Signal Processing 118 (2016) 159–176

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Applying Empirical Mode Decomposition and mutual information to separate stochastic and deterministic influences embedded in signals

Ricardo Araújo Rios [a,*], Rodrigo Fernandes de Mello [b]

[a] Department of Computer Science, Federal University of Bahia, Brazil
[b] Institute of Mathematics and Computer Science, University of São Paulo, Brazil

ARTICLE INFO

Article history:
Received 27 January 2015
Received in revised form
5 May 2015
Accepted 5 July 2015
Available online 14 July 2015

Keywords:
Empirical Mode Decomposition
Stochastic
Deterministic
Mutual information
Filter bank

ABSTRACT

Empirical Mode Decomposition (EMD) is a method to decompose signals into Intrinsic Mode Functions (IMFs) to be analyzed in terms of instantaneous frequencies and amplitudes. By comparing the phase spectra of IMFs, we observed that a subset of them contains more stochastic influences while the other is predominantly deterministic. Considering this observation, we claim that IMFs can be combined to form two additive components: one deterministic and another stochastic. Having both components separated, researchers can improve data modeling as well as forecasting. In this context, this paper presents a new approach to separate deterministic from stochastic influences embedded in signals, considering the mutual information contained in phase spectra of consecutive IMFs. As previous step of this study, we also proved that EMD works as a filter bank.

1. Introduction

The Empirical Mode Decomposition (EMD) method was designed to decompose signals into components referred to as Intrinsic Mode Functions (IMFs) in order to study and analyze their instantaneous frequencies and amplitudes by using the Hilbert Spectral Analysis (HSA) [1]. Among the main advantages of EMD are the extraction of a reduced number of components in comparison to other techniques and its application to linear and nonlinear signals [1].

By analyzing the EMD results, we empirically observed that this method produces IMFs at different frequency bandwidths. This same observation also caught the attention of other researchers such as Flandrin et al. [2], who studied EMD using numerical experiments. From this empirical observation, in this paper we prove that EMD works as a filter bank. While conducting this proof and experimenting with EMD on synthetic and real-world datasets we also confirmed that EMD separates deterministic from stochastic influences embedded in signals.

While experimenting of real-world data we noticed high-frequency IMFs could be better modeled using Stochastic tools [3], while low-frequency ones could be better modeled using Dynamical System tools [4]. In fact, frequency by itself cannot be used as an indicator of stochasticity or determinism; however, we observed that the way EMD produces IMFs could help to separate deterministic from stochastic components. In order to proceed with this separation, we designed a new approach that compares the phase spectra (complex Fourier coefficients) between consecutive IMFs to measure their similarities. As a consequence, similar IMFs share some degree of information what allows us to assume them as deterministic while dissimilar ones are taken as stochastic.

This phase spectra comparison was performed using the Mutual Information method (MI), which measures the dependency between variables. Experiments performed

* Corresponding author.
E-mail addresses: ricardorios@dcc.ufba.br (R.A. Rios), mello@icmc.usp.br (R.F.d. Mello).

http://dx.doi.org/10.1016/j.sigpro.2015.07.003
0165-1684/

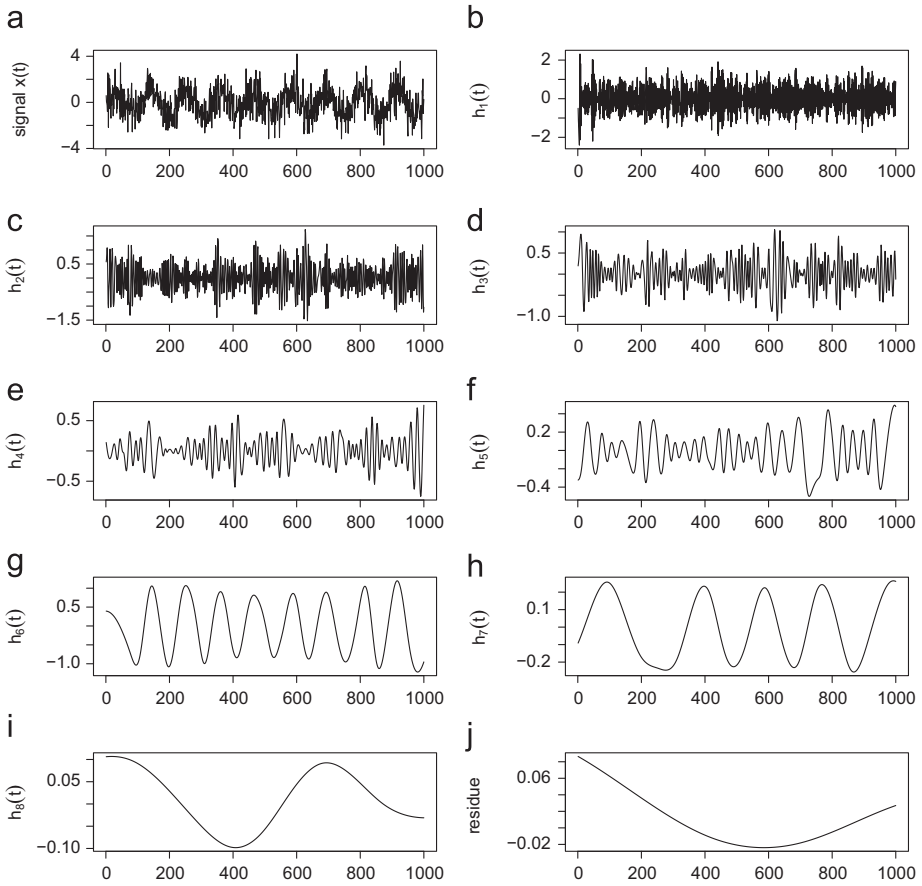

Fig. 1. The noisy signal $x(t) = \sin(2\pi t) + \varepsilon(0,1)$ is shown in plot (a). Plots from (b) to (i) show all IMFs $h_n(t)$ extracted at each iteration from the signal $x(t)$ using the EMD method. The last plot (j) shows the residue $r(t)$.

on signals with additive noise confirmed that the first high-frequency IMFs produced by the EMD method present lower levels of mutual information, i.e., they usually share almost no information; however, as next IMFs were produced (at lower frequencies), higher mutual information was observed. In this sense, when two IMFs are stochastic, no relevant information is shared between them and, consequently, their mutual information is low. On the other hand, higher levels of mutual information point out the presence of similar influences which we assume as determinism.

Considering this analysis, we state that the application of EMD on signals provides IMFs that can be separated into two classes: one with lower mutual information and another presenting higher mutual information levels. The first class corresponds to stochastic influences, whereas the second represents deterministic ones. Having the deterministic and the stochastic components separated, one can improve: (i) the modeling and prediction of each individual component [5] as well as (ii) the filtering, once noise is typically related to stochastic processes [6,7].

The remainder of this paper is organized as follows. In Section 2, we present an overview on the Empirical Mode Decomposition (EMD) method. The process of extracting components from signals, using EMD, was analyzed in Section 3 considering instantaneous frequencies and phase spectra. This analysis has motivated the development of this work, which starts discussing the Nyquist–Shannon sampling theorem in Section 4. This theorem was considered to prove EMD works as a filter bank as detailed in Section 5. Next (Section 6), we present the experimental results performed on synthetic and real-world datasets to illustrate our proof and confirm that IMFs can be combined to form two components: one stochastic and another deterministic. Finally, in Section 7, we draw conclusions and discuss future work.

2. On empirical mode decomposition

The Empirical Mode Decomposition (EMD) method supports the decomposition of signals into Intrinsic Mode Functions (IMFs) regardless of their linearity, stationarity, and stochasticity [1,2,8]. The key point to perform this decomposition is the sifting process [9,10], which initially analyzes a signal $x(t)$ and identifies local maxima and minima values of observations along time. The cubic spline method is then applied to the maxima and minima to compose the upper $u(t)$ and lower $l(t)$ envelopes, respectively [1]. The approximation values obtained using both cubic splines (upper and lower) are used to compute the mean envelope $m(t)$.

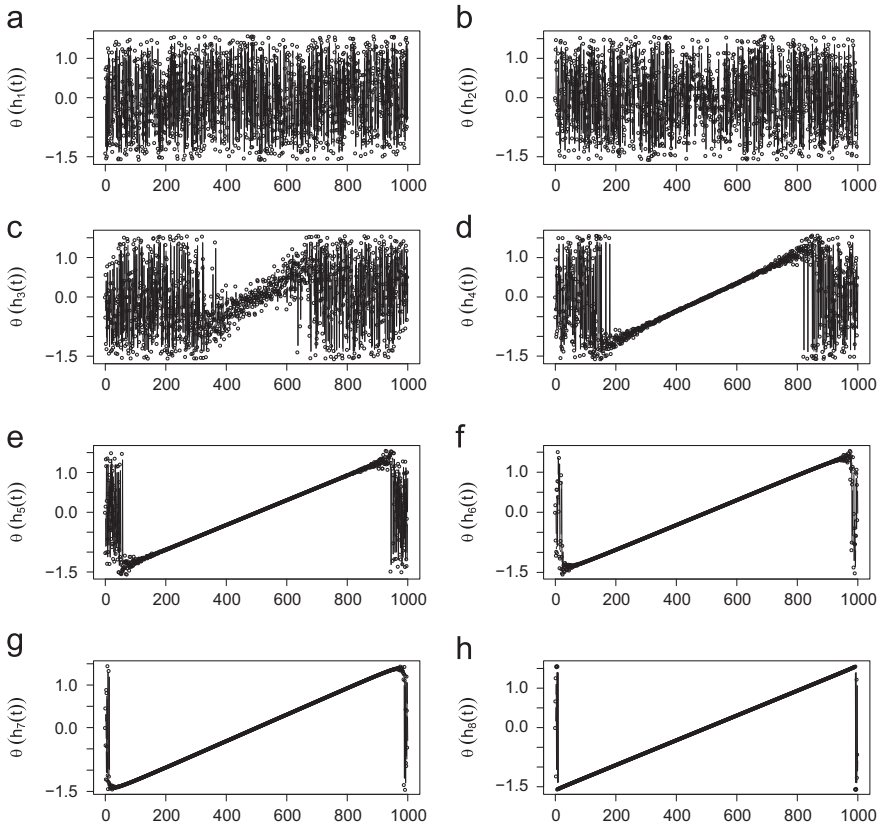

Fig. 2. The phase spectrum for the IMFs presented in Fig. 1.

Later, $m(t)$ is removed from the original signal $x(t)$, producing the first candidate component $h_{1,1}(t) = x(t) - m(t)$, in which the first index corresponds to the IMF identifier (as this is the first IMF to be extracted, the index is one) and, the second, to the candidate identifier (as this is the first candidate, the index is also one). This candidate is used in place of the original data and all sifting process is repeated until the candidate satisfies the IMF definition, which must agree with one of the following requirements: (i) the number of extrema and the number of zero-crossings must be either equal or differ at most by one or (ii) at every point, the mean $m(t)$ is zero.

