Advances in Adaptive Data Analysis
Vol. 3, No. 3 (2011) 363–383

DOI: 10.1142/S1793536911000696

TREND EXTRACTION FOR SEASONAL TIME SERIES USING ENSEMBLE EMPIRICAL MODE DECOMPOSITION

FAROUK MHAMDI*,‡, JEAN-MICHEL POGGI†,§ and MÉRIEM JAÏDANE*,¶

*Unité Signaux et Systèmes, Ecole Nationale d'Ingénieurs de Tunis,
BP 37, Le Belvédère 1002, Tunis, Tunisia
†Université d'Orsay, Lab. de Mathématiques,
bât. 425, 91405 Orsay, France
‡Farouk.Mhamdi@enit.rnu.tn
§Jean-Michel.Poggi@math.u-psud.fr
¶meriem.jaidane@planet.tn

In this paper, we investigate eligibility of trend extraction through the empirical mode decomposition (EMD) and performance improvement of applying the ensemble EMD (EEMD) instead of the EMD for trend extraction from seasonal time series. The proposed method is an approach that can be applied on any time series with any time scales fluctuations. In order to evaluate our algorithm, experimental comparisons with three other trend extraction methods: EMD-energy-ratio approach, EEMD-energy-ratio approach, and the Hodrick–Prescott filter are conducted.

Keywords: Trend extraction of time series; Ensemble empirical mode decomposition.

1. Introduction

Time series often contain many components such as seasonal and cyclical components, trends, and irregularities. Even if we assume an additive decomposition model, trend extraction and seasonal adjustments are difficult tasks of time series analysis, due to the extreme variety of time series with their own time scales. Thereby, the trend has fuzzy general definition, despite its great practical importance. Nevertheless, the trend is considered as a "smooth additive component that contains information about global change," see [Alexandrov *et al.* (2009)]. For long- or medium-term load forecasting, for example, it is of great interest to extract smoothed trend components from time series. However, this general definition makes trend extraction an ambiguous task since we can found several candidates that match this definition.

Several techniques have been traditionally used for time series components extraction and adjustment. As mentioned above, we will focus on trend extraction. The most frequently used approaches for trend extraction are local or global regressions, moving average filtering, Tramo-Seats, X-11, X-12, and the Hodrick–Prescott filter (see [Alexandrov *et al.* (2009)] for a recent review). Note that those methods need model specification or parameter adequacy related to time scale characteristics of the time series components. Therefore, each method has it applying assumption that is make comparison difficult.

Huang *et al.* [1998] have presented an attractive signal analysis method called the empirical mode decomposition (EMD). This method is particularly useful to deal with possibly nonstationarity and nonlinearity that often characterize time series. It considers the signal as a superposition of oscillatory components, extracted from upper and lower envelopes, so-called intrinsic mode functions (Imf). The Imfs are fully data-driven. This method is easy to implement and does not use any predetermined transform depending on the choice of a particular basis. The EMD is an adaptive method that is entirely empirical and captures the time series characteristics in separate Imfs, explaining why it has been successfully applied in many engineering fields, see e.g. [Zhou *et al.* (2008); Flandrin *et al.* (2004)]. All those attractive characteristics have motivated researchers to develop EMD-based trend extraction methods, since it allows to identify various trends at different time scales [Suling *et al.* (2009); Flandrin *et al.* (2004); Mhamdi *et al.* (2010); Moghtaderi *et al.* (2010); Wu *et al.* (2007)].

Flandrin *et al.* [2004] have investigated the potentialities and limitations of EMD-based methods in detrending and relating the trend with the statistical properties of the Imfs. The trend is then defined as the sum of the Imfs having nonzero means. Application to heart-rate data illustrates its potential detrending usefulness. Another definition is given by Suling *et al.* [2009], relating trend to time scales. In [Mhamdi *et al.* (2010)] we have investigated the eligibility of considering the residue of the EMD as the trend that can be extracted from time series with classical trends (linear or exponential). Indeed, at the end of the Imf extraction algorithm, the number of extremes in the residue does not exceed two these results make EMD algorithm very suitable to extract monotonic trends. Preliminary results of our simulations [Mhamdi *et al.* (2010)] indicate that the EMD approach gives similar results as the Hodrick Prescott filter (see e.g.[Pollock (2003)]).

Moghtaderi *et al.* [2010] have presented another attractive EMD based method with automatic identification of the Imfs that compose the trend extracted: the trend is obtained by aggregating some low-frequency Imfs, automatically chosen. This method [Moghtaderi *et al.* (2010)], as others based on EMD, still suffer from modes mixing and a lack of robustness in the criteria that automatically identify the Imfs that will compose the trend in the case of noisy signals. This affects trend extraction accuracy of Mogthaderi's method in the case of seasonal time series.

In this paper, we will first show some accuracy improvement in Moghtaderi's algorithm (EMD-energy-ratio approach) by using the ensemble EMD (EEMD) introduced by [Zhaohua and Huang (2004)] instead of the EMD is then introduced to deal with noisy data. This EMD variant is more stable and minimizes modes mixing [Niazy *et al.* (2009)]. Our proposed variant of Moghtaderi's trend extraction algorithm will be noted EEMD-energy-ratio approach.

In a second part of this article, we extend the EEMD-based algorithm dealing to seasonal time series. Trend extraction through this algorithm is based on Imfs seasonality checking (ISC) using a new criterion based on pattern seasonality checking. Our method dealing with seasonal time series will be noted EEMD-ISC. We investigate its performance improvements comparing three methods: EMD-energy-ratio approach, EEMD-energy-ratio approach, and the Hodrick–Prescott filter. These comparisons will be based on simulated and real seasonal time series with different seasonal patterns and various trend curves. For classical simple trends (exponential and linear), we refer to our studies presented in [Mhamdi *et al.* (2010)].

The outline of the paper is as follows. Section 2 recalls some facts about EMD and sketches how it is a good candidate for sensible components extraction. Section 3 presents the EMD-energy-ratio based trend extraction procedure and its limits. Section 4 investigates the performance improvement of applying the ensemble EMD instead of the EMD in Moghtaderi's algorithm for trend extraction. Section 5 presents our algorithm for trend extraction from noisy seasonal time series. In Section 6, experimental results on simulated seasonal time series are analyzed, and compared to the other three methods: EMD-energy-ratio approach, EEMD-energy-ratio approach, and the Hodrick–Prescott filter.

