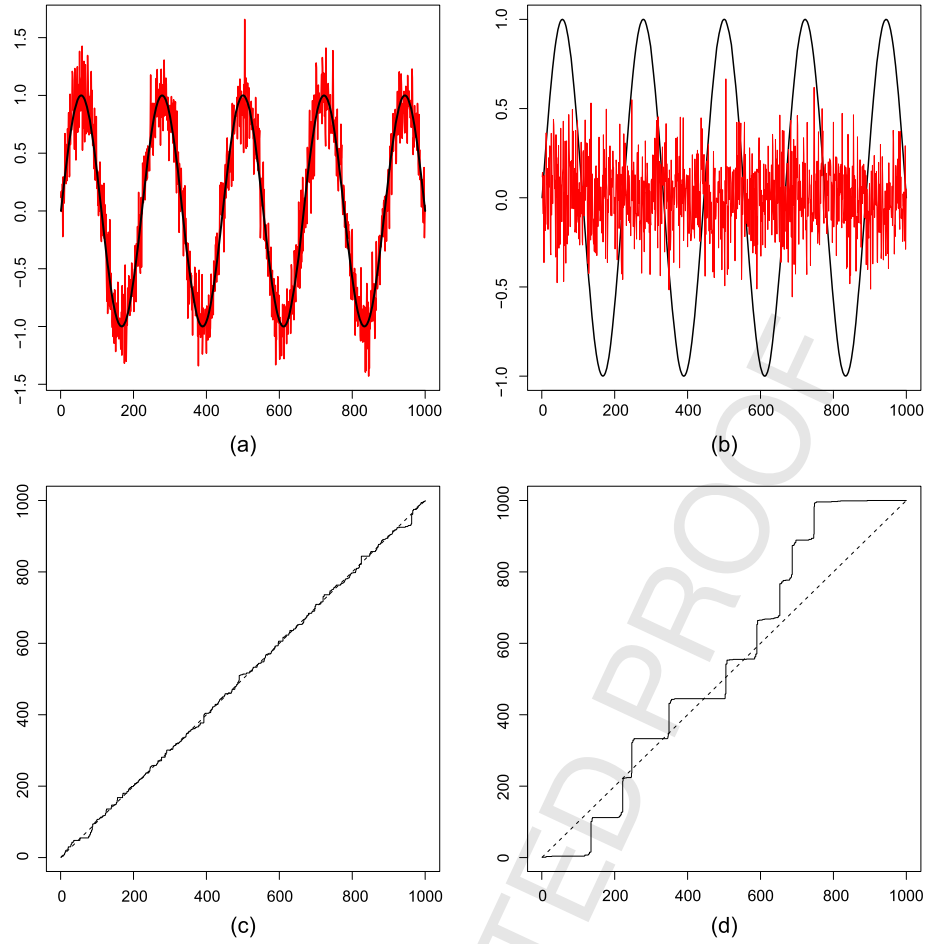



**Fig. 9.** (a) Comparison of a sinusoidal time series before and after addition of noise. (b) MDDL area for the time series shown in (a). (c) Comparison of a sinusoidal series and a white noise series. (d) MDDL area of the time series shown in (c).

returns the area of the isosceles triangle, whose sides are formed by $N$ expected and $N$ predicted observations, and the hypotenuse is formed by the diagonal line. Consequently, this interval can be written as $[0, +\frac{(N^2-N)}{2}]$. In practical situations, the number of observations is finite, therefore the Euclidean distance between the warping path and the diagonal line can be used for the calculation of the MDDL score [19].

Fig. 9 shows the way MDDL measures the similarity between series. Initially, two time series were produced by a sinusoidal function with $\pi$ angular frequency. One of them received some white noise according to a Normal distribution of 0 and 0.5 mean and standard deviation, respectively (both time series are shown in Fig. 9(a)). The MDDL distance between those series is 6.47, and the MDDL area formed between the warping path and the diagonal line is shown in Fig. 9(c). Both series exhibit similar seasonal behaviors, despite one of them being affected by noise. However, if a sinusoidal function is compared to the same white noise added to the sine function of the first example, neither the values, nor the general behavior are similar (see Fig. 9(b)). In such scenario, the MDDL distance is equal to 88.88, since the MDDL area increased, as shown in Fig. 9(d). The time complexity of MDDL depends on the DTW operations plus a linear comparison of $O(n)$ between the warping path and the diagonal line, therefore, MDDL cannot be considered a metric.

### 3.7. Mean distance from attractors

The Mean Distance from Attractors (MDA) [34] is a metric that evaluates the similarity of the phase spaces obtained from two time series. This strategy is promising since, it enables comparisons of recurrences, trends, and seasonalities embedded in phase spaces. In this sense, if two attractors have similar shapes, both time series exhibits similar behavior in the time domain.

Given time series $t^1$ and $t^2$, embedding dimension $m$ and time-delay separation $\tau$ must be estimated for the unfolding of series observations in the phase space, according to Taken's embedding theorem [38]. After the time series have been embedded, an Euclidean distance $d(.,.)$ is calculated on the states of attractors along their indices, i.e., the distance of the first state of attractor $A$ to the first state of attractor $B$, and so on. Finally, the average of all distances is computed and assigned as the MDA value. As a consequence, the lower the MDA value is, the more similar the time series. However, both attractors and time series must have the same length, which may be a problem if the decomposition technique does not produce components of same lengths of the original series.

### 3.8. Comparison

Several strategies can assess the decomposition process, which might be selected according to previous knowledge or conclusions drawn from data behaviors. However, the selection of a single strategy is a complex task or not feasible for the evaluation general-purpose decompositions. Towards a better understanding of such strategies, we performed several experiments to support the choice of the most suitable one for the scenarios under analysis.

| | MAE | RMSE | Minkowski P1 | Minkowski P2 | Minkowski P3 | Correlation | MDDL | MDA | Euclidean CID | DTW CID |
|---|---|---|---|---|---|---|---|---|---|---|
| MAE | 1,0000 | 0,9981 | 1,0000 | 0,9981 | 0,9944 | -0,9038 | 0,3858 | 0,9930 | 0,1450 | 0,1450 |
| RMSE | 0,9981 | 1,0000 | 0,9981 | 1,0000 | 0,9990 | -0,9057 | 0,3921 | 0,9951 | 0,1354 | 0,1354 |
| Minkowski P1 | 1,0000 | 0,9981 | 1,0000 | 0,9981 | 0,9944 | -0,9038 | 0,3858 | 0,9930 | 0,1450 | 0,1450 |
| Minkowski P2 | 0,9981 | 1,0000 | 0,9981 | 1,0000 | 0,9990 | -0,9057 | 0,3921 | 0,9951 | 0,1354 | 0,1354 |
| Minkowski P3 | 0,9944 | 0,9990 | 0,9944 | 0,9990 | 1,0000 | -0,9036 | 0,3944 | 0,9943 | 0,1287 | 0,1287 |
| Correlation | -0,9038 | -0,9057 | -0,9038 | -0,9057 | -0,9036 | 1,0000 | -0,5478 | -0,9216 | -0,1675 | -0,1675 |
| MDDL | 0,3858 | 0,3921 | 0,3858 | 0,3921 | 0,3944 | -0,5478 | 1,0000 | 0,4303 | 0,4571 | 0,4571 |
| MDA | 0,9930 | 0,9951 | 0,9930 | 0,9951 | 0,9943 | -0,9216 | 0,4303 | 1,0000 | 0,1438 | 0,1438 |
| Euclidean CID | 0,1450 | 0,1354 | 0,1450 | 0,1354 | 0,1287 | -0,1675 | 0,4571 | 0,1438 | 1,0000 | 1,0000 |
| DTW CID | 0,1450 | 0,1354 | 0,1450 | 0,1354 | 0,1287 | -0,1675 | 0,4571 | 0,1438 | 1,0000 | 1,0000 |