After obtaining the candidate that has satisfied the IMF definition, the first IMF is defined according to $h_1(t) = h_{1,k}(t)$, assuming k candidates were produced until reaching the IMF definition. This first IMF is then removed from data, i.e., $x(t) - h_1(t)$, and the resultant signal is again analyzed by the whole process, producing further IMFs until a stop criterion is reached. This criterion usually holds when the last IMF becomes a monotonic function, avoiding the extraction of further components. Therefore, this last component is called final residue, $r(t)$ [1]. In summary, according to EMD, a signal $x(t)$ is composed of a set of IMFs plus a residue, as defined in the following equation:

$$x(t) = \sum_{j=1} h_j(t) + r(t) \tag{1}$$

In the next section, we present an analysis on IMFs extracted using the EMD method. The analysis presented in the next section follows the approach presented in [11].

3. Analyzing IMF frequencies

Aiming at better understanding the IMFs produced by EMD, let $x(t) = x(1), x(2), \ldots, x(T)$ be a signal with additive noise created by combining a sine function with angular frequency 2π plus a Normal distribution with mean $\mu = 0$ and standard deviation equals to 1. This synthetic signal is shown in Fig. 1(a). After applying EMD on $x(t)$, a set of IMFs ($\{h_j(t) | 1 \le j \le N\}$[1]) and a residue $r(t)$ were extracted, as shown in Fig. 1(b–j).

By analyzing the EMD results, we noticed that the frequency bandwidths tend to reduce as new IMFs are extracted. In order to understand this reduction, we studied IMFs in the frequency domain. In this sense, the Fourier Transform $\mathcal{F}(\cdot)$ was applied on every IMF Eq. (2), but the residue, since it only contains a trend. For every IMF $h_j(t)$, we obtain

$$C_j(t) = \mathcal{F}(h_j(t)), \tag{2}$$

which is a set of complex coefficients $C_j(t) = \{c_{j,1}, c_{j,2}, \ldots, c_{j,k}, \ldots, c_{j,T}\}$ in frequency space, such that $c_{j,k}$ is

[1] In this example, N is equal to 8.

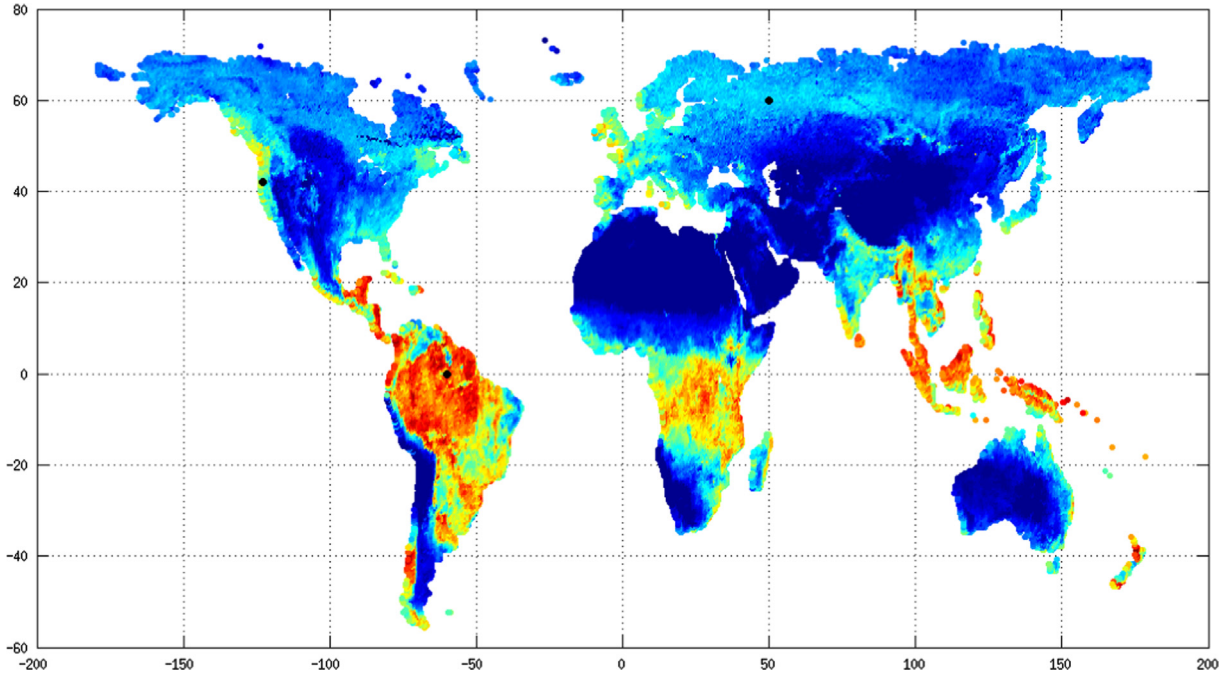

Fig. 3. FAPAR dataset: the three dots on the map represents 3 different signals used to demonstrate our proof (Section 5). (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

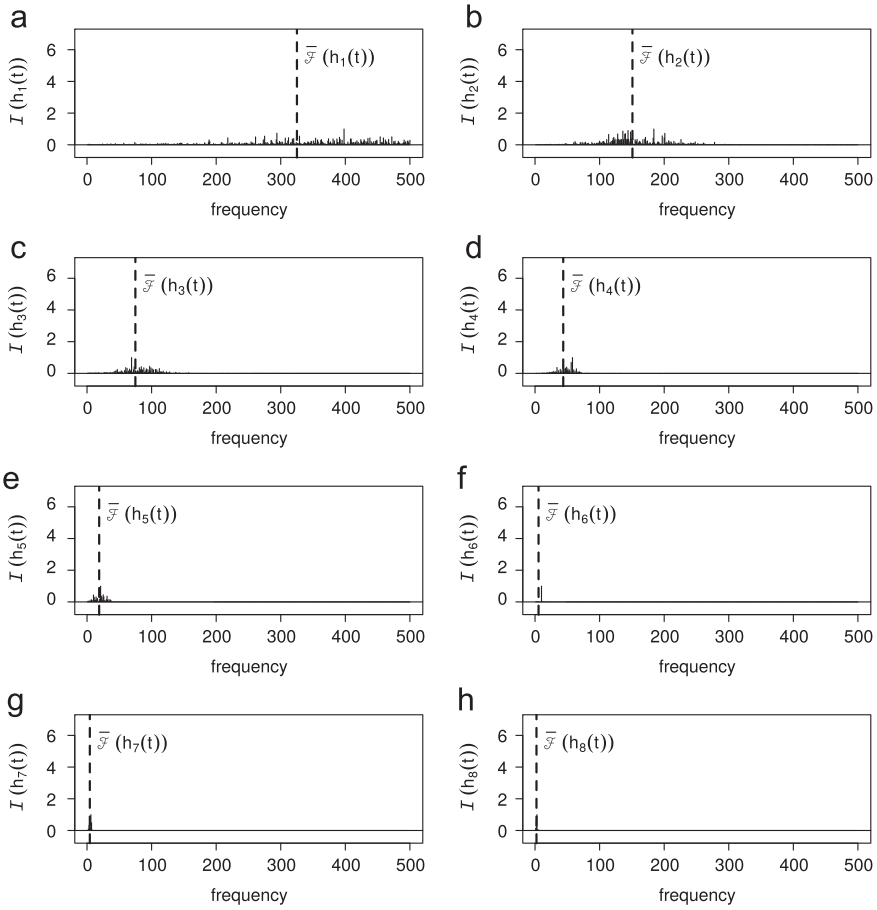

Fig. 4. Spectral density for each IMF extracted from signal *noisy-sine*. The vertical dotted line represents the weighted-mean frequency $\overline{\mathcal{F}}(\cdot)$ defined in Eq. (18).

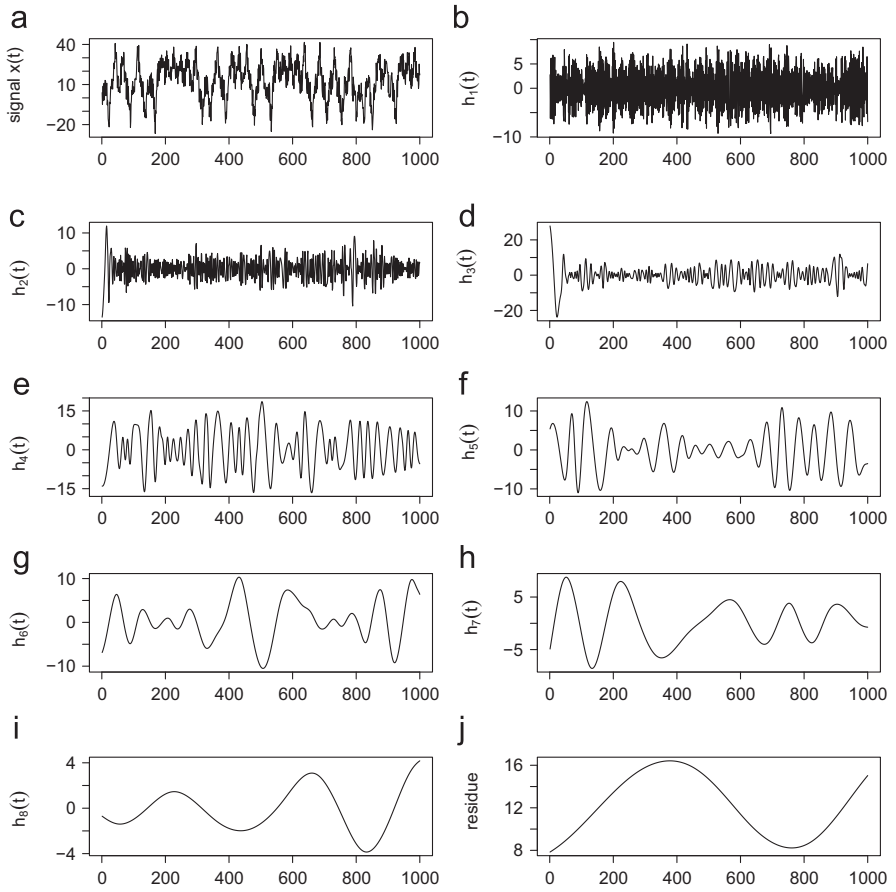

Fig. 5. Signal *noisy-lorenz* is illustrated in (a). Figures (b)–(i) show all IMFs $h_n(t)$ extracted at each EMD iteration. Last figure (j) shows residue $r(t)$.

calculated using the following equation:

$$c_{j,k} = \sum_{t=1}^{T} h_j(t) \cdot e^{-i2\pi(k/T)t}, \quad \forall k \in \{1, 2, \ldots, T\} \tag{3}$$

After obtaining the Fourier coefficients for every IMF, we calculate the phase spectrum for each component. For this purpose, the arctangent function is applied on the ratio between the imaginary $\Im(\cdot)$ and real $\Re(\cdot)$ parts of coefficients as presented in the following equation:

$$\theta(h_j(t)) = \arctan\left(\frac{\Im(C_j(t))}{\Re(C_j(t))}\right), \quad \forall j \in \{1, 2, \ldots, N\} \tag{4}$$

Fig. 2 shows phase spectra for all IMFs extracted from synthetic signal $x(t)$, which were presented in Fig. 1(b)–(i). According to this figure, we notice that the last phase spectra behaves more deterministically than the first ones.