2. EMD for Time Series Analysis

EMD has been introduced by Huang *et al.* [1998], as an alternative approach to traditional methods for analyzing time series such as wavelets (see [Misiti *et al.* (2007)]) or Fourier methods. The key idea of EMD is to locally decompose data $y(t)$ into oscillatory components so-called Imfs. This method is particularly useful to deal with possibly nonstationary and nonlinear components that often characterize time series. It considers the signal as a superposition of oscillatory components, which are extracted from upper and lower envelopes, so-called Imfs. The Imfs are fully data-driven. This method is easy to implement and does not use any analyzed predetermined transform.

2.1. *EMD algorithm*

The algorithm for the extraction of Imfs from a given time series $y(t)$ data is called sifting and it consists of the following steps: [Huang *et al.* (1998)]

(i) Initialize the residue $r_0(t) = y(t)$, set $g_0(t) = r_{k-1}(t)$, and $i = 1$; the index of Imf $k = 1$.

(ii) Construct the lower minima $I\min_{i-1}$ and the upper maxima $I\max_{i-1}$ envelopes of the signal by the cubic spline method.
(iii) Calculate the mean values by averaging the upper and lower envelopes. Set $m_{i-1} = [I\max_{i-1} + I\min_{i-1}]/2$.
(iv) Subtract the mean from the original signal: $g_i = g_{i-1} - m_{i-1}$ and $i = i + 1$, and repeat steps (ii)–(iv) until g_i being an Imf (see below for the definition). If so, the kth Imf is given by $IMF_k = g_i$.
(v) Update residue $r_k(t) = r_{k-1}(n) - IMF_k(t)$. This residual component is treated as a new data and subjected to the process described above to calculate the next IMF_{k+1}.
(vi) Repeat the steps above until the final residual component $r(t)$ becomes monotone.

It turns out that an Imf satisfies the two following properties. First, the upper and lower envelopes are symmetric and second, the number of zero-crossings and the number of extremum are equal or differ at most by one.

The advantage of this method is that the oscillatory modes, which are generated, are derived directly from the data, without any reference to a predetermined dictionary of functions.

At the end of this process, the initial time series is decomposed into K Imf components and r is the final residue:

$$y(t) = \sum_{k=1}^{K} IMF_k(t) + r(t). \tag{1}$$

2.2. *EMD and time scales of a time series*

The EMD is an adaptive method that is entirely empirical and captures the characteristics in separate Imfs. Let us consider observed additive time series $y = (y(1), y(2), \ldots, y(N))$ supposed to be of the form:

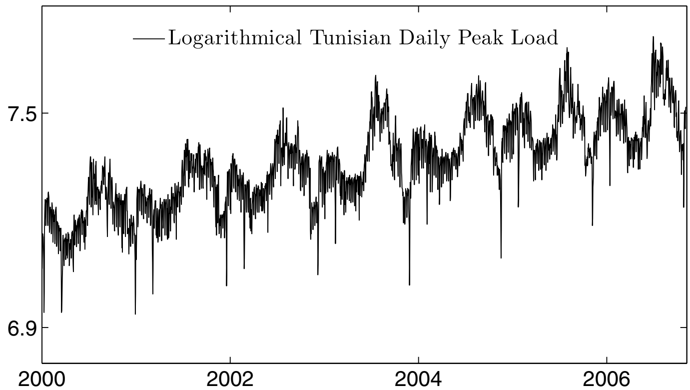
$$y(t) = T(t) + S(t) + C(t) + I(t), \tag{2}$$

where the different components are trend (T), seasonal components (S), cycles (C), and irregular term (I) for error modeling.

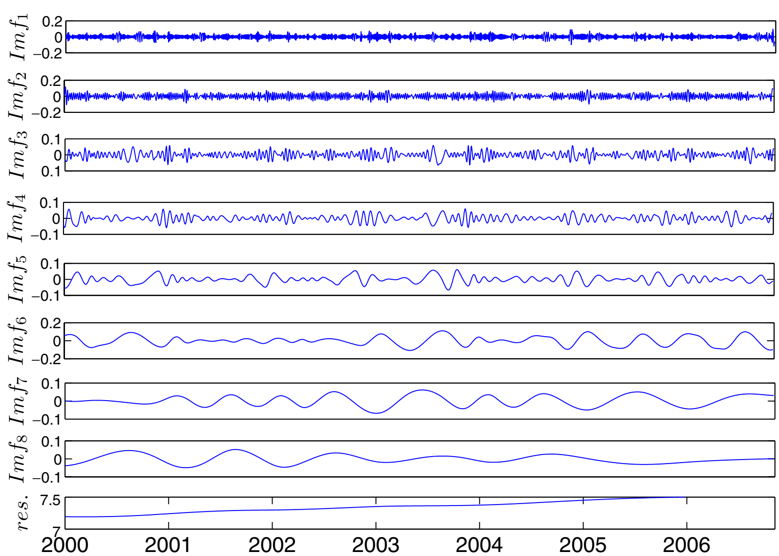
With respect to this usual form, we will not make distinction between trend and cyclical components that are a long-term nonstationary components, in order to make these components identifiable. Therefore, the signal decomposition reduces to:

$$y(t) = T(t) + S(t) + I(t). \tag{3}$$

According to this time series modeling (Eq. (3)), we can expect that the EMD can be considered as an attractive method for time series analysis and especially for time series trend extraction.



(a)



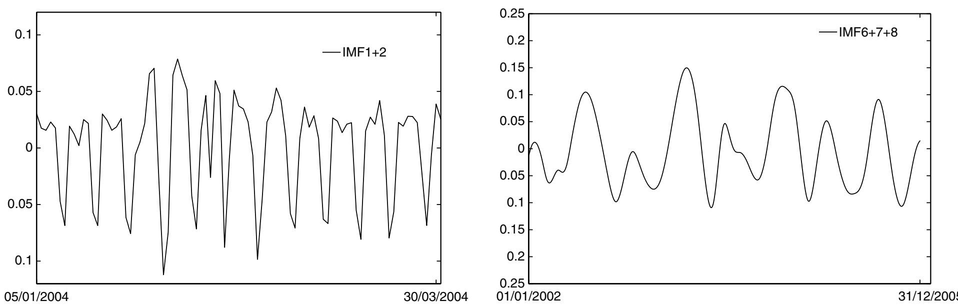
(b)

Fig. 1. EMD decomposition of the logarithmical daily peak load 2000–2006 from STEG utility: (a) The logarithmical daily peak load 2000–2006 and (b) Imf components and the final residue or preliminary identified trend.

To illustrate the ability of EMD to recover sensible time series components, we apply the EMD method to logarithmical Tunisian daily peak loads from 2000 to 2006 from STEG utility[a] (Fig. 1).

As noticed by Ould Mohamed *et al.* [2009], we note that Imfs 1 to 2 exhibit high frequency and can represent very short-term (weekly) fluctuations (Fig. 2(a)), Imfs 3 to 5 capture small percentage of variance (0.13%), indicating that such Imfs are

[a]Tunisian company of electricity and gas.