**Fig. 10.** Correlation matrix calculated for pairs of measures.

Each measure presented in this section is driven by specific biases, therefore, before proceeding with the experiments, we analyzed all measures by the pairwise Pearson correlation index, as shown in Fig. 10, for identifying similarities among their biases. The results were obtained after 2.3 million experiments conducted according to eight decomposition techniques on fifty-five synthetic time series. Different combinations of hyperparameters, as detailed in Section 4, were assessed for each technique.

As one compares the components towards absolute differences among observations, the measures MAE, RMSE, Minkowski P1, Minkowski P2, Minkowski P3, Pearson Correlation, and MDA resemble each other. Their similar results are due to the normalization applied on all time series, rescaling their observations within $[0, 1]$ range, and the values provided by decomposition step, which were very close to the expected ones. On the other hand, MDDL shows a relatively low correlation in comparison to other measurements, since it is focused on the comparison of the expected shapes, rather than on only the absolute values. CID has provided only high correlation results between its variations (see Equation (20)). Once we set two different measures, Euclidean and DTW, results emphasize the high correlation is typically associated with the correction factor than the distance measures. Therefore, by considering Fig. 10, it is likely that MDDL and MAE are enough to assess the decomposition process and analyze different biases, as detailed in the following section.

## 4. Experimental analysis

All state-of-the-art techniques are assessed from a wide range of tests towards capturing the performance of decomposition approaches in different scenarios. A grid search provided a fair sensitivity evaluation of hyperparameters, and stronger conclusions on the performance of the decomposition algorithms under study.

### 4.1. Datasets

Although some authors have applied decomposition techniques to model real-world time series [64–69,26], we used synthetic ones, in our experiments, due to three main reasons. Firstly, each dataset (synthetic or real) has its own bias, which may favor some particular methods. Therefore, the inclusion of only a few real-world datasets would influence our analyses. Secondly, the analysis of several datasets collected from different applications would make our work virtually impossible. Finally, the use of synthetic datasets enables the analyses of different types of decomposition bias created under controlled simulations. As a consequence, fifty-five synthetic time series were designed from combination of five different deterministic components with eleven stochastic ones. The deterministic components were chosen according to previous studies [12,70] and for describing real-world phenomena. Each component has a particular attractor shape in the phase-space, describing the general time series behavior.

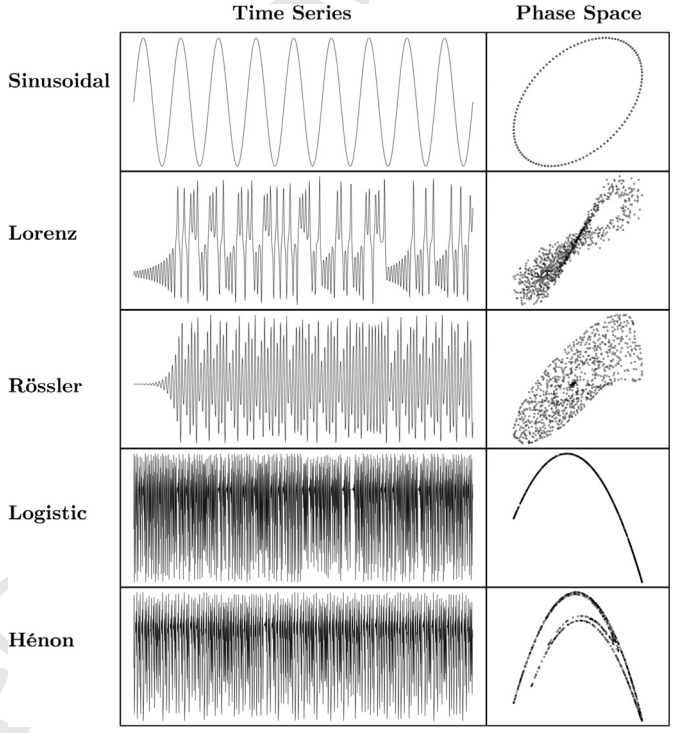



**Fig. 11.** Time series formed by the Sinusoidal function, Lorenz system, Rössler system, Logistic Map, and Hénon Map. The corresponding attractors in the phase space are also illustrated.

Typically used in climate models, as temperatures [71], sinusoidal functions are characterized by elliptical attractors. The Lorenz system and its simplification through the Rössler system describe a fluid circulation in a superficial layer, uniformly heated from the bottom and uniformly cooled from the top. Both have attractors with narrow trajectories and smooth transitions between states. Lorenz resembles an infinity symbol, while Rössler is similar to a disc with a raised side. Finally, Hénon and Logistic maps have attractors with abrupt transitions and shapes that remind a concave parabola. Fig. 11 shows such deterministic components (left column) and their phase spaces (right column).

Each deterministic component was rescaled within the $[0, 1]$ range and combined to 11 different stochastic components, which resulted in 55 evaluation scenarios. The stochastic components were formed by a zero constant time series, five series following a uniform distribution with $\{-0.01, -0.05, -0.10, -0.15, -0.20\}$ minimum values and $\{0.01, 0.05, 0.10, 0.15, 0.20\}$ maximum values, and five series following Normal distributions with zero mean and standard deviations set as $\{0.01, 0.05, 0.10, 0.15, 0.20\}$. The combination of the deterministic components with the stochastic ones forms one test set. Towards more robust and reliable results,

Please cite this article in press as: F.S.L.G. Duarte et al., Decomposing time series into deterministic and stochastic influences: A survey, Digit. Signal Process. (2019), https://doi.org/10.1016/j.dsp.2019.102582