Besides this visual inspection, we decided to employ a method to quantify the mutual information (MI) in phase spectra of consecutive IMFs. In an analysis presented in [11], it was adopted the mutual information estimator proposed by Darbellay and Vajda [12, 11], which is based on partitioning the data space into a nite number of non-overlapping rectangular cells. According to the authors, instead of partitioning data into xed-width bins, the technique must continuously partition it until achieving a conditional independence among cells. Hence, the mutual information estimated on these partitions is closer to its theoretical value.

Besides this visual inspection, we decided to employ a method to quantify the mutual information (MI) in phase spectra of consecutive IMFs [11]. The concept of mutual information (MI) was initially introduced by Shannon [12] to quantify the information shared by two variables. For example, let X and Y be two strictly independent random variables, their MI is zero because one variable cannot provide any information about the other [13].

The mutual information between continuous variables X and Y is formally presented in the following equation, in which $f_{X,Y}(x,y)$ is the joint Probability Density Function (PDF) for variables X and Y, and $f_X(x)$ and $f_Y(y)$ are the marginal PDFs of X and Y, respectively [14].

$$I(X;Y) = \int_Y \int_X f_{X,Y}(x,y) \log\left(\frac{f_{X,Y}(x,y)}{f_X(x) f_Y(y)}\right) dx\, dy \tag{5}$$

In all practical situations, one needs a (discrete) estimator for the mutual information (Eq. (5)). Most of the estimators consider three straightforward steps. First, the dataset is partitioned into fixed-width discrete bins (cells), then a histogram is created using the relative frequency of samples in each bin [14]. Finally, the mutual information is calculated on the histogram. The main problem faced by such

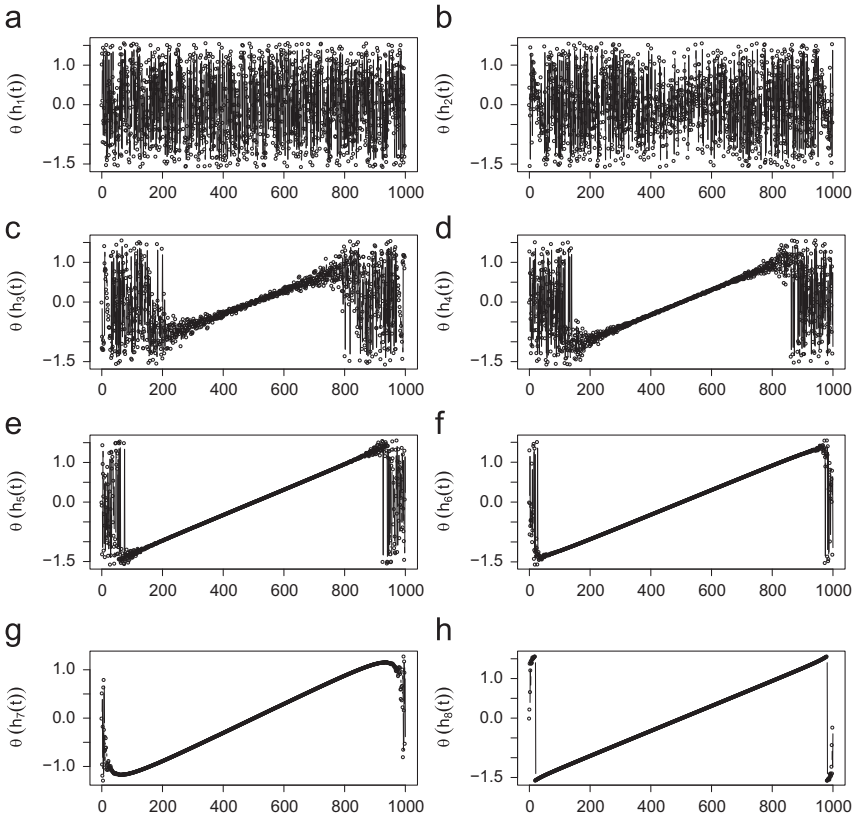

Fig. 6. The phase spectrum for every IMF extracted from signal *noisy-lorenz*.

estimators is the limitation of partitioning data into bins, which is performed using a fixed width, what produces an accumulated error.

Aiming at overcoming this drawback, Darbellay and Vajda [15] presented a new method based on partitioning the data space into a finite number of non-overlapping rectangular cells. According to the authors, instead of partitioning data into fixed-width bins, the technique must continuously partition it until achieving a conditional independence among cells. Hence, the mutual information estimated on these partitions is closer to the theoretical value.

The Darbellay and Vajda (DV) method is based on three steps [16]. Initially, data must be organized into a one-cell partition. Aiming at clarifying the presentation of the method, we consider the data is organized in two dimensions (the extension of the algorithm to higher dimensions is discussed in [16]). Then, the cell is partitioned into 2 equiprobable halves. Next, the two resultant subcells must be recursively partitioned, if the number of points in each new cell is greater than 4 and the mutual information (Eq. (5)) in each subcell takes approximately the same value. The similarity among the mutual information values in the last step is calculated using the χ^2 "goodness-of-fit" test using a five-percent significance level, as suggested in [16]. When there is no further possible partitioning, the mutual information is calculated on the final subcells.

Considering the Darbellay and Vajda (DV) method, we analyzed the phase spectra v_p of consecutive IMFs $h_p(t)$ and $h_{p+1}(t)$, such as $\forall p \in \{1, 2, \ldots, N-1\}$: $\theta(h_1(t))$ versus $\theta(h_2(t))$, $\theta(h_2(t))$ versus $\theta(h_3(t))$, …, $\theta(h_p(t))$ versus $\theta(h_{p+1}(t))$ (using the following equation):

$$v_p = I(\theta_p(t), \theta_{p+1}(t)) \tag{6}$$

As a main advantage, the use of mutual information on phase spectra allows the detection of behavior modifications in consecutive IMFs. As noticed, higher-frequency IMFs usually (but not necessarily) present lower mutual information what allows us to characterize them as stochastic, whereas mutual information increases as a next lower-frequency IMF is obtained. According to these results, deterministic IMFs present phase spectra with more similar information, once there is a strong entropy among them. The difference between frequency bandwidths among IMFs can be justified by the successive subtractions between the signal and the mean envelope, which is calculated by the local maxima and minima of signal values.

As a consequence, we state that IMFs could be decomposed into two classes: one having lower-mutual-information IMFs and another containing higher-mutual-information ones. Based on this analysis, we assume first class corresponds to stochastic influences, whereas the second represents deterministic ones. To demonstrate this decomposition, in the next section, we present the Nyquist–Shannon sampling theorem, which was considered in the formal analysis presented in this work to prove that EMD acts as a filter bank.

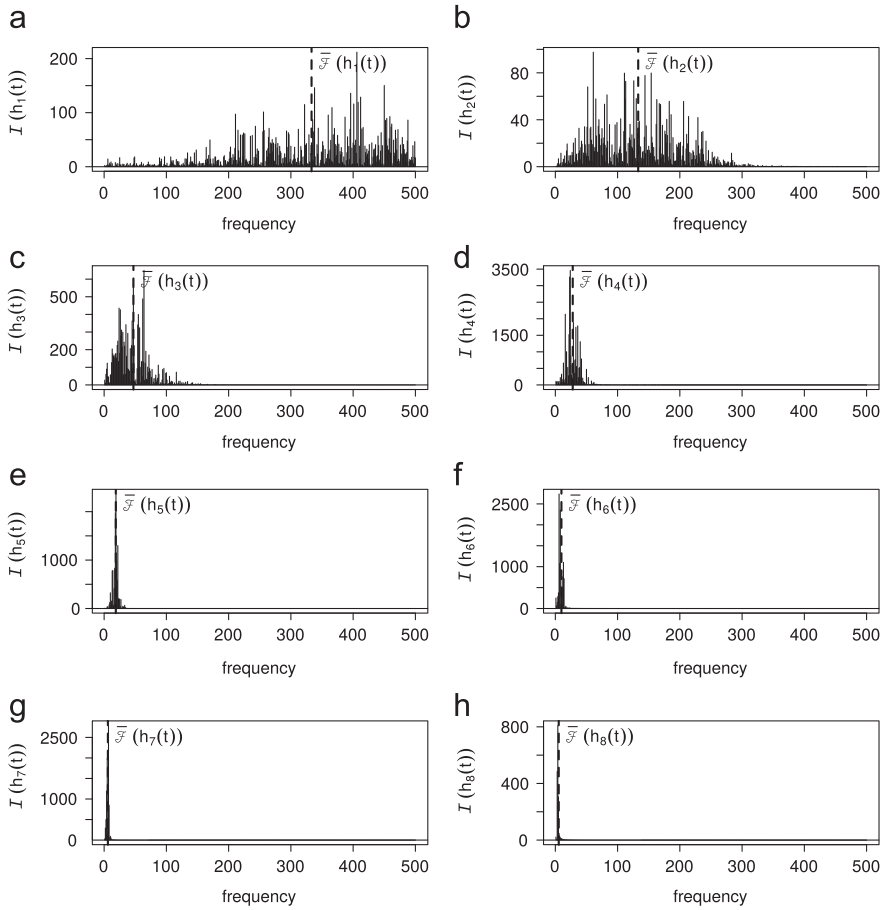

Fig. 7. Spectral density for every IMF extracted from signal *noisy-lorenz*. The vertical dotted line represents the weighted-mean frequency $\overline{\mathcal{F}}(\cdot)$ defined in Eq. (18).

Table 1
Mutual information calculated on phase spectra of consecutive IMFs extracted from signals *noisy-sine* and *noisy-lorenz*.

Mutual information	Noisy-sine	Noisy-lorenz
$I(\theta(h_1(t),\theta(h_2(t))$	0.00051	$8e-04$
$I(\theta(h_2(t),\theta(h_3(t))$	0.01452	0.02022
$I(\theta(h_3(t),\theta(h_4(t))$	0.44547	1.0797
$I(\theta(h_4(t),\theta(h_5(t))$	2.2147	2.0168
$I(\theta(h_5(t),\theta(h_6(t))$	3.3778	3.3755
$I(\theta(h_6(t),\theta(h_7(t))$	4.0156	3.9835
$I(\theta(h_7(t),\theta(h_8(t))$	4.0954	3.8857

Table 2
Weighted-mean frequencies for the synthetic signals.