(a) (b)

Fig. 2. Sensible Imfs aggregation according to interpretable seasonal load components: (a) Sum of Imfs 1–2: weekly component and (b) Sum of Imfs 6–7–8: annual component.

not significant. Imfs 6 to 8 capture mid-term effects described by seasonal annual variations (Fig. 2(b)). Finally, the residue component in EMD could represent the major trend of load demand in the long term that may be related to economic growth in Tunisia.

This attractive property motivate researchers to develop trend extraction EMD-based methods, since it allows identifying various trends at different time scales [Suling *et al.* (2009); Flandrin *et al.* (2004); Mhamdi *et al.* (2010); Moghtaderi *et al.* (2010); Wu *et al.* (2007)]. A powerful method presented in [Moghtaderi *et al.* (2010)] is detailed below.

3. EMD-Energy-Ratio Approach for Trend Extraction

In [Moghtaderi *et al.* (2010)], the discrete time series $(y(t))$ is assumed to be of the form:

$$y(t) = T(t) + X(t), \tag{4}$$

where $X(t)$ has nonzero and continuous power spectrum and the most of his spectrum is concentrated on some interval $[f_{\min}, f_{\max}]$, $0 \leq f_{\min} \ll f_{\max} < 1/2$ and $T(t)$ is supposed slowly varying sequence with respect to X. This means that the discrete-time Fourier of X is nonzero and continuous in $[0, f_T]$, where $f_T < f_{\min}$.

The trend $T(t)$ is obtained by aggregating the Imfs of the lowest frequency: $\widehat{T}(t) = \sum_{k=k^\star}^{K} IMF_k(t) + r(t)$, where $k^\star$ is automatically chosen, based on Imfs energy-ratio approach analysis as described below.

3.1. *EMD-energy-ratio approach*

Let $y(t) = \sum_{k=1}^{K} IMF_k(t) + r(t)$ be the EMD decomposition of y. Let $G^k = \sum_{t=0}^{N-1} |Imf_k(t)|^2$ denotes the energy of the kth Imf, $k = 1, \ldots, K$ and Z^k denotes the number of zero-crossings in the kth Imf $(k = 1, \ldots, K)$. For each EMD decomposition, $R^k = Z^{k-1}, Z^k$ denotes the kth ratio of zero-crossing numbers, $k = 2, \ldots, K$.

Based on various simulations, [Moghtaderi *et al.* (2010)] concluded that in the absence of a trend (T):

- $G^k < G^{k-1}$ for all $k = 1, \ldots, K$.
- $R^k \approx 2$ for all $k = 2, \ldots, K$.

Based on these results, the estimated trend is the result of IMFs and residue aggregation as follow:

$$\widehat{T}(t) = \sum_{k=k^\star}^{K} IMF_k(t) + r(t), \tag{5}$$

where $k^\star$ is the smallest k verifying:

$$\begin{cases} G^{k^\star} > G^{k^\star - 1} \\ R^{k^\star} \text{ is significantly different from 2.} \end{cases} \tag{6}$$

Note that confidence intervals CI_{1-p} with different significance levels p ($0 \leq p \leq 100$) were performed.[b]

3.2. *Limitations of the EMD-energy-ratio trend extraction approach*

The EMD-energy-ratio approach, as described above, shows two limitations: sensitivity to the noise contained in the data and also the IMFs aggregation criteria is not optimized to trend extraction from seasonal time series. To circumvent these drawbacks, we will investigate its performance through experimental studies. We first consider two simulated models for monthly electrical data containing classical trends (linear and exponential), even if they are unrealistic.

- Nonseasonal time series y^1:

$$\begin{cases} y^1(t) = T(t) + \epsilon(t). \\ \epsilon(t) = \nu(t) + \theta\nu(t-1) \quad \nu(t) \sim \aleph(0, \sigma^2) iid \\ T(t) = \beta_0 + e^{\alpha t} \end{cases} \tag{7}$$

- Seasonal time series y^2:

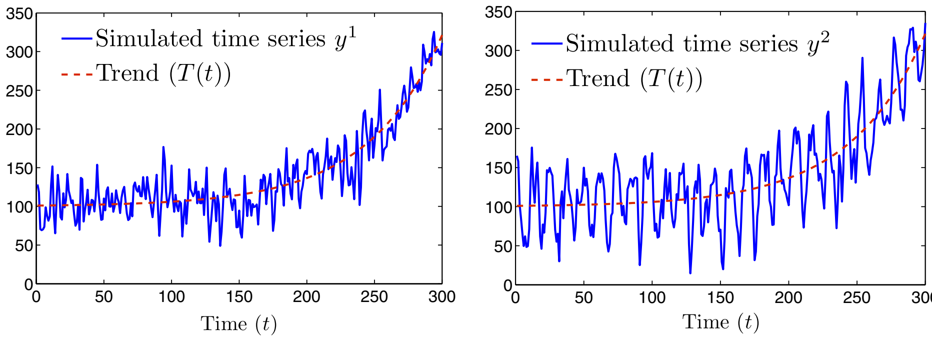
$$\begin{cases} y^2(t) = y^1(t) + S(t). \\ S(t) = \beta_1 \cos(\frac{2\pi t}{12}) + \beta_2 \sin(\frac{2\pi t}{12}) \end{cases} \tag{8}$$

where $t = (1, 2, \ldots N)$, $N = 300$, $\beta_0 = 100$, $\beta_1 = 24$, $\beta_2 = 32$, $\theta = 0.8$, $\sigma^2 = 50$, and $\alpha \in \{0.015, 0.0151, \ldots, 0.025\}$.

In Fig. 3, we present the two monthly load (y^1 and y^2) for $\alpha = 0.018$.

Let SNR denotes the signal to noise ratio which is defined as the power ratio between a signal (y^1 or y^2) and the noise ϵ. For each value of SNR $\in \{100, 33.33, 20, 12.5, 10\}$, we create 100 realizations of each time series models for

[b] In our study, we choose p equal to 0.09. In this case, $CI_{0.91} = [1.81, 2.73]$.



(a) (b)

Fig. 3. Simulated monthly power patterns with exponential trend ($\alpha = 0.18$): (a) 25 years simulated non seasonal time series y^1 and (b) 25 years simulated seasonal time series y^2.

y^1 and y^2 ($\alpha = 0.018$) of length $T = 300$. For each realization, we apply the EMD to extract Imfs and we apply the energy-ratio approach (with $p. = .09$) to each EMD decomposition in order to evaluate $k^\star$.