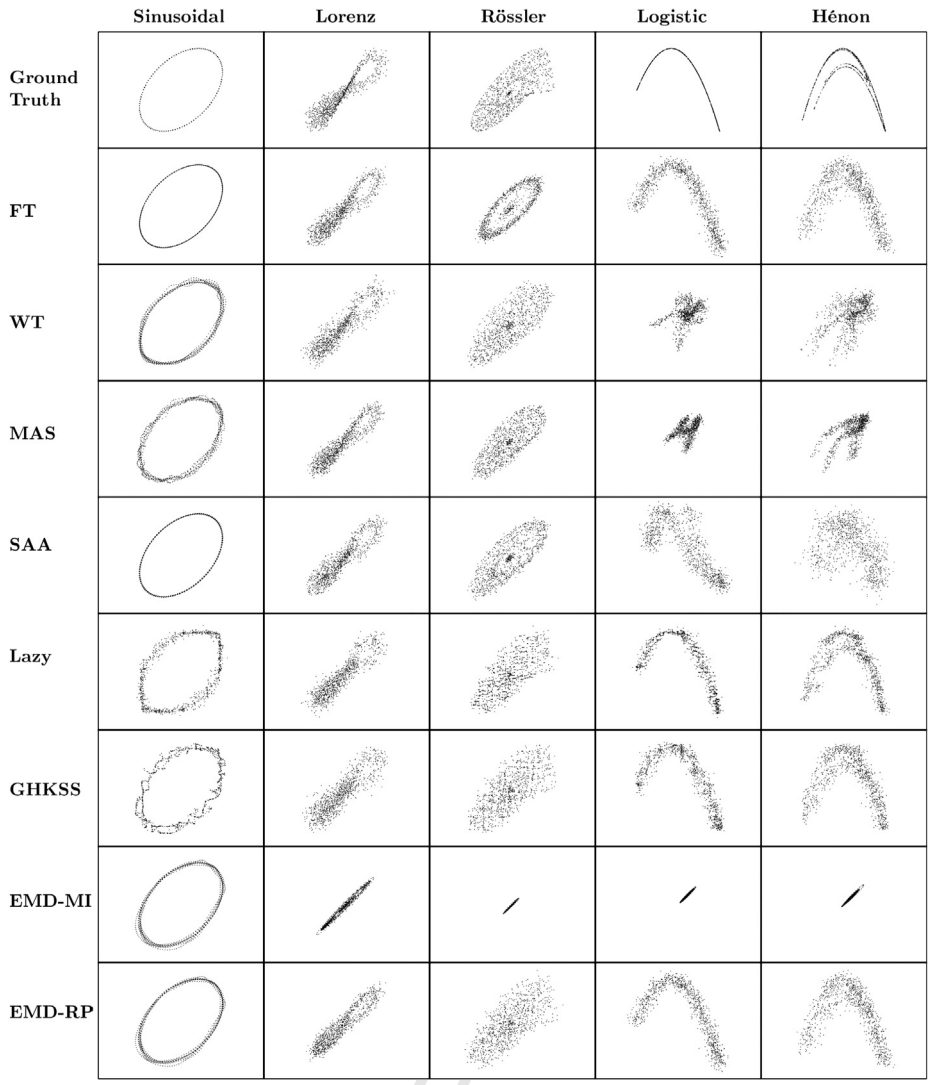



**Fig. 12.** Visual inspection of decomposition techniques applied over five different deterministic scenarios. Each column represents a deterministic component and each row corresponds to the best result for all eight decomposition techniques, namely Fourier, Wavelet, Moving Average Smoothing, SSA, Lazy, GHKSS, EMD-MI, and EMD-RP.

thirty test sets were generated and used by the decomposition strategies. The next section addresses some interesting conclusions drawn from a visual inspection on the results.

### 4.2. Visual inspection

The addition of a stochastic component affects the deterministic attractor is in the phase space. However, its topological shape is usually kept, which enables a visual inspection and provides important information on the decomposition process. In general, the decomposition process is well performed when the attractor shape of the extracted deterministic component is close to the expected behavior and free of noise. Based on this assumption, we visually compare the algorithms under different settings.

To illustrate this behavior, all the (five) deterministic components (described in Section 4.1) were combined with stochastic time series following Normal distributions. Eight decomposition techniques, namely Fourier, Wavelet, Moving Average Smoothing, SSA, Lazy, GHKSS, EMD-MI, and EMD-RP were employed and the decomposed deterministic components were unfolded into their phase spaces, as illustrated in Fig. 12. The plots enable assessments of trends, recurrences, and deterministic behaviors. The differences illustrate biases of the decomposition techniques.

Each column in Fig. 12 represents a deterministic component. The first row shows the expected attractor, i.e., the attractor prior to the noise addition representing the best decomposition result. Aiming at a fair visual inspection, all phase spaces in Fig. 12 were normalized in the $[-1, 1]$ range for avoiding false conclusions and visual pitfalls.

The visual inspection revealed all decomposition techniques showed good performances for sinusoidal time series and provided attractors similar to the expected ones. Although Lazy and GHKSS yielded the worst results, they showed the best performance for the remaining chaotic ones. A comparison of the extracted attractor and the ground truth promotes a better understand of the bias of each decomposition technique.

In most chaotic scenarios (last two columns of Fig. 12), WT, MAS and EMD-MI deformed the attractor shape completely. The reduction in the size of phase spaces indicates an amplitude shrinkage on the time space. Based on that, EMD-MI clearly changes the amplitudes of deterministic components regardless the scenario – except by considering sinusoidal time series. Finally, although such visual inspection is useful, numerical results are essential for a better understanding of the potential and drawbacks of the studied decomposition approaches, as discussed in the next section.

**Table 1**
Results of the Mean Absolute Error (MAE). Rows correspond to decomposition techniques and columns to deterministic components.

| | Mean Absolute Error – MAE | | | | | |
|---|---|---|---|---|---|---|
| | Sinusoidal | Lorenz | Rössler | Logistic | Hénon | Total |
| FT | $0.04 \pm 0.04$ | $0.07 \pm 0.04$ | $0.22 \pm 0.07$ | $0.08 \pm 0.07$ | $0.07 \pm 0.06$ | $0.09 \pm 0.08$ |
| WT | $0.05 \pm 0.04$ | $0.10 \pm 0.05$ | $0.22 \pm 0.05$ | $0.53 \pm 0.00$ | $0.39 \pm 0.01$ | $0.26 \pm 0.18$ |
| MAS | $0.03 \pm 0.02$ | $0.09 \pm 0.05$ | $0.20 \pm 0.07$ | $0.50 \pm 0.02$ | $0.40 \pm 0.02$ | $0.24 \pm 0.18$ |
| SSA | $0.02 \pm 0.02$ | $0.07 \pm 0.03$ | $0.13 \pm 0.02$ | $0.24 \pm 0.04$ | $0.24 \pm 0.04$ | $0.14 \pm 0.09$ |
| Lazy | $0.05 \pm 0.04$ | $0.08 \pm 0.03$ | $0.09 \pm 0.04$ | $0.06 \pm 0.04$ | $0.06 \pm 0.04$ | $0.07 \pm 0.04$ |
| GHKSS | $0.05 \pm 0.05$ | $0.06 \pm 0.05$ | $0.06 \pm 0.05$ | $0.06 \pm 0.05$ | $0.06 \pm 0.05$ | $0.06 \pm 0.05$ |
| EMD-MI | $0.13 \pm 0.22$ | $0.26 \pm 0.02$ | $0.39 \pm 0.00$ | $0.55 \pm 0.00$ | $0.48 \pm 0.00$ | $0.36 \pm 0.18$ |
| EMD-RP | $0.03 \pm 0.03$ | $0.08 \pm 0.06$ | $0.30 \pm 0.13$ | $0.07 \pm 0.08$ | $0.06 \pm 0.06$ | $0.11 \pm 0.13$ |