IMFs	Noisy-sine	Noisy-lorenz
$\overline{\mathcal{F}}(h_1(t))$	325.08	333.2
$\overline{\mathcal{F}}(h_2(t))$	150.84	133.1
$\overline{\mathcal{F}}(h_3(t))$	74.88	47.06
$\overline{\mathcal{F}}(h_4(t))$	43.67	27.95
$\overline{\mathcal{F}}(h_5(t))$	18.82	18.78
$\overline{\mathcal{F}}(h_6(t))$	5.45	9.8
$\overline{\mathcal{F}}(h_7(t))$	4.34	6.05
$\overline{\mathcal{F}}(h_8(t))$	2.23	5.47

4. On the Nyquist–Shannon sampling theorem

According to the Nyquist–Shannon sampling theorem [17], a band-limited signal $x(t)$ can be correctly reconstructed from an infinite sequence of samples. A signal is considered band-limited when its frequencies are restricted to a bandwidth of B hertz, determined by the difference between the highest (f_h) and lowest (f_l) frequencies, i.e., $B=f_h-f_l$.

In order to reconstruct a signal $x(t)$, the sampling frequency f_s should be greater than twice the highest frequency, $f_s>2f_h$ [17,18]. By assuming the lowest frequency is equal to zero, the bandwidth can be rewritten as $B=f_h$ and the sampling frequency is commonly defined as $f_s>2B$. The Nyquist frequency B and rate f_s are defined as $f_s/2$ and $2B$, respectively; therefore, a band-limited signal can be correctly reconstructed when the sampling rate has exceeded f_s.

In order to formally describe the Nyquist–Shannon sampling theorem, consider the Fourier transform $X(f)$ for a continuous signal $x(t)$, as defined in the following equation:

$$X(f)=\int_{-\infty}^{\infty} x(t)\, e^{-2\pi itf}\, dt \tag{7}$$

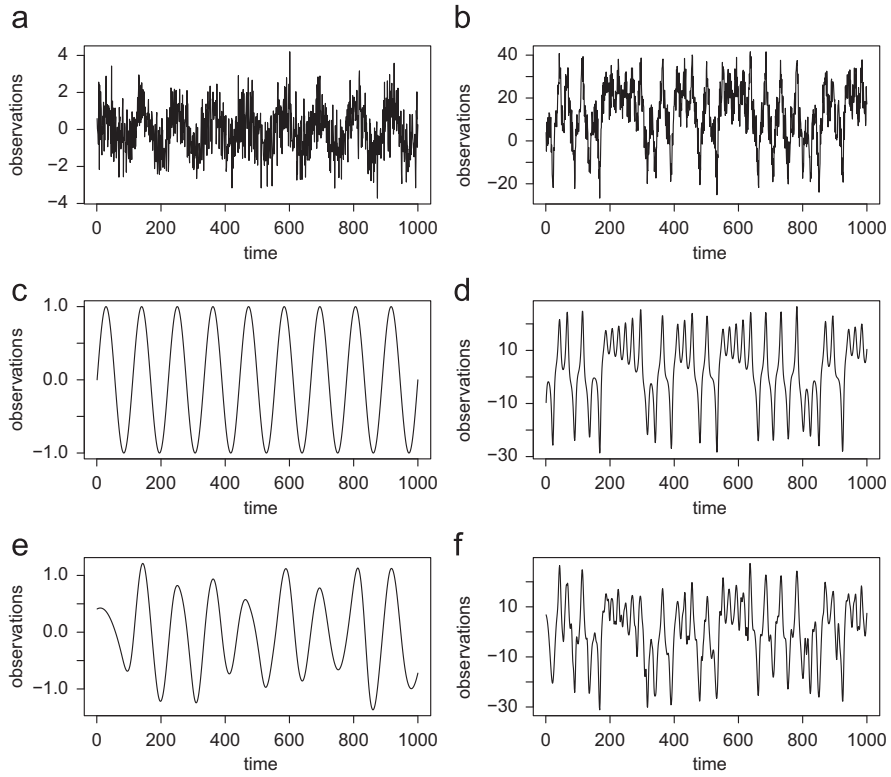

Fig. 8. Figures (a) and (b) illustrate signals *noisy-sine* and *noisy-lorenz* respectively. Figures (c) and (d) show the expected deterministic behavior for the sine and lorenz signals (without noise), respectively. Figures (e) and (f) show the estimated deterministic components for both signals after EMD decomposition and the use of Mutual Information to separate them.

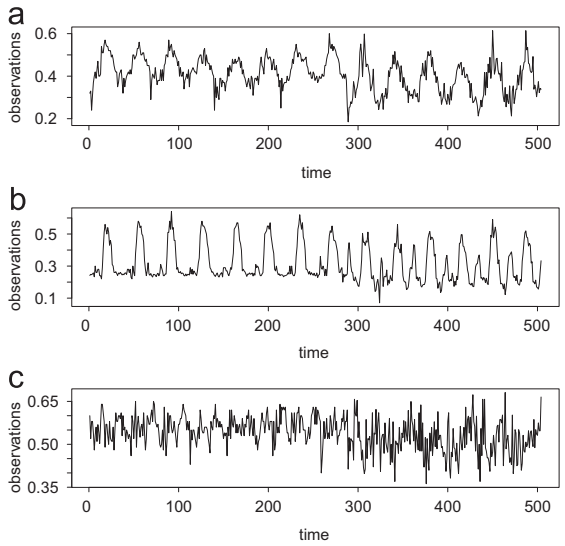

Fig. 9. FAPAR signals considered in experiments. They correspond to the following coordinates: (a) [−123.25, 42.75]; (b) [50.25, 60.75]; (c) [−60.75, 0.25].

According to the Nyquist–Shannon sampling theorem, a signal $x(t)$ is sufficiently reconstructed if time-spaced observations are collected at regular and constant time intervals, as defined in Eq. (8), in which f is the maximum represented frequency when observations are collected at T time intervals.

$$T \stackrel{\text{def}}{=} \frac{1}{f} \tag{8}$$

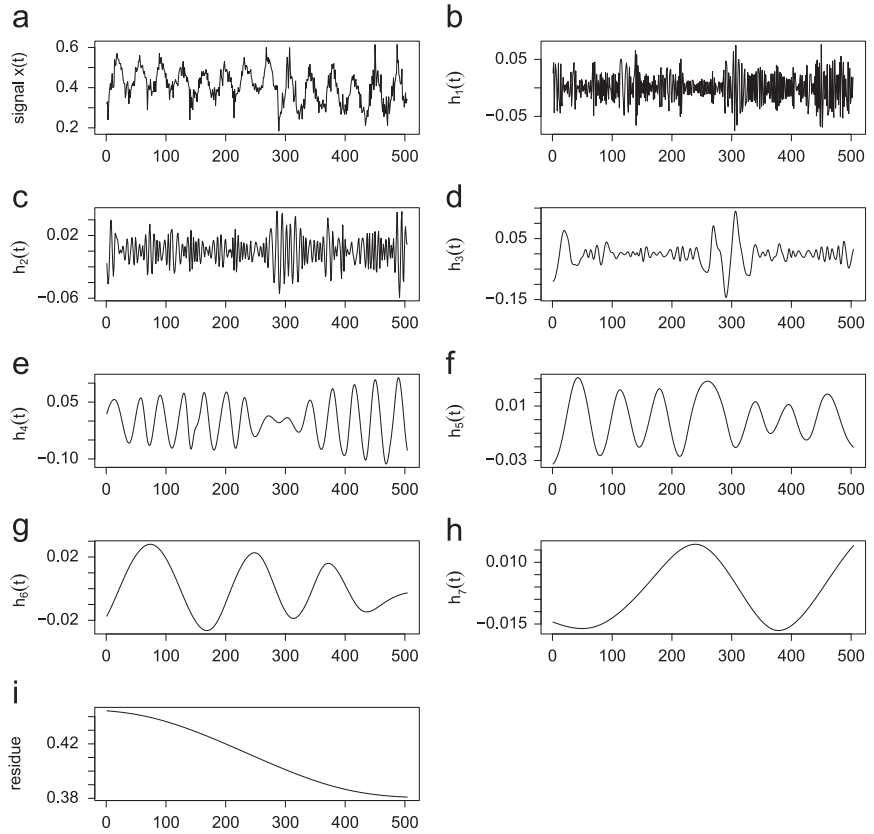

Fig. 10. Signal *fapar-north* (Fig. 9(a)) is shown in plot (a). Plots from (b) to (i) show all IMFs $h_n(t)$ extracted along EMD iterations. The last plot (j) shows residue $r(t)$.

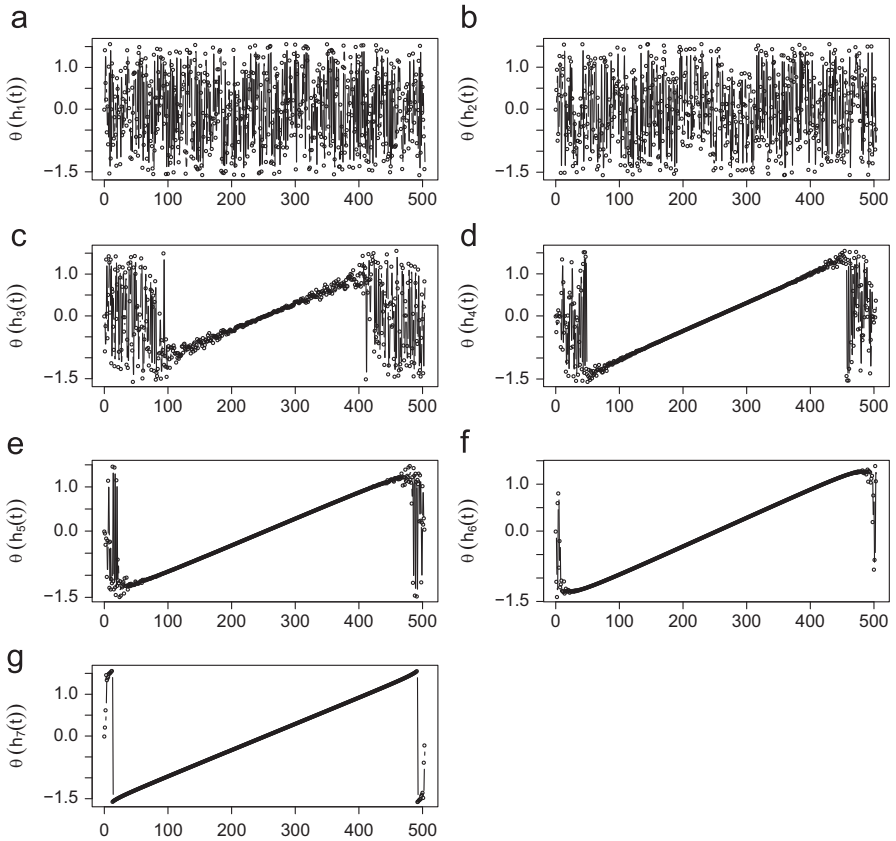

Fig. 11. The phase spectra for all IMFs extracted from signal *fapar-north*.