Similarly, we estimate the optimal value k^{opt} of $k \in \{1, 2, \ldots, K\}$ by minimizing the Euclidean distance ($d = \frac{\|T - \widehat{T}\|_2}{\|T\|_2}$) between the estimated trend $\widehat{T}(t) = \sum_{i=k}^{K} IMF_i(t) + r(t)$ and the simulated T. Let $\overline{d}$ denotes the mean error of 100 realizations for each value of SNR and θ denotes the number of good detections ($k^\star = k^{\text{opt}}$) obtained with the EMD-energy-ratio approach.

We observe, in Table 1, that the mean error increases with the SNR. This is due to the EMD algorithm that suffers from serious problems such as instability on the number of components (Imfs) obtained, modes mixing, and breaks on the Imfs curves since it is highly sensitive to the noise contained in the data. In addition, Table 2 shows that the existence of seasonal components deteriorates EMD-energy-ratio accuracy, since the number of false detections is more important in the case of a seasonal time series. Therefore, it is important to propose another trend extraction criteria for seasonal time series.

In the next section, we show the ability of improving EMD-energy-ratio approach by using a modified EMD variant called the EEMD [Zhaohua and Huang 2004] instead of EMD, leading to robust techniques for trend extraction for noisy seasonal time series.

Table 1. Comparison of mean error ($\overline{d}$) of the trend extraction through the EMD-energy-ratio approach.

SNR	Nonseasonal time series (y^1)	Seasonal time series (y^2)
100	0.038	0.112
33.33	0.055	0.096
20	0.066	0.090
12.5	0.074	0.093
10	0.084	0.093

Table 2. Comparison of good detections indicator (θ) (in %) obtained with the EMD-energy-ratio approach.

SNR	Nonseasonal time series (y^1)	Seasonal time series (y^2)
100	49	29
33.33	42	31
20	38	30
12.5	37	33
10	52	44

4. EEMD Based Energy-Ratio Approach

4.1. *EEMD*

The EEMD is a variant of the EMD algorithm introduced by [Zhaohua and Huang (2004)]. This EMD variant is more robust than the EMD [Niazy *et al.* 2009]. The principle is to add N realizations of Gaussian white noise $b = (b^{(1)}(t), b^{(2)}(t), \ldots, b^{(N)}(t))$ to the signal $y(t)$ in order to obtain N noisy pseudo signals $(y^{(1)}(t), y^{(2)}(t), \ldots, y^{(N)}(t))$. Then, we apply EMD algorithm to each signal $y^{(i)}, i = 1, \ldots, N$. We note $(\mathrm{IMF}_1^{(i)}, \mathrm{IMF}_2^{(i)}, \ldots, \mathrm{IMF}_{K^{(i)}}^{(i)})$ the Imfs vector obtained through each EMD decomposition of $y^{(i)}(t)$, where $K^{(i)}, i = 1, \ldots, N$ is the number of Imfs. Finally, the EEMD Imfs $(\overline{\mathrm{IMF}_k})$ decompositions is obtained by averaging EMD Imfs $(\mathrm{IMF}_k^{(i)})$ derived from each signal $y^{(i)}(t)$. Let us recall that in [Zhaohua and Huang (2004)], we reject all EMD decompositions whose number of Imfs obtained $(K^{(i)})$ is different from K that is previously fixed by user.

In addition, in our case, the EEMD algorithm is slightly modified to achieve better stability robustness. Two parameters of the EEMD algorithm are changed. First, the noise added to the signal reflects the physical measurement noise of the considered time series. Second, the number of Imfs (K_0) is set to the more recurrent number of Imfs $K_0 = \mathrm{mode}\{K^{(1)}, K^{(2)}, \ldots, K^{(N)}\}$ and all EMD decompositions whose number of Imfs obtained $K^{(i)}$ is different from K_0 are rejected. This variant is described in Fig. 4.

Since $y(t) \neq \sum_{k=1}^{K_0} \overline{\mathrm{IMF}_k}(t)$, due to the noise added to the original signal, in order to reconstruct the original considered signal $y(t)$, the first $\overline{\mathrm{IMF}}_1$ is set to $\overline{\mathrm{IMF}}_1^*(t) = y(t) - \sum_{k=1}^{K_0} \overline{\mathrm{IMF}_k}(t)$.

As illustration, we have applied the EEMD algorithm to the logarithmical daily peak load 2000–2006 from STEG utility (Fig. 5).

In the sequel, we show the performance improvement achieved using this variant of EEMD instead of EMD through the EMD-energy-ratio algorithm.

4.2. *EEMD-energy-ratio performance*

We investigate the performance improvement achieved using the EEMD instead of EMD through the EMD-energy-ratio algorithm. We consider two cases: without seasonal component y^1 and with seasonal component y^2 (Eq. (8)). The

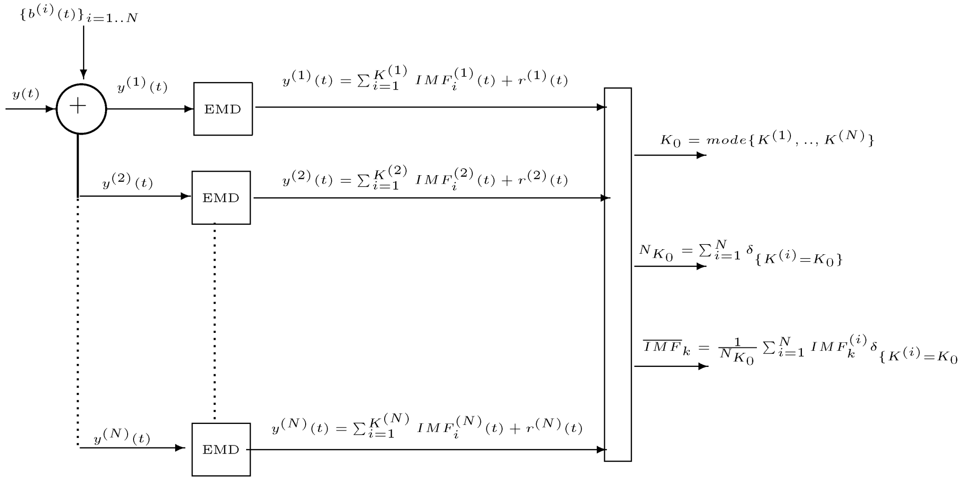


Fig. 4. Modified EEMD.