**Table 2**
Results of Mean Distance from Diagonal Line (MDDL). Rows correspond to decomposition techniques and columns denote deterministic components.

| | Mean Distance from Diagonal Line – MDDL | | | | | |
|---|---|---|---|---|---|---|
| | Sinusoidal | Lorenz | Rössler | Logistic | Hénon | Total |
| FT | $1.54 \pm 0.65$ | $3.08 \pm 1.20$ | $12.58 \pm 4.03$ | $1.92 \pm 0.83$ | $1.18 \pm 0.65$ | $4.06 \pm 4.73$ |
| WT | $1.66 \pm 0.70$ | $2.57 \pm 0.72$ | $5.74 \pm 2.60$ | $3.66 \pm 1.13$ | $3.65 \pm 0.50$ | $3.46 \pm 1.93$ |
| MAS | $2.33 \pm 0.75$ | $3.63 \pm 0.77$ | $4.64 \pm 2.40$ | $4.21 \pm 1.62$ | $3.23 \pm 0.68$ | $3.61 \pm 1.62$ |
| SSA | $2.31 \pm 1.75$ | $3.44 \pm 0.91$ | $22.12 \pm 2.71$ | $5.40 \pm 7.26$ | $3.83 \pm 0.66$ | $7.42 \pm 8.24$ |
| Lazy | $2.15 \pm 0.54$ | $3.26 \pm 1.20$ | $23.29 \pm 2.40$ | $2.14 \pm 0.92$ | $1.30 \pm 0.90$ | $6.43 \pm 8.56$ |
| GHKSS | $1.63 \pm 0.82$ | $4.08 \pm 1.94$ | $27.08 \pm 9.71$ | $2.00 \pm 1.04$ | $1.30 \pm 0.95$ | $7.22 \pm 10.94$ |
| EMD-MI | $37.99 \pm 72.38$ | $11.95 \pm 3.55$ | $67.80 \pm 31.14$ | $39.04 \pm 21.00$ | $53.38 \pm 28.63$ | $42.03 \pm 42.87$ |
| EMD-RP | $2.87 \pm 1.34$ | $5.21 \pm 2.38$ | $18.97 \pm 7.95$ | $2.82 \pm 1.43$ | $2.05 \pm 1.55$ | $6.38 \pm 7.46$ |

### 4.3. Quantitative evaluation

Despite the relevance of the visual inspection, it is not scalable in our scenario, which considers the assessment of a wide range of hyperparameters and strategies. As explained in Section 3, many measurements support users decisions, and most of them evaluate the same bias and trends, thus yielding similar and redundant conclusions. In attempt to avoid an overbiased situation, we here analyse a series of experiments considering two measurements that have different biases, namely (i) Mean Absolute Error (MAE) and (ii) Mean Distance from Diagonal Line (MDDL).

Experiments were conducted according to eight state-of-the-art techniques.[3] Grid search was employed to find suitable settings of hyperparameters that decompose the time series. Each setting was used in experiments on all datasets, which resulted in $1,650$ scenarios — as described in Section 4.1, 2.3 million experiments were run.

The best hyperparameter setting was chosen for each test scenario for reducing both MAE and MDDL, relying on the Pareto Frontier [72]. This approach is not sensitive to data scales, which is an important property, since such measurements have different magnitudes. Table 1 shows a summary of results for MAE, where rows correspond to the decomposition techniques and columns denote deterministic sources. Since deterministic series were added to eleven stochastic influences, high variances were observed in this benchmark composed of thirty sets. Therefore, we decided to report averages and standard deviations to support our conclusions.

Confirming the visual inspection, GHKSS and Lazy did not present the best results for sinusoidal scenarios. This was expected once the bias imposed by the other techniques matches with the sinusoidal behavior. However, they obtained the lowest averages and standard deviations, which suggest the decomposition process performed in the phase space helps the preservation of both attractor shapes and deterministic and overall behavior such a characteristic is important in the context of time series forecasting, since improves the prediction performance. In addition, note the results for Rössler in which they had, on average, errors five times smaller than the others. The worst decomposition results were produced by EMD-MI, especially due to its non-parametric (automatic) approach for the decomposition of time series, a cost to pay under some scenarios.

Taking into account the MDDL results (Table 2), once again EMD-MI showed the worst results as compared to the other studied techniques. WT and MAS yielded, on average, the lowest overall averages and standard deviations, ranking them as the best approaches to decompose series in terms of MDDL. However, by analyzing individual results, we observe that Lazy and GHKSS presented good results, except in case of Rössler, whose results somehow strongly influenced their overall performances.

By observing that the results got in the Rössler dataset biased our analysis, we decided to employ violin plots as auxiliary tools, thereby supporting the analysis of distributions per decomposition technique. This was necessary, since we are dealing with multimodal distributions whose averages and standard deviations may not be enough to describe and compare them. So, once lower MDA and MDDL values mean a better fit, we should look for the distribution that concentrates most of its results around the bottom region.

From the MAE violin (Fig. 13(a)), it is notably how the results of WT, MAS, and EMD-MI have high variances, indicating their strong link with the data scenarios. In spite of the same high variance for EMD-RP, most of the results is concentrated in the base. This indicates that the variation is not related to the data scenarios, but to some other characteristic of the technique. The high concentration of results in the lower region for Lazy and GHKSS confirmed they are the best approaches for general-purpose decompositions.

The most interesting visualization is provided by MDDL violins (Fig. 13(b)). According to Table 2, Lazy and GHKSS produced the worst results, however, their distributions show a higher concentration of values in the bottom region, as desired. Aiming at a better visualization of the violin area of interest (i.e. the lower values), all results larger than 30 were cut off. Therefore, the results for EMD-MI, concentrated around 30, were even worse. Another relevant observation is some distributions are multimodal, according to the data under analysis, and prone to provide the best results depending on the data behavior. Therefore, the less curled the distribution, the more general-purpose (or robust) such a decomposition to different input data.

[3] All scripts are available at https://github.com/felipelageduarte/Survey.