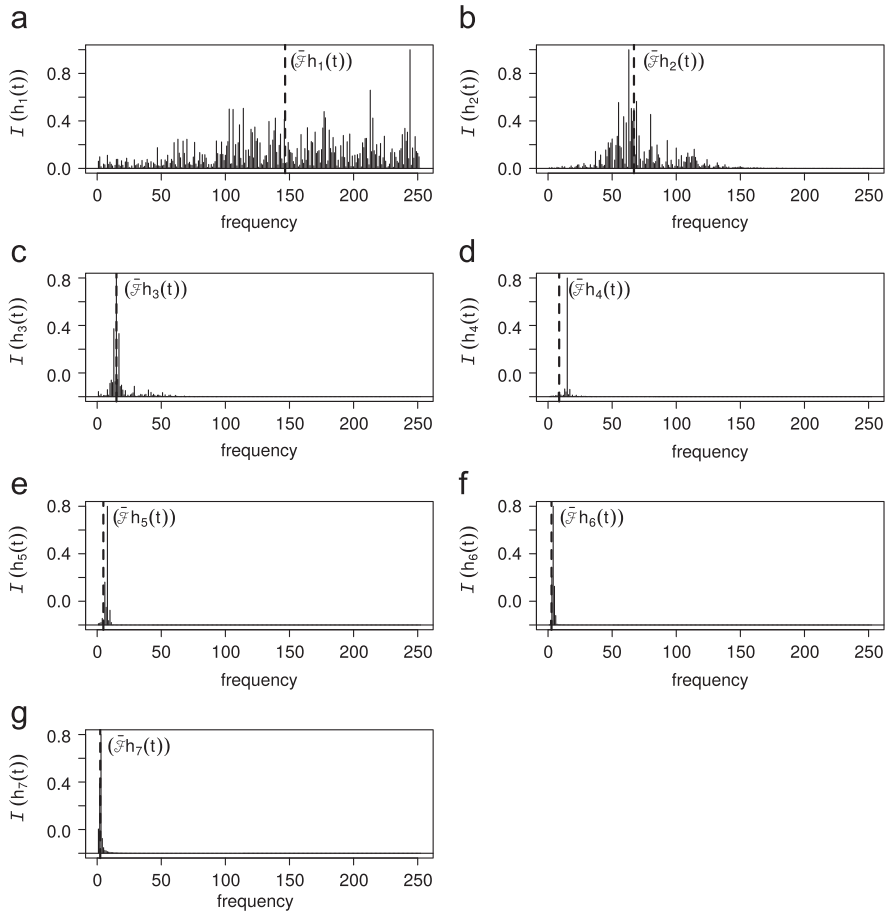

Fig. 12. Spectral density for every IMF extracted from signal *fapar-north*. The vertical dotted line represents the weighted-mean frequency $\overline{\mathcal{F}}(\cdot)$ defined in Eq. (18).

When interpolating any signal $x(t)$, a subset of observations $D(x(t))$ of $x(t)$ are selected. Assuming $x(t)$ is continuous, and therefore, has an infinite number of observations, $D(x(t))$ has a finite number of observations, which is used to construct new data points within the time range of $x(t)$. As a consequence of the Nyquist–Shannon sampling theorem, by increasing the time interval between consecutive observations, the interpolation of $D(x(t))$ will represent lower frequencies than those of $x(t)$. This statement can be verified by Eq. (8), i.e., once the sampling interval increases the maximal signal frequency is reduced.

5. Analyzing frequencies of intrinsic mode functions

We start our formal approach to show EMD works as a filter bank by remembering our observation that a next IMF contains lower frequencies than previous ones. To clarify it, consider a signal $x(t)$ composed of j IMFs, $\{h_1(t), h_2(t), \ldots, h_j(t)\}$, which were extracted using EMD. Let $\mathcal{F}_{\max}(\cdot)$ be a function that calculates the maximum frequency of signals. We can define the following hypothesis: $\mathcal{F}_{\max}(h_1(t)) > \mathcal{F}_{\max}(h_2(t)) > \cdots > \mathcal{F}_{\max}(h_{j-1}(t)) > \mathcal{F}_{\max}(h_j(t))$, i.e., the maximum frequency of a next IMF is always lower than the previous one.

By proving this hypothesis, we can demonstrate that the EMD sifting process acts as a filter, removing components according to their frequencies. After this result, we can apply the mutual information (MI) method to separate deterministic from stochastic influences embedded in each IMF. Furthermore, we observed deterministic influences increase as the frequency is reduced; this happens because sinusoidal characteristics are reinforced by EMD as next IMFs are obtained. Next, we start defining the following theorem:

Theorem 1. *If the maximum frequency of a next Intrinsic Mode Function produced by EMD is lower than the previous one,* $\mathcal{F}_{\max}(h_{j-1}(t)) > \mathcal{F}_{\max}(h_j(t))$*, then EMD works as a filter bank.*

Proof. Considering the sifting process (presented in Section 2), the first EMD step is performed by calculating the signal envelopes considering the following definitions:

Definition 1 (*Envelopes*). First, the signal extrema are selected. Then, all maxima are connected by using the cubic spline function, producing the upper envelope $u(t)$

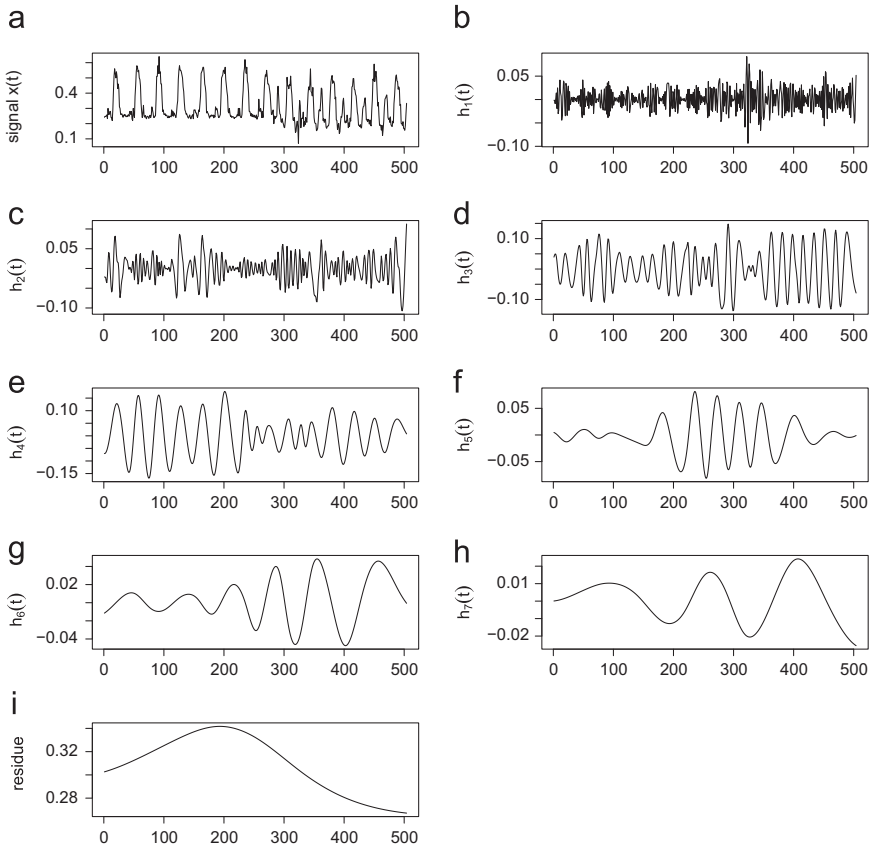

Fig. 13. Signal *fapar-asia* is shown in plot (a). Plots from (b) to (i) show all IMFs $h_n(t)$ extracted along EMD iterations. The last plot (j) shows residue $r(t)$.

(Eq. (9)). The same occurs with all minima connected to form the lower envelope $l(t)$ (Eq. (10)).

$$u(t) = a'x(t)^3 + b'x(t)^2 + c'x(t) + d' \tag{9}$$

$$l(t) = a''x(t)^3 + b''x(t)^2 + c''x(t) + d'' \tag{10}$$

Next, the sifting process estimates the mean envelope, as shown in the following definition:

Definition 2 (*Mean envelope*). The mean envelope is calculated by the following equation:

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

$$m(t) = \frac{[a'x(t)^3 + b'x(t)^2 + c'x(t) + d'] + [a''x(t)^3 + b''x(t)^2 + c''x(t) + d'']}{2}$$

$$m(t) = \frac{(a'+a'')}{2}x(t)^3 + \frac{(b'+b'')}{2}x(t)^2 + \frac{(c'+c'')}{2}x(t) + \frac{(d'+d'')}{2} \tag{11}$$

Remark 1. According to Eq. (11), the mean envelope is still considered a cubic spline function, once its properties and characteristics have been preserved.

After calculating the mean envelope, the next step is to remove it from signal $x(t)$, producing an IMF candidate $h_{1,1}(t)$ (the first index represents the IMF identifier and the second corresponds to the candidate identifier). This IMF candidate is used in place of $x(t)$ and all previous steps are then repeated (Eq. (12)) until reaching a stop criterion, as shown in Section 2.

$$\begin{aligned} h_{1,1}(t) &= x(t) - m_1(t) \\ h_{1,2}(t) &= h_{1,1} - m_2(t) \\ &\vdots \\ h_{1,k}(t) &= h_{1,k-1} - m_k(t) \end{aligned} \tag{12}$$

Therefore, considering one of the IMF requirements was reached after k iterations, the first real IMF is obtained by the following equation:

$$h_1(t) = h_{1,k}(t) \tag{13}$$

Definition 3 (*Intrinsic Mode Function*). IMF can be defined in terms of intermediary mean envelopes as shown in Eq. (14). In such equation, we only formalized how the first IMF $h_1(t)$ was obtained, but the remaining ones are produced by using the same steps (the qth IMF, i.e., $h_q(t)$, is obtained by replacing the original signal $x(t)$ by $x(t) - \sum_{i=1}^{q-1} h_i(t)$, as discussed in Section 2).

$$h_1(t) = x(t) - m_1(t) - m_2(t) - m_3(t) - \cdots - m_k(t)$$

$$h_1(t) = x(t) - \sum_{i=1}^{k} m_i(t) \tag{14}$$

The next step is the analysis of the definition resulting from the Nyquist–Shannon sampling theorem, shown in