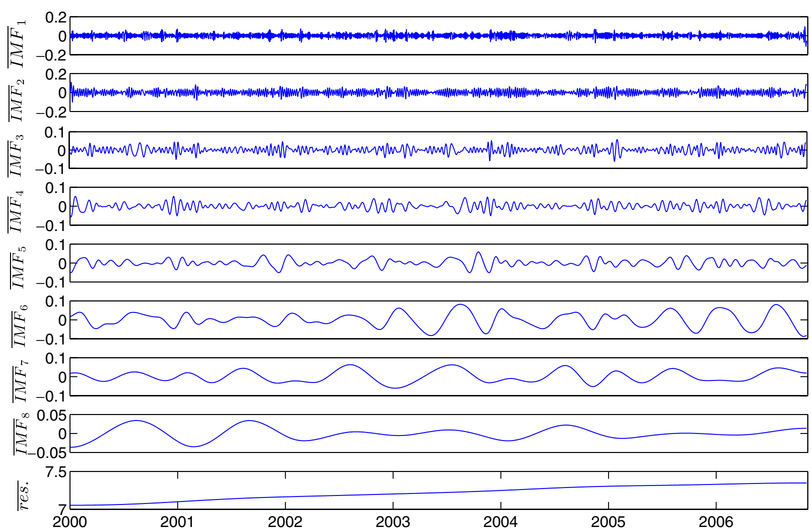


Fig. 5. EEMD of the logarithmical daily peak load 2000–2006 from STEG utility: $\overline{\mathrm{IMF}}_i$ components and the final residue or monotonic identified trend.

comparison will be based on the estimated Euclidean distance ($d = \frac{\|T-\widehat{T}\|_2}{\|T\|_2}$) between simulated trend T and the trend $\widehat{T}$ extracted through the two approaches: EMD-energy-ratio and EEMD-energy-ratio. Let min, mean, and max denote, respectively, the minimum, mean, and maximum Euclidean distances estimated from 100 realizations through the EMD-energy-ratio and EEMD-energy-ratio approaches.

Table 3. Statistics of estimated errors obtained through EMD-energy-ratio and EEMD-energy-ratio approaches in the case of non-seasonal simulated time series (y^1 with $\alpha = 0.018$).

	EMD-energy-ratio			EEMD-energy-ratio		
SNR	Min	mean	Max	Min	Mean	Max
100.00	0.018	0.038	0.112	0.017	0.035	0.078
33.33	0.023	0.055	0.126	0.026	0.050	0.112
20.00	0.033	0.066	0.140	0.036	0.056	0.130
12.50	0.029	0.074	0.179	0.035	0.068	0.153
10.00	0.046	0.084	0.214	0.040	0.076	0.183

Table 4. Statistics of estimated errors obtained through EMD-energy-ratio and EEMD-energy-ratio approaches in the case of noisy seasonal simulated time series (y^2 with $\alpha = 0.018$).

	EMD-energy-ratio			EEMD-energy-ratio		
SNR	Min	Mean	Max	Min	Mean	Max
100.00	0.024	0.112	0.227	0.019	0.083	0.215
33.33	0.031	0.086	0.272	0.031	0.084	0.227
20.00	0.035	0.090	0.320	0.032	0.090	0.316
12.50	0.042	0.093	0.359	0.039	0.087	0.278
10.00	0.037	0.093	0.372	0.038	0.087	0.284

We can conclude that applying the EEMD variant instead of the EMD increases the accuracy of Moghtaderi's techniques in the two cases: seasonal and nonseasonal time series (Tables 3 and 4). However, seasonal components deteriorate energy-ratio accuracy (see Tables 4). Therefore, it is important to define another trend extraction criterion for seasonal time series.

In the next section, we present our proposed trend extraction method for seasonal time series.

5. Imfs Seasonality Checking Algorithm (EEMD-ISC)

Let us consider a time series coming from the full considered given by model Eq. (2). In time series analysis, seasonal components are usually identified as components with constant or slightly varying period. Based on this idea, we will develop an "Imfs seasonality checking criteria" allowing the EEMD based trend extraction approach to deal with seasonal time series. In the following, we describe the three steps of this technique:

- Step 1: Ensemble empirical mode decomposition.
- Step 2: Imfs seasonality checking.
- Step 3: Trend extraction.

Step 1: In our approach, we first apply the modified EEMD as described in Sec. 4.1 to the considered time series. Let $(\overline{\mathrm{IMF}_1^*}, \overline{\mathrm{IMF}_2}, \ldots, \overline{\mathrm{IMF}_{K_0}})$ denote the EEMD Imfs obtained from Step 1. We note l_k and p_k the numbers of maxima and minima for each EEMD Imf ($\overline{\mathrm{IMF}_k}$), $1 \leq k \leq K_0$). Let, also, d^M and d^m denote the distance vectors calculated between indices t_i^M, $1 \leq i \leq l_k$, of the maxima and between indices t_j^m, $1 \leq j \leq p_k$, of the p_k minima ($d_j^M = t_j^M - t_{j-1}^M$ and $d_j^m = t_j^m - t_{j-1}^m$) finding on each EEMD Imfs ($\overline{\mathrm{IMF}_k}$).

Let us remark that, if an $\overline{\mathrm{IMF}_k}$ is (or contains) a periodic pattern with period equal to p, the sum of the distances $D^M = \sum_{i=1}^{l_k} d_i^M$ (respectively, $D^m = \sum_{i=1}^{l_k} d_i^m$) will be "proportional" with p, then seasonality checking test is the following:

Step 2: The kth EEMD Imf ($\overline{\mathrm{IMF}_k}$) will belong to the estimated seasonal component $\widehat{S}$ if:

$$1 - \beta \leq \frac{\frac{D^M}{\widetilde{d}^M}}{l_k - 1} \leq 1 + \beta \quad or \quad 1 - \beta \leq \frac{\frac{D^m}{\widetilde{d}^m}}{p_k - 1} \leq 1 + \beta, \tag{9}$$

where $\widetilde{d}^M$ (respectively, $\widetilde{d}^m$) denotes the mode distance observed on d^M (respectively, d^m) and $\beta \geq 0$. This algorithm parameter is typically small with respect to periodic component properties.

Step 3: Let $E^s = \{k_1, k_2, \ldots, k_m\}$, $1 < k_1 < k_2 < \cdots < k_m < K_0$ denotes the index set of the m seasonal EEMD Imfs. The EEMD-ISC extracted trend $\widehat{T}$ is then defined as the sum (from $k_m + 1$ to K_0) of EEMD Imfs: $\widehat{T}(t) = \sum_{k=km+1}^{K_0} \overline{\mathrm{IMF}_k}(t) + r(t)$.

Note that it is important to make correction to the estimated trend $T(t)$ at the end of the algorithm since the first EEMD Imf $\overline{\mathrm{IMF}_1^*}$ (see Sec. 4.1) does not has necessary zero mean. Therefore, the final trend component will be reset to $T(t) = T(t) + \frac{1}{N}\sum_{i=1}^{N} \overline{\mathrm{IMF}_1^*}(t_i)$.