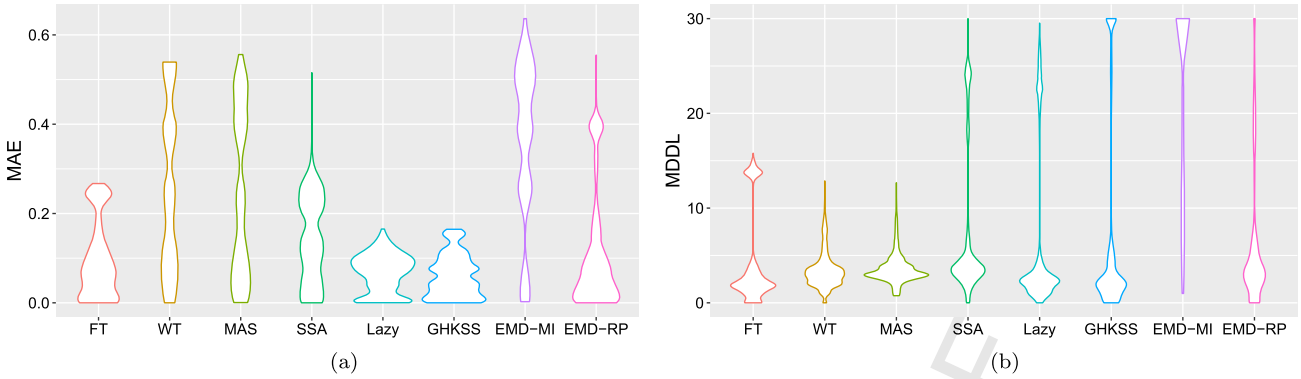



**Fig. 13.** Analyzing the results using violin plots: (a) Mean Absolute Error (MAE); (b) Mean Distance from the Diagonal Line (MDDL).

| | | FT | WT | MAS | SSA | Lazy | GHKSS | EMD-MI | EMD-RP |
|---|---|---|---|---|---|---|---|---|---|
| MAE | FT | 100% | 82% | 70% | 55% | 44% | 41% | 89% | 38% |
| | WT | 18% | 100% | 40% | 9% | 18% | 15% | 88% | 26% |
| | MAS | 30% | 60% | 100% | 14% | 24% | 18% | 86% | 33% |
| | SSA | 45% | 91% | 86% | 100% | 36% | 30% | 96% | 43% |
| | Lazy | 56% | 82% | 76% | 64% | 100% | 43% | 85% | 50% |
| | GHKSS | 59% | 85% | 82% | 70% | 57% | 100% | 86% | 52% |
| | EMD-MI | 11% | 12% | 14% | 4% | 15% | 14% | 100% | 24% |
| | EMD-RP | 65% | 74% | 67% | 57% | 50% | 48% | 80% | 100% |
| MDDL | FT | 100% | 61% | 74% | 86% | 75% | 72% | 100% | 98% |
| | WT | 41% | 100% | 61% | 77% | 52% | 46% | 100% | 63% |
| | MAS | 26% | 39% | 100% | 63% | 37% | 33% | 99% | 55% |
| | SSA | 16% | 25% | 37% | 100% | 33% | 38% | 97% | 48% |
| | Lazy | 29% | 48% | 63% | 67% | 100% | 50% | 98% | 74% |
| | GHKSS | 40% | 56% | 67% | 64% | 54% | 100% | 99% | 80% |
| | EMD-MI | 0% | 0% | 1% | 3% | 2% | 1% | 100% | 5% |
| | EMD-RP | 16% | 38% | 45% | 54% | 31% | 33% | 99% | 100% |

**Fig. 14.** Number of times a decomposition technique (rows) was better than or equal to the remaining ones (columns).

As discussed elsewhere, the technique of best results of MAE and MDDL is considered the most appropriate for tackling general-purpose decomposition problems. Although some techniques provide good results for both measures, by analyzing the violin plots, we conclude that techniques having a dispersion along MAE concentrate values at the bottom of MDDL, and vice versa. This confirms the relevance of those two measurements, given they are somehow complementary to each other, making difficult the definition of a single best strategy.

Although the violin plots complement the analysis of the results, they are still rather subjective, only giving a visualization of the distribution. Note that it is not possible to see in which scenario a specific technique is better than another. In this context, the table in Fig. 14 addresses this need by showing how many times a given decomposition technique (row) was better or equal to others (column). Lazy, GHKSS and Fourier were the best approaches according to MAE and MDDL measurements. However, none of them had the best results for all scenarios, suggesting that the decomposition results depend on the bias imposed by each technique, the evaluation measure, and specific time series characteristics.

We performed statistical tests following the approach described in [73] for assuring the validity and non-randomness of the results Friedman test [73], which is a non-parametric randomized block analysis of variance, followed by Nemenyi test, which is a post-hoc that finds the groups of data that differ from the multiple comparisons run by the Friedman test were applied. Considering the pair-wise comparisons, we get p-values less than $10^{-16}$, thereby strongly suggesting our results are statistically reliable.

## 5. Summary of decomposition approaches

This section summarizes all decomposition approaches analyzed in this manuscript. Firstly, we discussed and detailed the advantages and limitations of using transformation-based methods. The Fourier Transform showed the best time complexity, however it is used to decompose periodic signals without locating frequencies along time. To solve this issue, Wavelet transform was proposed based on a infinite set of basis functions, allowing to improve FT results. The main problem faced by the methods is the change in the original behavior for non-periodic time series.

Such a problem motivated the development of Phase-Space-based methods, which aim at decomposing the stochastic and deterministic components taking into account high-dimensional spaces. The methods are especially required for the modeling of time series with chaotic behaviors.

By analyzing the state-of-the-art researches, we also noticed there are other decomposition methods interested in directly extracting stochastic and deterministic components in time domain. The main advantage of using those methods is the possibility of analyzing time series without requiring any transformation. Finally, it is worth to emphasizing several real-world problems have been addressed by decomposing time series components by using subspace-based method. In this scenario, Singular Spectrum Analysis has been widely adopted, providing fair results even in nonstationary time series.

Towards a better understanding of the strategies addressed, Table 3 provides a summary of their biases, computational complexities, and decomposition results. The first three columns enable a comparison of each technique according to the working space (i.e., time or phase), number of parameters, and computational com-

**Table 3**
Summary of the surveyed decomposition techniques, highlighting: (i) space in which the decomposition happens (i.e., time or phase), (ii) the number of hyper-parameters, (iii) computational complexities, and (iv) experimental results.

| | Working Space | #Parameters | Computational Complexity | #Wins | | |
|---|---|---|---|---|---|---|
| | | | | MAE | MDDL | Overall |
| FT | Transformation-based method | 1 | $O(n\ log\ n)$ | 4 | 7 | 11 |
| WT | Transformation-based method | 2 | $O(n)$ | 1 | 5 | 6 |
| MA | Time-Domain method | 1 | $O(n)$ | 2 | 3 | 5 |
| SSA | Subspace-based method | 2 | $O(n^3)$ | 3 | 1 | 4 |
| Lazy | Phase-Space-based method | 4 | $O(n^2)$ | 5 | 3 | 8 |
| GHKSS | Subspace-based method | 6 | $O(n^3)$ | 7 | 6 | 13 |
| EMD-MI | Time-Domain method | 0 | $O(n^2)$ | 0 | 0 | 0 |
| EMD-RP | Time-Domain method | 4 | $O(n^2)$ | 5 | 2 | 7 |

plexities.[4] This may help the reader to adopt a technique based on data size and/or time constraints.