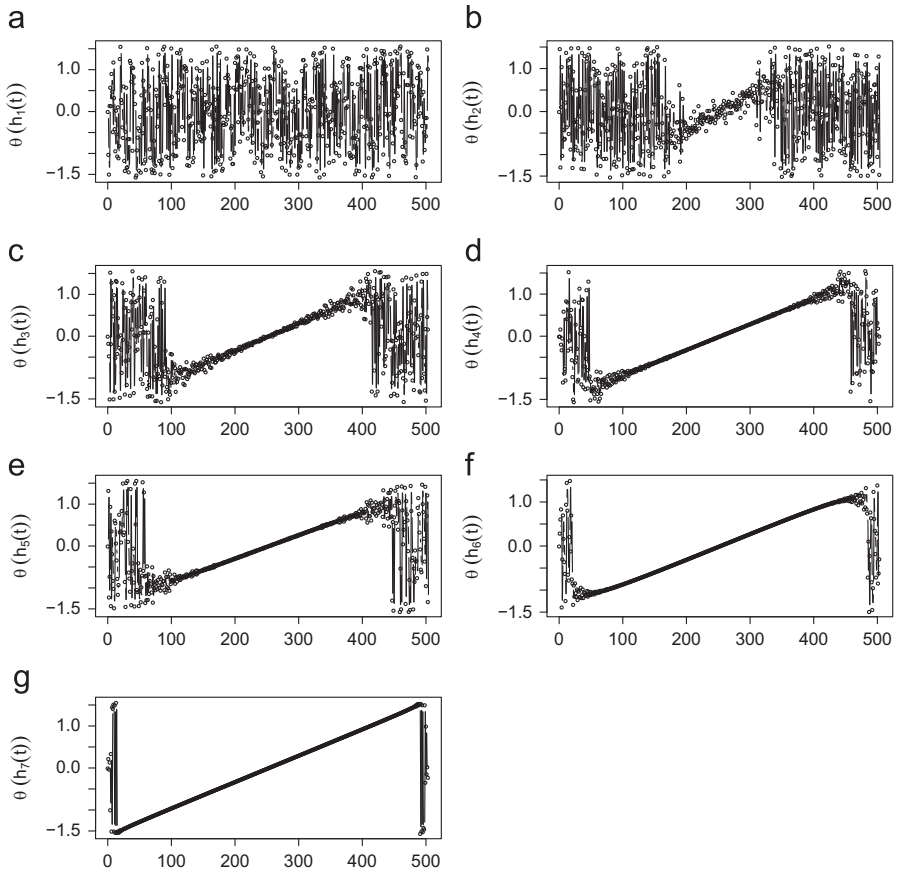

Fig. 14. The phase spectra for all IMFs extracted from signal *fapar-asia*.

Eq. (8). This definition is used to establish a relationship between the number of observations selected to compose the extrema and the signal frequency. By selecting a subset of observations to represent the minima and maxima values, the average time interval ΔT between consecutive observations increases in relation to the original signal, reducing the maximum frequency. This behavior is reflected on the mean envelope, whose maximum frequency is lower than that of the original signal.

In order to prove this assumption, consider $\Delta T_{x(t)}$ and $\Delta T_{m_1(t)}$ the time intervals for the original signal $x(t)$ and the first mean envelope $m_1(t)$, which is an interpolation based on a subset of the original observations in $x(t)$. The relation between these intervals is expressed in the following equation:

$$\Delta T_{x(t)} < \Delta T_{m_1(t)} \tag{15}$$

Considering the Nyquist–Shannon definition, in which the time interval ΔT and frequency f in the signals are inversely proportional, we extended Eq. (15) into the following equation:

$$\frac{1}{f_{x(t)}} < \frac{1}{f_{m_1(t)}} \quad \text{thus } f_{x(t)} > f_{m_1(t)} \tag{16}$$

Therefore, the frequency of the first mean envelope ($m_1(t)$ in Eq. (12)) is lower than that of the original signal $x(t)$. In the next sifting step, the signal is replaced by the first IMF candidate, $h_{1,1}(t) = x(t) - m_1(t)$, and the frequencies of the next mean envelope, $m_2(t)$, will be even lower than $m_1(t)$. Thus, the frequency of the first IMF ($h_1(t)$, in Eq. (14)) will be greater than $m_1(t)$ and greater than every mean envelope $m_i(t)$, $\forall i = 1, \ldots, k$. The next step in the EMD process is performed by removing the first IMF from the signal, i.e., $x(t) - h_1(t)$, and the resultant signal is again analyzed by the whole process, producing further IMFs until reaching the stop criterion. Therefore, the maximum frequency of the next IMF will be lower than the previous one. As a result, we state EMD works as a filter bank, removing components from high to low frequencies.□

After concluding a next IMF produced by EMD always contains lower frequency, we employed this method on synthetic and real-world datasets. As part of the experimental results, we also observed that the first IMFs usually share less information while the last ones are more similar; this is the case when noise is present (if no noise is present, then even the first IMFs will share enough information). That motivated us to design a new approach to separate deterministic and stochastic IMFs by computing their mutual information (MI). As a next step, IMFs sharing information above a given threshold are added to produce a deterministic component, while the others are added to form the stochastic one.

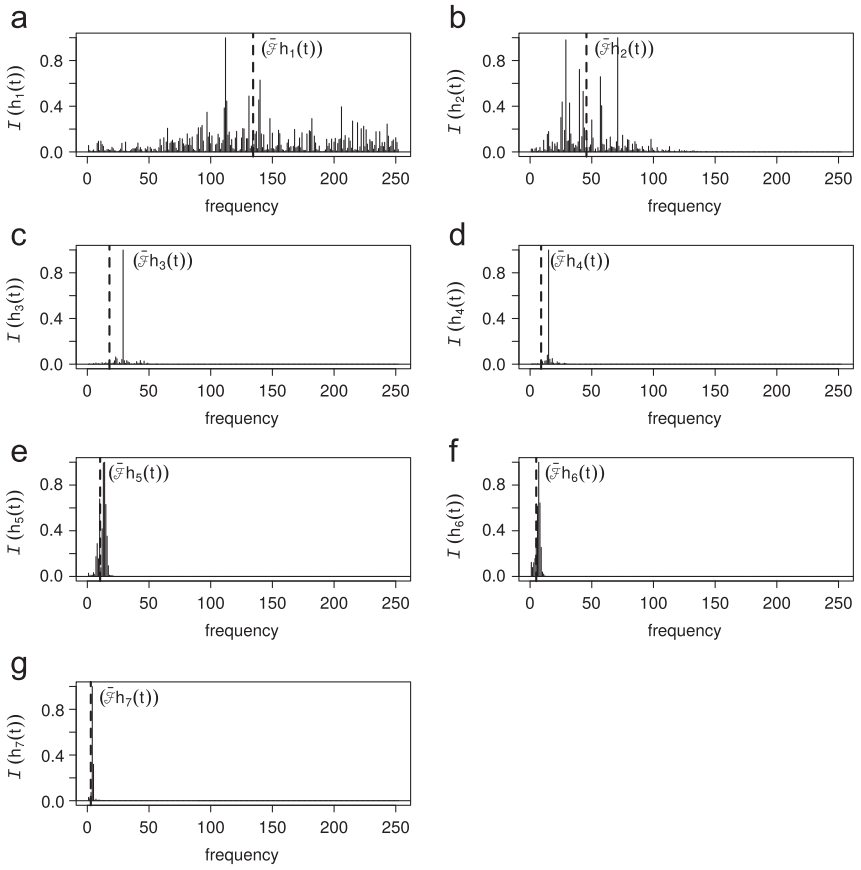

Fig. 15. Spectral density for every IMF extracted from signal *fapar-asia*. The vertical dotted line represents the weighted-mean frequency $\overline{\mathcal{F}}(\cdot)$ defined in Eq. (18).

6. Simulation results

This section presents experimental results performed on synthetic and real-world scenarios to illustrate the proof that EMD acts as a filter bank. Besides confirming our proof, the experiments presented in this section also confirm that IMFs can be combined to form two components: one stochastic and another deterministic.

In the first experimental scenario, two signals were produced using: (i) a sine function with angular frequency 2π plus a Normal distribution with mean $\mu = 0$ and standard deviation equals to 1 (referred to as *noisy-sine*) and (ii) a Lorenz system with parameters $\sigma = 10$, $\rho = 28$ and $\beta = 8/3$, producing a chaotic signal, plus a Uniform distribution on the interval from 0 to 10 (referred to as *noisy-lorenz*). The sine function and the Lorenz system were used to represent the deterministic influences, whereas the Normal and Uniform distributions represent the stochastic ones.

Our proof was also demonstrated in a real-world scenario, in which we analyzed the fraction of absorbed photosynthetically active radiation (FAPAR[2]). The dataset used in this scenario is one of the Land Surface products provided by the Institute for Environment and Sustainability (IES) of the European Commission's Joint Research Centre (JRC). In summary, this data represents the fraction of the incoming solar radiation between 400 and 700 nm which is absorbed by plants and other photosynthetic organisms on the earth surface [19–21,11].

The FAPAR dataset was compiled from two satellite sensors (SeaWifs onboard OrbitalViewer-2 and MERIS onboard Envisat) for the period of January 1998 to December 2011 at a spatial resolution of $0.5^\circ \times 0.5^\circ$. The plot of this dataset forms a map of 51,912 pixels, in which each pixel contains a signal. Each signal observation is a spatial and temporal average value for a ten-day period. Thus, the dataset considered in this work contains 168 monitored months, with 3 observations per month, producing a signal with 504 observations for every pixel in the 2D World map [19–21,11].

Aiming at better understanding this dataset, Fig. 3 shows the resulting global 2D World map for a time instant of FAPAR. The indices in this plot range from 0 to 1, which are represented by a color scale from blue to red, respectively. In the experiments presented in this section, we selected only 3 signals from FAPAR for two reasons: (i) due to space restrictions in this paper and (ii) our goal is not to draw conclusion about the whole dataset, but to demonstrate signals that can be decomposed into stochastic and deterministic components. The selected signals are also shown in

[2] Further details about the dataset and data processing can be found at http://fapar.jrc.ec.europa.eu/Home.php.

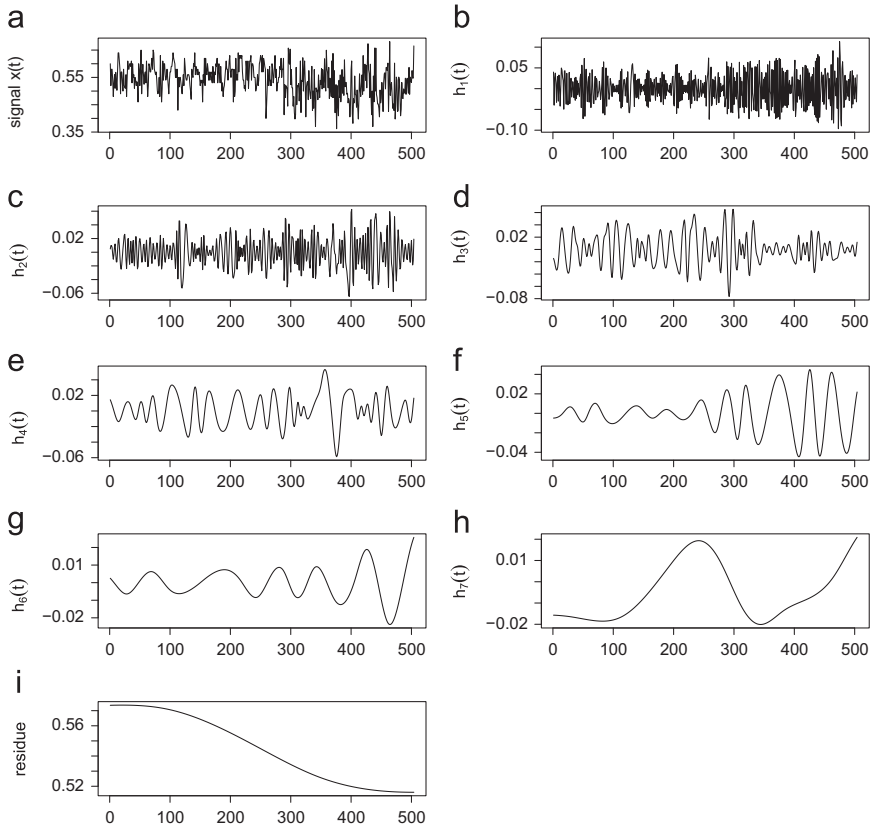

Fig. 16. Signal *fapar-south* is shown in plot (a). Plots from (b) to (i) show all IMFs $h_n(t)$ extracted along EMD iterations. The last plot (j) shows the residue $r(t)$.