As we can remark the performance of our method depend on the optimal choice of β. The results of various simulations based on different trend and seasonal component combinations (y^1, y^2, linear trend with sinusoidal component) indicate that β can be set to a default value equal to 0.3.

As an illustration, we present, in Table 5, the number of false detections estimated through the EEMD-energy-ratio and EEMD-ISC approaches in the cases of y^1 and y^2.

By comparing results recorded in Tables 3, 4, and 6, we conclude that the two algorithms perform similarly well for nonseasonal time series with performance superiority of the EEMD-energy-ratio when increasing the noise power in the signal.

Table 5. Comparison of good detections indicator (θ) (in %).

	Nonseasonal time series (y^1)		Seasonal time series (y^2)	
SNR	EEMD-energy-ratio	EEMD-ISC	EEMD-energy-ratio	EEMD-ISC
100.00	45	46	34	68
33.33	50	49	36	58
20.00	49	45	39	55
12.50	45	38	40	48
10.00	49	36	47	40

Table 6. Statistics of estimated errors obtained through EEMD-ISC approach in the case of y^1 and y^2.

	Nonseasonal time series (y^1)			Seasonal time series (y^2)		
SNR	Min	Mean	Max	Min	Mean	Max
100.00	0.018	0.031	0.051	0.017	0.043	0.093
33.33	0.027	0.047	0.102	0.030	0.059	0.098
20.00	0.036	0.061	0.128	0.033	0.062	0.107
12.50	0.034	0.069	0.131	0.030	0.076	0.107
10.00	0.039	0.079	0.149	0.037	0.080	0.124

In the case of seasonal time series, the results show that our method, based on EEMD-ISC, is more efficient than the two others methods (EMD-energy-ratio and EEMD-energy-ratio). This is clearer in the case of low noise power compared to the signal.

6. Performance Evaluation

In order to evaluate our algorithm, we compare EEMD-ISC and three other trend extractors: EEMD-energy-ratio approach, EEMD-energy-ratio approaches examined first, and the Hodrick–Prescott filter examined finally.

6.1. *Simulated seasonal time series*

In this section, we investigate EEMD-based algorithm performance for trend extraction from seasonal time series. The analyzed signal is modeled by the following equation:

$$y^{(i)}(t) = T^{(j)}(t) + \exp(\alpha t)S^{(k)}(t) + I^{(i)}(t), \tag{10}$$

where $i = 3, \ldots, 6$, $j = 1, 2$, and $k = 1, 2$. The exponential term ($\exp(\alpha t)$) is used in order to simulate heteroscedastic time series [Alexandrov *et al.* (2009)]. I denotes irregularities.

The elementary simulated components ($S(t)$ and $T(t)$) of the time series are as follows:

- Simulated seasonal components:

We consider two simulated monthly and quarterly periodic seasonal components (S^k, $k = 1, 2$) in order to make comparison with the Hodrick–Prescott filter with optimal values of his parameter λ (14,400 and 1,600). The two seasonal patterns are presented in Fig. 6.

- Simulated trends:

As a complement of our work, [Mhamdi *et al.* (2010)] in which we analyzed basic linear and exponential trends, we consider more complex trends $T^{(1)}(t)$ also considered in [Alexandrov *et al.* (2009)] and $T^{(2)}(t)$ shown in Fig. 7.

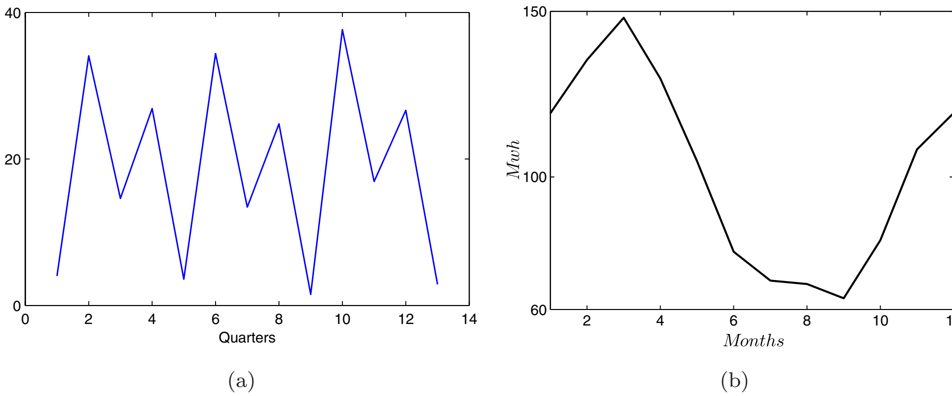


Fig. 6. Simulated quarterly and monthly seasonal components: (a) Three years simulated quarterly seasonal component $S^{(1)}$ and (b) One year simulated monthly seasonal component $S^{(2)}$.

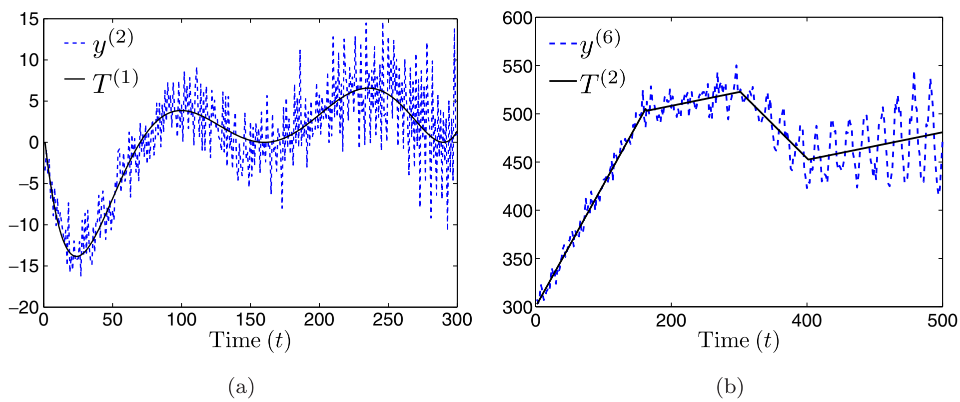


Fig. 7. Examples of simulated seasonal time series: (a) Simulated time series ($y^{(3)}$) and (b) Simulated time series ($y^{(6)}$).

We present, in Fig. 7, two examples of simulated time series $y^{(3)}$ obtained for $j = k = 1$ and $y^{(6)}$ obtained for $j = k = 2$.

6.2. *EEMD-ISC, EMD-energy-ratio, and EEMD-energy-ratio accuracy comparison*

To investigate the EEMD-ISC trend extraction performance, a comparison with trends extracted by EEMD-energy-ratio approach and EEMD-energy-ratio approach is performed.