The last three columns of the table show the number of wins — defined as number of times a given technique was better than or equal to — according to MAE, MDDL, and Overall (MAE + MDDL). This summary suggests GHKSS, Fourier, and Lazy provide, on average, the best results, especially when dealing with chaotic time series, since the expected attractor trajectories were better maintained. The results from EMD-MI are never better than any other decomposition approach, which is expected for a non-parametric methodology that decomposes the time series with no *a priori* users knowledge.

## 6. Concluding remarks

Time series decomposition supports the separation of deterministic and stochastic influences, taking advantage of Statistics and Dynamical System concepts for modeling and forecasting tasks. From this standpoint, we surveyed decomposition techniques and assessed them according to different evaluation criteria, in particular trying to maintain the underlying deterministic bias over phase spaces. Our study is also helpful for practical decision-making processes regarding the most adequate approaches for decomposing time series and assessing them.

Common evaluation criteria identified which of them tend to capture similar biases. The widely known Mean Absolute Error (MAE) and Mean Distance from the Diagonal Line (MDDL) seem to be the best criteria to assess differences between time series, since they analyze the overall data trends over time, and not simply the squared and absolute local differences. In our test scenarios, all time series were normalized into the [0,1] interval, which justifies why MAE and Root Mean Squared Error (RMSE) provided similar results.

Our results can help for practitioners choose more suitable decomposition strategy for a given target problem. By analyzing time series with no previous knowledge, GHKSS confirmed to properly decompose stochastic and deterministic components besides presenting high time complexity.

As future work, we aim at building a taxonomy that connects our conclusions on decomposition approaches to different application scenarios, signal-to-noise ratios, and time series characteristics.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

[4] The estimations of the computational complexities were based on the original algorithm with no optimization.

## Acknowledgments

This work was supported by FAPESP Research Foundation, São Paulo, Brazil under grants #2013/07375-0, #2014/21636-3, #2017/16548-6, CNPq, Brazil under grant #302077/2017-0. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of CNPq, FAPESP, and Itaú-Unibanco.

## References

[1] J.D. Hamilton, Time Series Analysis, vol. 2, Princeton University Press, Princeton, NJ, 1994.
[2] Y. Amihud, Illiquidity and stock returns: cross-section and time-series effects, J. Financ. Mark. 5 (1) (2002) 31–56, https://doi.org/10.1016/s1386-4181(01)00024-6.
[3] C. Su, X. Yin, R. Tao, O.-R. Lobonț, N.-C. Moldovan, Are there significant linkages between two series of housing prices, money supply and short-term international capital? – evidence from China, Digit. Signal Process. 83 (2018) 148–156, https://doi.org/10.1016/j.dsp.2018.08.017.
[4] V. Chandola, A. Banerjee, V. Kumar, Anomaly detection: a survey, ACM Comput. Surv. 41 (3) (2009) 15, https://doi.org/10.1145/1541880.1541882.
[5] K.-M. Lau, H. Weng, Climate signal detection using wavelet transform: how to make a time series sing, Bull. Am. Meteorol. Soc. 76 (12) (1995) 2391–2402, https://doi.org/10.1175/1520-0477(1995)076<2391:csduwt>2.0.co;2.
[6] O. Raaf, A.E.H. Adane, Pattern recognition filtering and bidimensional FFT-based detection of storms in meteorological radar images, Digit. Signal Process. 22 (5) (2012) 734–743, https://doi.org/10.1016/j.dsp.2012.04.008.
[7] J.S. Richman, J.R. Moorman, Physiological time-series analysis using approximate entropy and sample entropy, Am. J. Physiol., Heart Circ. Physiol. 278 (6) (2000) H2039–H2049, https://doi.org/10.1152/ajpheart.2000.278.6.h2039.
[8] J.a. Gama, I. Žliobaitė, A. Bifet, M. Pechenizkiy, A. Bouchachia, A survey on concept drift adaptation, ACM Comput. Surv. 46 (4) (2014) 44, https://doi.org/10.1145/2523813.
[9] R.F. de Mello, Y. Vaz, C.H. Grossi, A. Bifet, On learning guarantees to unsupervised concept drift detection on data streams, Expert Syst. Appl. 117 (2019) 90–102, https://doi.org/10.1016/j.eswa.2018.08.054.
[10] F.G. da Costa, F.S. Duarte, R.M. Vallim, R.F. de Mello, Multidimensional surrogate stability to detect data stream concept drift, Expert Syst. Appl. 87 (2017) 15–29, https://doi.org/10.1016/j.eswa.2017.06.005.
[11] G.E.P. Box, G. Jenkins, Time Series Analysis, Forecasting and Control, 3rd edition, Holden-Day, Incorporated, 1976.
[12] H. Kantz, T. Schreiber, Nonlinear Time Series Analysis, Cambridge Nonlinear Science Series, Cambridge University Press, 2004.
[13] R.A. Rios, Improving Time Series Modeling by Decomposing and Analysing Stochastic and Deterministic Influences, Ph.D. thesis, University of São Paulo, 10 2013, http://www.teses.usp.br/teses/disponiveis/55/55134/tde-18112013-143708/publico/ricardorios.pdf.
[14] E.J. Kostelich, T. Schreiber, Noise reduction in chaotic time-series data: a survey of common methods, Phys. Rev. E 48 (3) (1993) 1752–1763, https://doi.org/10.1103/physreve.48.1752.
[15] J.L. Holloway, Smoothing and Filtering of Time Series and Space Fields, Advances in Geophysics, vol. 4, Elsevier, 1958, pp. 351–389.
[16] J. Gao, H. Sultan, J. Hu, W.-W. Tung, Denoising nonlinear time series by adaptive filtering and wavelet shrinkage: a comparison, IEEE Signal Process. Lett. 17 (3) (2010) 237–240, https://doi.org/10.1109/lsp.2009.2037773.
[17] N. Murata, S. Ikeda, A. Ziehe, An approach to blind source separation based on temporal structure of speech signals, Neurocomputing 41 (1–4) (2001) 1–24, https://doi.org/10.1016/s0925-2312(00)00345-3.
[18] A. Subasi, EEG signal classification using wavelet feature extraction and a mixture of expert model, Expert Syst. Appl. 32 (4) (2007) 1084–1093, https://doi.org/10.1016/j.eswa.2006.02.005.