Fig. 3 by three small dots located at South and North America, and Asia. These signals are referred, in this section, to as *farpar-south*, *farpar-north*, and *farpar-asia*, correspondingly.

We start analyzing the signal *noisy-sine*. Hence, in the first step, the signal was decomposed by the EMD method, producing a set of IMFs $h_j(t)$ and a residue, as previously presented in Fig. 1. Next, we analyzed the phase spectrum for every extracted IMF as illustrated in Fig. 2, observing that frequency is reduced as new IMFs are extracted. Then, we analyzed each decomposed IMF calculating their spectral densities [22] according to Eq. (17), in which $X(h_q(t))$ corresponds to the Fourier coefficients for the qth IMF (Eq. (7)).

$$\mathcal{I}(h_q(t)) = |X(h_q(t))|^2 \tag{17}$$

In order to illustrate our proof, we compare the maximum frequencies between consecutive IMFs. This comparison is performed analyzing how frequency densities change between consecutive IMFs using the weighted-mean frequency $\overline{\mathcal{F}}(\cdot)$ expressed in Eq. (18), which considers frequency $(f, \forall f \in \{1, 2, \ldots, t/2\})$ and spectrum $(\mathcal{I}(\cdot))$ values obtained using the spectral densities. In this equation, the spectral densities were normalized in range $[0, 1]$.

$$\overline{\mathcal{F}}(h_q(t)) = \frac{\sum_{n=1}^{t} f \cdot \mathcal{I}(h_q(n))}{1 + \sum_{n=1}^{t} \mathcal{I}(h_q(n))} \tag{18}$$

Fig. 4 shows the spectral densities and the weighted-mean frequencies $\overline{\mathcal{F}}(\cdot)$ for each IMF obtained for signal *noisy-sine*. The results shown in this figure emphasize that the maximum frequency is reduced as new IMFs are iteratively extracted.

Similar to the previous experiment, we also analyzed signal *noisy-lorenz*. As a first step, the signal was decomposed using EMD. The obtained IMFs and residue are shown in Fig. 5.

Secondly, we analyzed the phase spectrum for every IMF as presented in Fig. 6. As the previous result, this figure also emphasizes that there is a frequency reduction as new IMFs are extracted.

Finally, the spectral density for every decomposed IMF was estimated as shown in Fig. 7. By analyzing this figure, we also observe that the maximum frequency was reduced as new IMFs were iteratively extracted.

Tables 1 and 2 summarize the results obtained for the synthetic signals. The first table shows the Mutual information values calculated on phase spectra of consecutive IMFs and the last one sums up the weighted-mean frequencies for IMFs.

To emphasize the frequency reduction along IMF extraction, we performed a complementary experiment on these noisy signals as shown in Fig. 8. In this experiment, we selected a cut-off point, in which all IMFs before this point were added to form the stochastic component and the remaining IMFs plus the residue were added to form the deterministic one. For signal *noisy-sine*, the deterministic component was created selecting all IMFs that share more information which are IMFs from 5 to 8, plus the residue (see Mutual Information values in Table 1). IMFs from 1 to 4

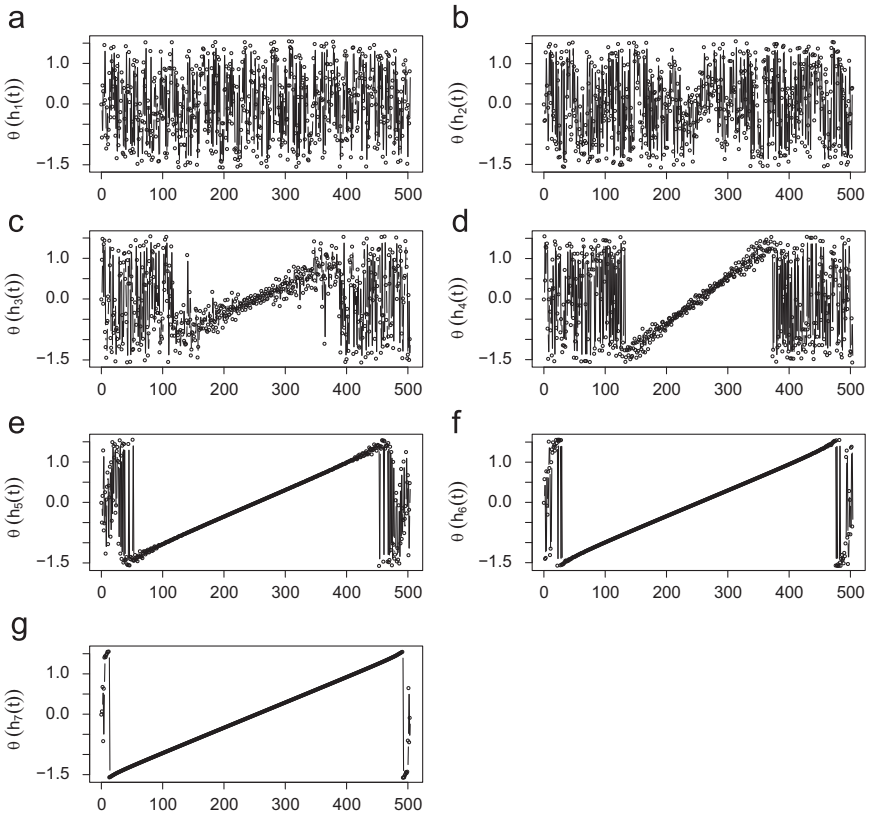

Fig. 17. The phase spectra for all IMFs extracted from signal *fapar-south*.

are added to form the stochastic component. Similarly, a deterministic component was also extracted from signal *noisy-lorenz* considering IMFs from 3 to 8, plus the residue. In this second signal, IMFs from 1 to 2 were added to produce the stochastic component.

According to this figure, the deterministic component (Fig. 8(e)) extracted from signal *noisy-sine* (Fig. 8(a)) is closer to the expected behavior as illustrated in Fig. 8(c). Similarly, the deterministic component (Fig. 8(f)) extracted from signal *noisy-lorenz* (Fig. 8(b)) is closer to the expected behavior shown in Fig. 8(d).

The next experiment was performed on a real-world scenario. As previously presented, three signals (Fig. 9) were randomly selected from the FAPAR dataset and analyzed considering the same steps applied to the synthetic scenarios.

The steps used to analyze these three signals were the same as applied on synthetic data. In this sense, we first analyzed signal *fapar-north* by decomposing it into a set of IMFs plus a residue as shown in Fig. 10.

Then, the phase spectra of all IMFs were plotted in Fig. 11 and, by analyzing them, we confirm that the maximum frequency reduces as new IMFs are extracted from the original signal.

The previous conclusion is confirmed by analyzing the spectral density for every decomposed IMF (Fig. 12), i.e., the maximum frequency is reduced as new IMFs were iteratively extracted.

The second real-world signal analyzed in this section was *fapar-asia* signal. After applying the EMD method on this time series, we obtained the IMFs presented in Fig. 13.

The phase spectra for all IMFs were plotted in Fig. 14, which confirmed the frequency reduction for consecutive IMFs.

The frequency reduction discussed in previous plot can also be highlighted by analyzing the spectral density for every decomposed IMF as illustrated in Fig. 15.

The last real-world signal was analyzed considering the same steps adopted for all results presented so far. Firstly, signal *fapar-south* was decomposed into a set of 7 IMFs plus a residue as illustrated in Fig. 16.

Secondly, we analyzed the phase spectrum and the spectral density for every IMFs extracted from this signal, as shown in Figs. 17 and 18, respectively. According to the results, we notice the maximum frequencies reduces as new IMFs are extracted.

The results presented in Figs. 17 and 18 were also summarized in Tables 3 and 4, respectively. There is a phenomenon that may happen along the extraction of IMFs, as we can see in Fig. 13(e) and (f), which is called *mode mixing* [23]. Mode mixing usually happens when two IMFs contain very close maximum frequencies, thus the decomposition may produce two IMFs when it should produce only one. Our proof does not consider this situation what makes the weighted-mean frequencies between IMFs $h_4(t)$ and $h_5(t)$ of signal *fapar-asia* present some

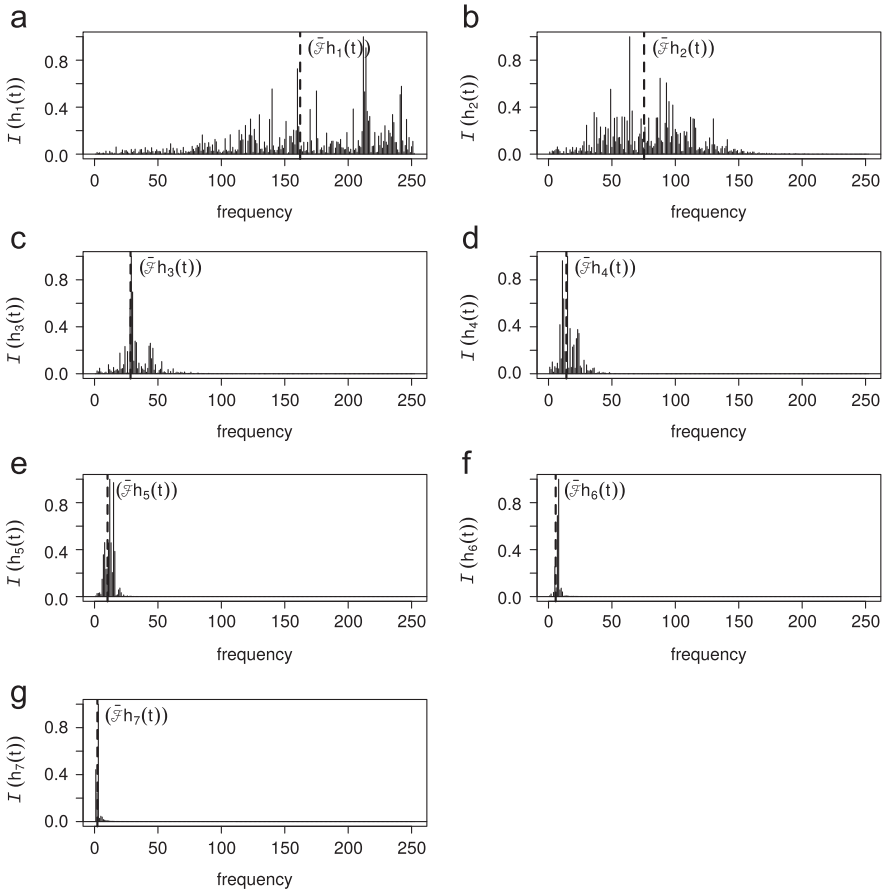

Fig. 18. Spectral density for every IMF extracted from signal *fapar-south*. The vertical dotted line represents the weighted-mean frequency $\overline{\mathcal{F}}(\cdot)$ defined in Eq. (18).