In order to simulate time series with various seasonality degrees, we make vary the noise variance. Therefore, let $\delta = \sum_{t=1}^{T}[I(t)]^2/\sum_{t=1}^{T}[CS(t)]^2$ denotes the ratio of noise (I) power corrupting the signal to the seasonal component (S) power. For each time series, we start from a time series with very prominent seasonality component ($0 \leq \delta \leq 5$) to a time series with very mild seasonality embedded with much irregularity ($\delta \geq 20$).

In Figs. 8–10 and Tables 7–9, we report, for three methods (EMD-energy-ratio, EEMD-energy-ratio, and EEMD-ISC), the average of Euclidean norm and the percentage of good detections over 100 runs of $y^{(3)}$ (Fig. 8 and Table 7), $y^{(4)}$ (Fig. 9 and Table 8) and $y^{(6)}$ (Fig. 10 and Table 11).

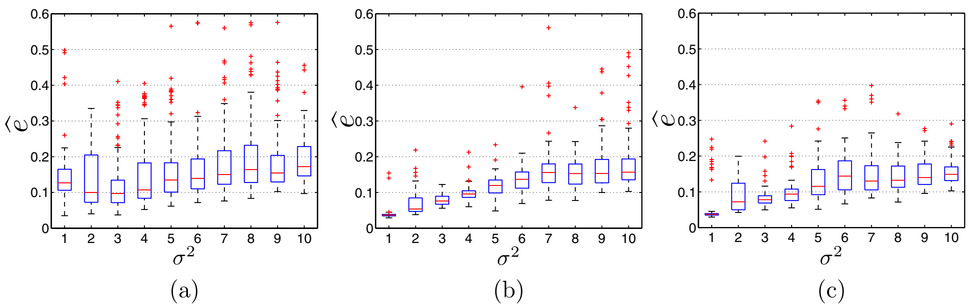


Fig. 8. (a) EMD-energy-ratio, (b) EEMD-energy-ratio, and (c) EEMD-ISC accuracy comparison in the case of $y^{(3)}$.

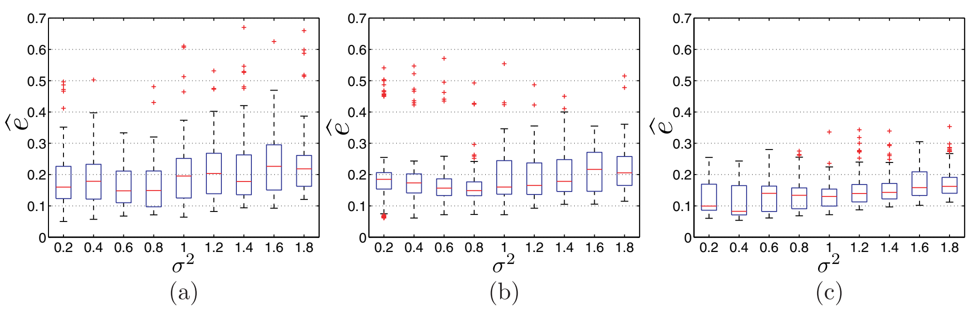


Fig. 9. (a) EMD-energy-ratio, (b) EEMD-energy-ratio, and (c) EEMD-ISC accuracy comparison in the case of $y^{(4)}$.

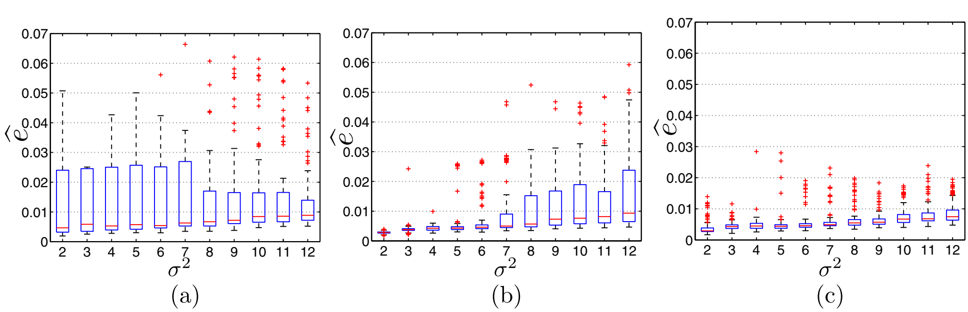


Fig. 10. (a) EMD-energy-ratio, (b) EEMD-energy-ratio, and (c) EEMD-ISC accuracy comparison in the case of $y^{(6)}$.

Table 7. Comparison of good detections (θ) (in %) $y^{(3)}$.

σ^2	δ(%)	EEMD-energy-ratio	EEMD-ISC
0.2	1,2	96	92
0.4	5,3	60	50
0.6	11	44	42
0.8	17,5	32	40
1.0	33	25	40
1.2	45,5	30	42
1.4	62	18	50
1.6	72	25	40
1.8	98,2	33	42

Table 8. Comparison of good detections (θ) (in %) $y^{(4)}$.

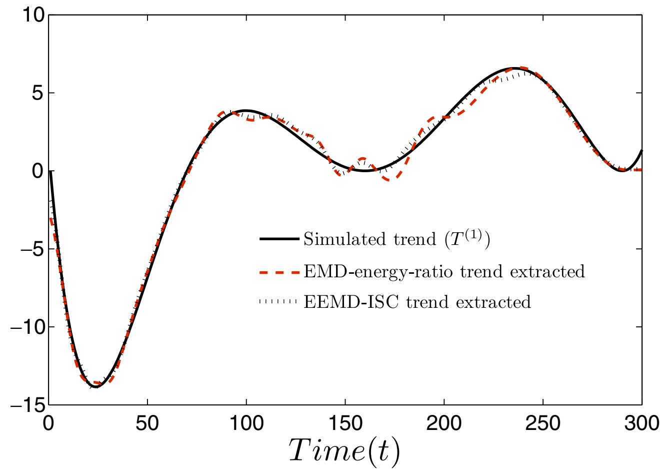
σ^2	δ (%)	EEMD-energy-ratio	EEMD-ISC
0.2	0,5	11	25
0.4	1,5	11	32
0.6	3,3	7	38
0.8	7,5	8	33
1.0	10,5	4	33
1.2	13	6	25
1.4	17,4	7	25
1.6	24	16	24
1.8	29	9	40

Table 9. Comparison of good detections (θ) (in %) $y^{(6)}$.