[19] R.A. Rios, R.F. Mello, Improving time series modeling by decomposing and analyzing stochastic and deterministic influences, Signal Process. 93 (11) (2013) 3001–3013, https://doi.org/10.1016/j.sigpro.2013.04.017.
[20] R.F. de Mello, M.A. Ponti, Machine Learning - A Practical Approach on the Statistical Learning Theory, Springer, 2018.
[21] S. Bochner, K. Chandrasekharan, Fourier Transforms, Princeton University Press, 1949.
[22] P. Bloomfield, Fourier Analysis of Time Series: An Introduction, John Wiley & Sons, 2004.
[23] C. Candan, M.A. Kutay, H.M. Ozaktas, The discrete fractional Fourier transform, IEEE Trans. Signal Process. 48 (2000) 1329–1337, https://doi.org/10.1109/78.839980.
[24] R. Palaniappan, Utilizing gamma band to improve mental task based brain-computer interface design, IEEE Trans. Neural Syst. Rehabil. Eng. 14 (3) (2006) 299–303, https://doi.org/10.1109/tnsre.2006.881539.
[25] D. Gabor, Theory of communication. Part 1: the analysis of information, J. Inst. Electr. Eng., Part 3, Radio Commun. Eng. 93 (26) (1946) 429–441, https://doi.org/10.1049/ji-3-2.1946.0074.
[26] R.A. Rios, L. Parrott, H. Lange, R.F. de Mello, Estimating determinism rates to detect patterns in geospatial datasets, Remote Sens. Environ. 156 (2015) 11–20, https://doi.org/10.1016/j.rse.2014.09.019.
[27] N.E. Huang, Z. Shen, S.R. Long, M.C. Wu, H.H. Shih, Q. Zheng, N.-C. Yen, C.C. Tung, H.H. Liu, The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, Proc. R. Soc. Lond., Ser. A, Math. Phys. Eng. Sci. 454 (1971) (1998) 903–995, https://doi.org/10.1098/rspa.1998.0193.
[28] A. Graps, An introduction to wavelets, IEEE Comput. Sci. Eng. 2 (2) (1995) 50–61, https://doi.org/10.1109/99.388960.
[29] M. Stocchi, M. Marchesi, Fast wavelet transform assisted predictors of streaming time series, Digit. Signal Process. 77 (2018) 5–12, https://doi.org/10.1016/j.dsp.2017.09.014.
[30] D. Percival, Wavelet Methods for Time Series Analysis, Cambridge University Press, Cambridge, New York, 2000.
[31] A. Haar, Zur theorie der orthogonalen funktionensysteme, Math. Ann. 69 (3) (1910) 331–371, https://doi.org/10.1007/BF01456326.
[32] J. Azcarraga, M.T. Suarez, Predicting academic emotions based on brainwaves, mouse behaviour and personality profile, in: P. Anthony, M. Ishizuka, D. Lukose (Eds.), PRICAI 2012: Trends in Artificial Intelligence, Springer, Berlin, Heidelberg, 2012, pp. 728–733.
[33] B. Rubik, Neurofeedback-enhanced gamma brainwaves from the prefrontal cortical region of meditators and non-meditators and associated subjective experiences, J. Altern. Complement. Med. 17 (2) (2011) 109–115, https://doi.org/10.1089/acm.2009.0191.
[34] F.S.L.G. Duarte, R.A. Rios, E.R. Hruschka, R.F. de Mello, Time series decomposition using spring system applied on phase spaces, in: 2018 7th Brazilian Conference on Intelligent Systems (BRACIS), 2018, pp. 504–509.
[35] N. Golyandina, Analysis of Time Series Structure: SSA and Related Techniques, 1st edition, Chapman and Hall/CRC, 2001.
[36] R. Vautard, M. Ghil, Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series, Phys. D: Nonlinear Phenom. 35 (3) (1989) 395–424, https://doi.org/10.1016/0167-2789(89)90077-8.
[37] M. Leles, J. Sansão, L. Mozelli, H. Guimarães, Improving reconstruction of time-series based in singular spectrum analysis: a segmentation approach, Digit. Signal Process. 77 (2018) 63–76, https://doi.org/10.1016/j.dsp.2017.10.025.
[38] F. Takens, Detecting strange attractors in turbulence, in: Dynamical Systems and Turbulence, Warwick 1980, Springer, 1981, pp. 366–381.
[39] V.Y. Pan, Structured Matrices and Polynomials: Unified Superfast Algorithms, Springer Science & Business Media, 2012.
[40] F. De la Torre, M.J. Black, A framework for robust subspace learning, Int. J. Comput. Vis. 54 (1) (2003) 117–142, https://doi.org/10.1023/A:1023709501986.
[41] T. Schreiber, Extremely simple nonlinear noise-reduction method, Phys. Rev. E 47 (4) (1993) 2401–2404, https://doi.org/10.1103/PhysRevE.47.2401.
[42] C.M. Bishop, Pattern Recognition and Machine Learning (Information Science and Statistics), Springer-Verlag, New York, Inc., Secaucus, NJ, USA, 2006.
[43] R. Hegger, H. Kantz, T. Schreiber, Practical implementation of nonlinear time series methods: the TISEAN package, Chaos 9 (1999) 413.
[44] P. Grassberger, R. Hegger, H. Kantz, C. Schaffrath, T. Schreiber, On noise reduction methods for chaotic data, Chaos 3 (2) (1993) 127–141.
[45] H. Kantz, T. Schreiber, I. Hoffmann, T. Buzug, G. Pfister, L.G. Flepp, J. Simonet, R. Badii, E. Brun, Nonlinear noise reduction: a case study on experimental data, Phys. Rev. E 48 (1993) 1529–1538, https://doi.org/10.1103/PhysRevE.48.1529.
[46] H.K.R. Hegger, T. Schreiber, Practical implementation of nonlinear time series methods: the tisean package, Chaos 9 (1999) 413.
[47] T. Schreiber, M. Richter, Nonlinear projective filtering in a data stream, Int. J. Bifurc. Chaos 9 (1999).
[48] S. Beveridge, C.R. Nelson, A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the 'business cycle', J. Monet. Econ. 7 (2) (1981) 151–174, https://doi.org/10.1016/0304-3932(81)90040-4.
[49] V. Gomez, A. Maravall, Programs tramo (time series regression with arima noise, missing observations, and outliers) and seats (signal extraction in arima time series). instructions for the user, Documento de Trabajo 9628.
[50] S.E. Said, D.A. Dickey, Testing for unit roots in autoregressive-moving average models of unknown order, Biometrika 71 (3) (1984) 599–607, https://doi.org/10.1093/biomet/71.3.599.
[51] A. Savitzky, M.J.E. Golay, Smoothing and differentiation of data by simplified least squares procedures, Anal. Chem. 36 (8) (1964) 1627–1639, https://doi.org/10.1021/ac60214a047.
[52] P. Flandrin, G. Rilling, P. Goncalves, Empirical mode decomposition as a filter bank, IEEE Signal Process. Lett. 11 (2) (2004) 112–114, https://doi.org/10.1109/LSP.2003.821662.
[53] R.A. Rios, R.F. de Mello, Applying empirical mode decomposition and mutual information to separate stochastic and deterministic influences embedded in signals, Signal Process. 118 (2016) 159–176, https://doi.org/10.1016/j.sigpro.2015.07.003.
[54] A. Jerri, The Shannon sampling theorem – its various extensions and applications: a tutorial review, Proc. IEEE 65 (11) (1977) 1565–1596, https://doi.org/10.1109/PROC.1977.10771.
[55] J.-P. Eckmann, S.O. Kamphorst, D. Ruelle, Recurrence plots of dynamical systems, Europhys. Lett. 4 (9) (1987) 973.
[56] N. Marwan, M.C. Romano, M. Thiel, J. Kurths, Recurrence plots for the analysis of complex systems, Phys. Rep. 438 (5–6) (2007) 237–329, https://doi.org/10.1016/j.physrep.2006.11.001.
[57] J.P. Zbilut, C.L.W. Jr., Embeddings and delays as derived from quantification of recurrence plots, Phys. Lett. A 171 (3–4) (1992) 199–203, https://doi.org/10.1016/0375-9601(92)90426-M.
[58] C.L.W. Jr., J.P. Zbilut, Dynamical assessment of physiological systems and states using recurrence plot strategies, J. Appl. Physiol. 76 (2) (1994) 965–973.
[59] G. Darbellay, I. Vajda, Estimation of the information by an adaptive partitioning of the observation space, IEEE Trans. Inf. Theory 45 (4) (1999) 1315–1321, https://doi.org/10.1109/18.761290.
[60] G.E.A.P.A. Batista, E.J. Keogh, O.M. Tataw, V.M.A. de Souza, Cid: an efficient complexity-invariant distance for time series, Data Min. Knowl. Discov. 28 (3) (2014) 634–669.
[61] J. Benesty, J. Chen, Y. Huang, I. Cohen, Pearson correlation coefficient, in: Noise Reduction in Speech Processing, Springer, 2009, pp. 1–4.
[62] P. Tormene, T. Giorgino, S. Quaglini, M. Stefanelli, Matching incomplete time series with dynamic time warping: an algorithm and an application to post-stroke rehabilitation, Artif. Intell. Med. 45 (1) (2009) 11–34, https://doi.org/10.1016/j.artmed.2008.11.007.
[63] M. Müller, Information Retrieval for Music and Motion, vol. 2, Springer, 2007.
[64] R. Rios, R. de Mello, A systematic literature review on decomposition approaches to estimate time series components, INFOCOMP J. Comput. Sci. 11 (3–4) (2012) 31–46, http://www.dcc.ufla.br/infocomp/index.php/INFOCOMP/article/view/361.
[65] E. de Lima, A. Andrade, J. Pons, P. Kyberd, S. Nasuto, Empirical mode decomposition: a novel technique for the study of tremor time series, Med. Biol. Eng. Comput. 44 (2006) 569–582, https://doi.org/10.1007/s11517-006-0065-x.
[66] S. Amman, M. Das, An efficient technique for modeling and synthesis of automotive engine sounds, IEEE Trans. Ind. Electron. 48 (1) (2001) 225–234, https://doi.org/10.1109/41.904583.
[67] N.P. Macciotta, A. Cappio-Borlino, G. Pulina, Time series autoregressive integrated moving average modeling of test-day milk yields of dairy ewes, J. Dairy Sci. 83 (5) (2000) 1094–1103, http://www.biomedsearch.com/nih/Time-series-autoregressive-integrated-moving/10821585.html.
[68] G. Tzagkarakis, M. Papadopouli, P. Tsakalides, Singular spectrum analysis of traffic workload in a large-scale wireless lan, in: Proceedings of the 10th ACM Symposium on Modeling, Analysis, and Simulation of Wireless and Mobile Systems, MSWiM '07, ACM, New York, NY, USA, 2007, pp. 99–108.
[69] M. Tiemeyer, J.S.K. Wong, A task migration algorithm for heterogeneous distributed computing systems, J. Syst. Softw. 41 (3) (1998) 175–188, Elsevier Science.
[70] L.d.C. Pagliosa, R.F.d. Mello, Applying a kernel function on time-dependent data to provide supervised-learning guarantees, Expert Syst. Appl. 71 (C) (2017) 216–229, https://doi.org/10.1016/j.eswa.2016.11.028.
[71] Y. Zhang, M. Zou, S. Yang, Y. Lin, C. Ye, Colder-winter monthly-temperature forecasting in general trend of global warming via multiple sine functions decomposition, in: 2017 36th Chinese Control Conference (CCC), 2017, pp. 4551–4555.
[72] P. Godfrey, R. Shipley, J. Gryz, Algorithms and analyses for maximal vector computation, VLDB J. 16 (1) (2007) 5–28, https://doi.org/10.1007/s00778-006-0029-7.
[73] J. Demšar, Statistical comparisons of classifiers over multiple data sets, J. Mach. Learn. Res. 7 (2006) 1–30, http://dl.acm.org/citation.cfm?id=1248547.1248548.