Table 3
Mutual information calculated on phase spectra of consecutive IMFs extracted from signals *fapar-north*, *fapar-asia*, and *fapar-south*.

Mutual information	fapar-north	fapar-asia	fapar-south
$I(\theta(h_1(t),\theta(h_2(t))$	0.00028	0.00533	0.00255
$I(\theta(h_2(t),\theta(h_3(t))$	$5e-04$	0.43647	0.00154
$I(\theta(h_3(t),\theta(h_4(t))$	1.4381	1.5678	0.52888
$I(\theta(h_4(t),\theta(h_5(t))$	2.6853	2.0431	1.1256
$I(\theta(h_5(t),\theta(h_6(t))$	2.7992	2.126	2.8967
$I(\theta(h_6(t),\theta(h_7(t))$	3.4648	2.8061	3.6727

Table 4
Weighted-mean frequencies for the real-world datasets.

IMFs	fapar-north	fapar-asia	fapar-south
$\overline{\mathcal{F}}(h_1(t))$	146.58	134.38	162.13
$\overline{\mathcal{F}}(h_2(t))$	67.01	45.69	75.37
$\overline{\mathcal{F}}(h_3(t))$	15.05	17.93	28.4
$\overline{\mathcal{F}}(h_4(t))$	8.69	8.95	14.1
$\overline{\mathcal{F}}(h_5(t))$	4.83	10.38	10.26
$\overline{\mathcal{F}}(h_6(t))$	2.7	4.98	5.66
$\overline{\mathcal{F}}(h_7(t))$	2.4	2.81	2.25

increase. We assume this is not a problem with our proof, but an EMD issue when extracting IMFs at closer frequencies.

Finally, Fig. 19 shows the deterministic components obtained from the real-world signals after combining the respective sets of deterministic IMFs. The IMFs used to create these three components were selected considering the phase spectra and the spectral density for every signal. As the real-world signals have similar behavior, the deterministic components were created considering IMF 2 as the cut-off point, i.e., all IMFs whose identifier is greater than 2 were taken to form the respective deterministic component.

The experiments presented in this section showed that the weighted-mean frequency reduces as further IMFs are extracted, confirming the theorem proved in the previous section. In addition, we also decomposed the original signals into deterministic and stochastic components considering the Mutual Information method applied on phase spectra of consecutive IMFs. We claim that this approach is useful to improve modeling of real-world data as confirmed in [5].

7. Conclusion

This paper presented a proof of the Empirical Mode Decomposition method works as a filter bank and supports the separation of stochastic and deterministic influences, based on the mutual information of consecutive

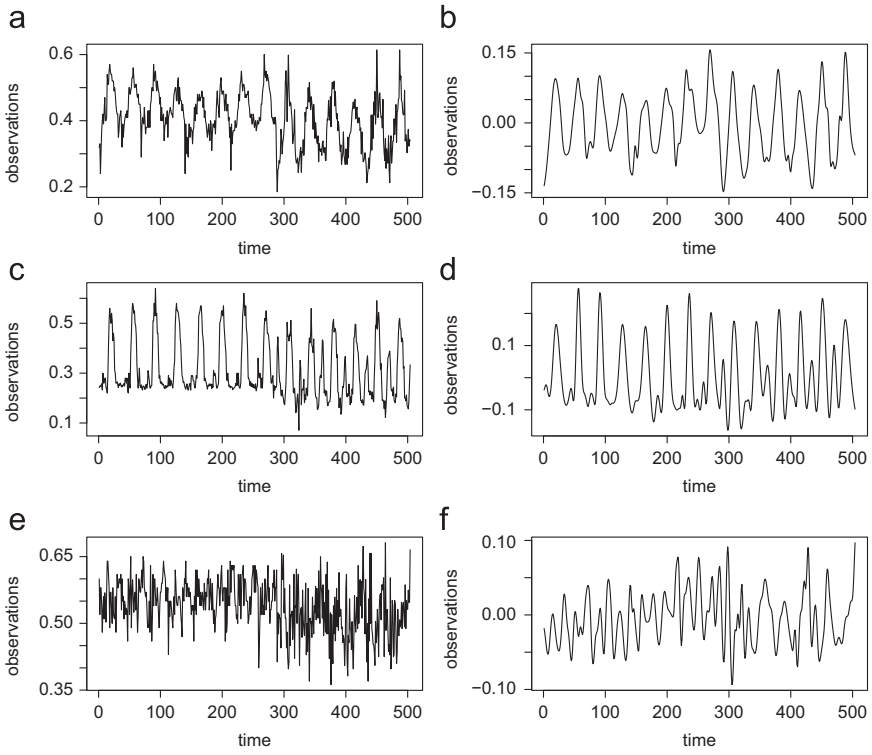

Fig. 19. Figures (a) and (b) illustrate signal *fapar-north* and its deterministic component. Figures (c) and (d) illustrate signal *fapar-asia* and its deterministic component. Finally, Figures (e) and (f) illustrate signal *fapar-south* and its deterministic component. All those deterministic components were obtained after considering the IMFs extracted using the EMD method.

Intrinsic Mode Functions. By combining IMFs according to their influences, an original signal can be separated into two components: one stochastic and another deterministic. Experiments were conducted on synthetic and real-world signals having an additive mixture of stochastic and deterministic influences. In all those scenarios, our approach was capable of separating both components.

In addition, one might wonder what happens if our approach is applied on a purely deterministic or on a purely stochastic signal. As a first step, signals will be decomposed in IMFs. When analyzing the first signal, i.e., the deterministic one, high-levels of mutual information will happen since the first IMF, what confirms all of them share relevant information. When analyzing the second, i.e., the stochastic one, no IMF will share relevant information.

As a consequence of this study, researchers can decompose signals into deterministic and stochastic components, improving their individual modeling and forecasting as confirmed in [5]. Furthermore, the study of the deterministic and stochastic components allows to understand patterns embedded in signal influences such as confirmed in [11].

Acknowledgments

This paper is based upon work supported by FAPESP (Sao Paulo Research Foundation), Brazil, under grant nos. 2014/13323-5 and 2009/18293-9.

References

[1] 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, R. Soc. Lond. Proc. Ser. A 454 (1998) 903–995.
[2] P. Flandrin, G. Rilling, P. Goncalves, Empirical mode decomposition as a filter bank, IEEE Signal Process. Lett. 11 (2) (2004) 112–114.
[3] G. Box, G.M. Jenkins, G. Reinsel, Time Series Analysis: Forecasting Control, 3rd Edition, Prentice Hall, 1994.
[4] K.T. Alligood, T.D. Sauer, J.A. Yorke, Chaos: An Introduction to Dynamical Systems, Springer, 1997.
[5] 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.
[6] F. Marvasti, Nonuniform sampling theorems for bandpass signals at or below the Nyquist density, IEEE Trans. Signal Process. 44 (3) (1996) 572–576.
[7] T. Moon, T. Weissman, Universal FIR MMSE filtering, IEEE Trans. Signal Process. 57 (3) (2009) 1068–1083.
[8] N. Tsakalozos, K. Drakakis, S. Rickard, A formal study of the nonlinearity and consistency of the empirical mode decomposition, Signal Process. 92 (2012) 1961–1969.
[9] E. Delechelle, J. Lemoine, O. Niang, Empirical mode decomposition: an analytical approach for sifting process, IEEE Signal Process. Lett. 12 (11) (2005) 764–767.
[10] O. Niang, A. Thioune, M.C.E. Gueirea, E. Delechelle, J. Lemoine, Partial differential equation-based approach for empirical mode decomposition: application on image analysis, IEEE Trans. Image Process. 21 (9) (2012) 3991–4001.
[11] R.A. Rios, L. Parrott, H. Lange, R.F. de Mello, Estimating determinism rates to detect patterns in geospatial datasets, Remote Sens. Environ. 156 (0) (2015) 11–20.
[12] C.E. Shannon, A mathematical theory of communication, Bell Syst. Tech. J. 27 (1948) 379–423, 623–656. ⟨http://cm.bell-labs.com/cm/ms/what/shannonday/shannon1948.pdf⟩.

[13] A. Kraskov, H. Stgbauer, P. Grassberger. Estimating mutual information, Phys. Rev. E: Stat. Nonlinear Soft Matter Phys. 69 (6 Pt 2).
[14] A. Papana, D. Kugiumtzis, Evaluation of mutual information estimators on nonlinear dynamic systems, Complex Phenom. Nonlinear Syst. 11 (2) (2008) 225–232.
[15] 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.
[16] G.A. Darbellay, P. Tichavský, Independent component analysis through direct estimation of the mutual information, in: Proceedings of 2nd International Workshop on ICA and Blind Source Separation, IEEE Signal Processing Society, Helsinki, Finland, 2000, pp. 69–74.
[17] A. Jerri, The Shannon sampling theorem – its various extensions and applications: a tutorial review, Proc. IEEE 65 (11) (1977) 1565–1596.
[18] J.G. Proakis, D.G. Manolakis, Digital signal processing, Principles, Algorithms, and Applications, 3rd Ed. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1996.
[19] R. Fensholt, I. Sandholt, M.S. Rasmussen, Evaluation of MODIS LAI, fAPAR and the relation between fAPAR and NDVI in a semi-arid environment using in situ measurements, Remote Sens. Environ. 91 (3–4) (2004) 490–507.
[20] N. Gobron, W. Knorr, A.S. Belward, B. Pinty, Fraction of absorbed photosynthetically active radiation (FAPAR), Bull. Am. Meteorol. Soc. 91 (7) (2010) S50–S51.
[21] M.D. Mahecha, L.M. Fürst, N. Gobron, H. Lange, Identifying multiple spatiotemporal patterns: a refined view on terrestrial photosynthetic activity, Pattern Recognit. Lett. 31 (14) (2010) 2309–2317.
[22] R.H. Shumway, D.S. Stoffer, Time Series Analysis and Its Applications: With R Examples, 2nd Edition, Springer, 2006.
[23] N.E. Huang, M.-L.C. Wu, S.R. Long, S.S. Shen, W. Qu, P. Gloersen, K.L. Fan, A confidence limit for the empirical mode decomposition and Hilbert spectral analysis, Proc. R. Soc. Lond. A: Math. Phys. Eng. Sci. 459 (2037) (2003) 2317–2345.