**Felipe S.L.G. Duarte** received the B.Sc. degree in Computer Science from the University of São Paulo, Brazil, in 2012. He is currently a Ph.D. candidate at the Computer Science Department of the Institute for Mathematics and Computer Science, at the University of São Paulo and Data

Scientist Manager in the Finance Industry. His research interests include information visualization, Machine Learning, Time Series Analysis and Data Streams.

**Ricardo Rios** has concluded his PhD degree in 2013 from University of São Paulo. Nowadays, he is assistant professor of Computer Science at the Department of Computer Science of the Federal University of Bahia, Brazil. Ricardo is head of the Computational Intelligence and Optimization Research laboratory and his research interests include Signal Processing, Time Series Analysis, and Machine Learning. Besides theoretical studies, he is also interested in applying his research in, for example, volcanology, medicine, and geophysics.

**Eduardo Hruschka** is currently head of data science at Itaú Bank. From 2013 to 2016, he was chief data scientist of the startup 'Big Data'. Prior to that, he had an academic career. He received his Ph.D. degree in Computational Systems from COPPE/Federal University of Rio de Janeiro in 2001. Then he worked for several private universities until 2007, when he joined the University of São Paulo (USP), where he is currently a part-time associate professor of the Department of Computer Engineering and Digital Systems. From 2010 to 2012, he worked as a postdoctoral researcher at the University of Texas at Austin. In his academic career, he has published over than 100 scientific papers on data science, machine learning, and artificial intelligence.

**Rodrigo de Mello** is Associate Professor with the Institute of Mathematics and Computer Sciences, Department of Computer Science, University of São Paulo, São Carlos, Brazil. He completed his PhD degree from University of São Paulo, São Carlos in 2003. His research interests include the theoretical learning proofs for both Supervised and Unsupervised scenarios, the Statistical Learning Theory, Machine Learning, Applications in Dynamical Systems concepts, Time Series Analysis and Data Streams.