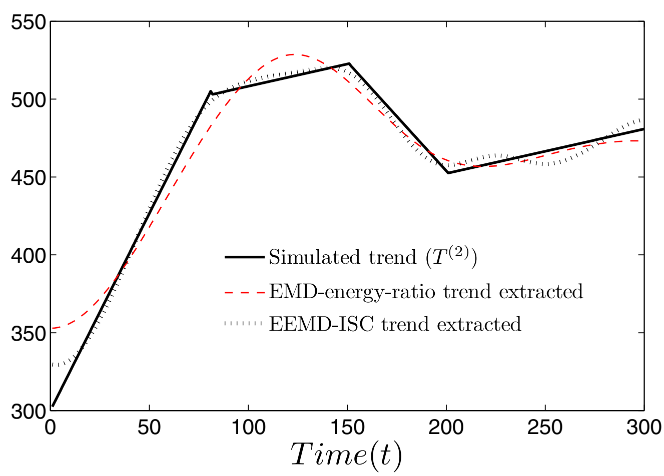
σ^2	δ(%)	EEMD-energy-ratio	EEMD-ISC
2	1,5	20	10
3	6	22	11
4	11	17	21
5	17	22	26
6	25	37	38
7	40	50	53
8	52	52	54
9	61	47	68
10	68	44	58
11	93	47	61
12	110	43	61

These results show that the trend extraction accuracy gap between our method and the EEMD-energy-ratio approach is varying with the degrees of the seasonality in the data. Our method gives more accurate results especially in the case of prominent seasonality component. We also observe high errors related to edge effects (Fig. 11), which are out of the scope of this paper.

We conclude that the EMD (EEMD) is an attractive alternative approach for trend extraction and the efficiency of the EEMD-ISC algorithm for trend extraction that can be applied, unlike other methods, to a large family of standard time series.



(a)



(b)

Fig. 11. Examples of trend extracted from simulated time series: EMD-energy-ratio and EEMD-ISC trends extracted from a realization of $y^{(3)}$ (a) ($\delta = 5\%$) and (b) $y^{(6)}$ ($\delta = 5\%$).

6.3. *EEMD-ISC vs. Hodrick–Prescott filter*

In time series analysis, there are different approaches to deal with time series trend extraction: Moving Average, X-12, Tramo–Seats, and Hodrick–Prescott filter are the most widely used trend extraction methods by economists [Alexandrov *et al.* (2009)].

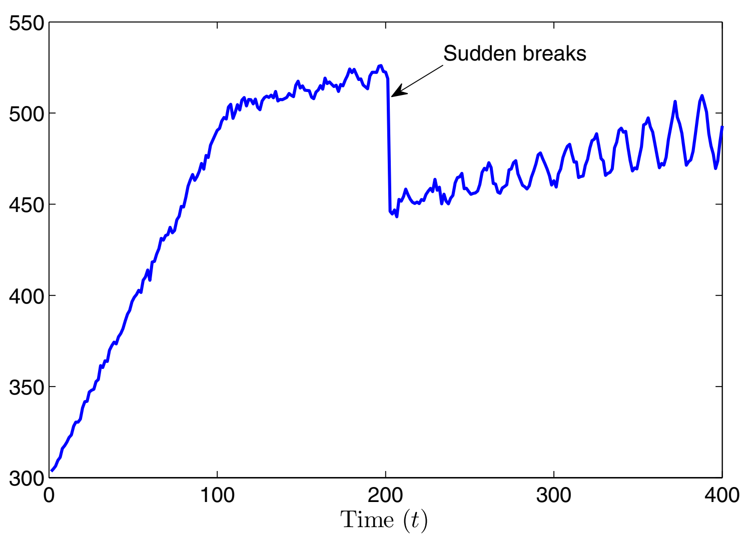
In this section, we make comparison between EEMD-ISC and the Hodrick–Prescott filter. Let us recall that the Hodrick–Prescott strategy consists, for a time series $y(t) = T(t) + C(t)$ supposed to contain a trend $(T(t))$ and a cyclical component $(C(t))$, to extract trend $(\widehat{T}(t))$ by finding the solution of:

$$\min_{\{\widehat{T}(t)\}_{t=1}^{N-1}} \left\{ \sum_{t=1}^{N-1} (y(t) - \widehat{T}(t))^2 + \lambda \sum_{t=2}^{N-1} [(\widehat{T}(t+1) - \widehat{T}(t)) - (\widehat{T}(t) - \widehat{T}(t-1))]^2 \right\}, \tag{11}$$

where the parameter λ is a positive number that penalizes variability in the growth rate of the trend component. The larger value of λ, the smoother the trend extracted, and then a good extraction of a trend requires a suitably chosen value of λ. If $C(t)$ as well as the second difference of $T(t)$ is normally and independently distributed, the HP filter is an "optimal filter" and the optimal value of λ is equal to the ratio of the two variances $\sigma^2_{C(t)}/\sigma^2_{\Delta^2_{T(t)}}$ of the cyclical component $C(t)$ and the second difference of the trend $\Delta^2_{T(t)}$ (see [Schlicht (2005)]).

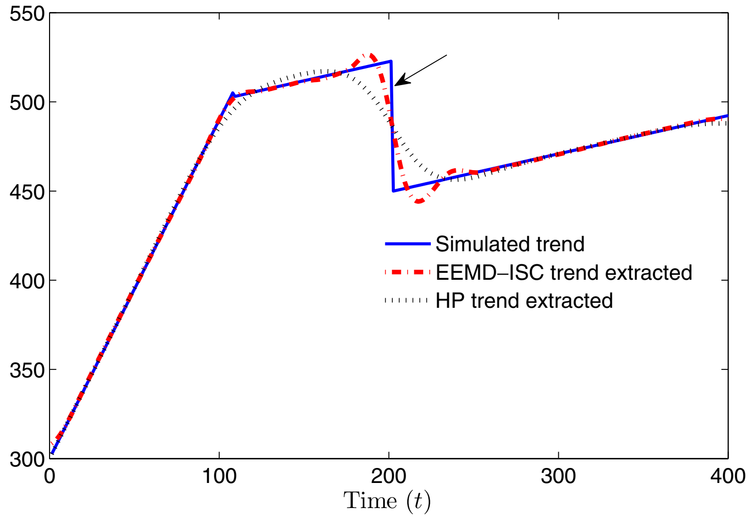
In the case of real data, those variances are unknown and the noise is not necessary normally distributed. This makes ambiguous λ optimal choice. In the case of monthly data, the classical value of λ is 14,400.

Results of various simulations seem to show that, in the case of seasonal time series, only our method gives results that can be considered close to the optimal Hodrick–Prescott ones (see [Mhamdi *et al.* (2010); Moghtaderi *et al.* (2010)]). However, it is also important to recall that, unlike Hodrick–Prescott, our method can



(a)

Fig. 12. Adjustment improvement when applying EEMD-ISC method to extract trends with sudden breaks in trend component: (a) Simulated seasonal time series with sudden breaks and (b) Delay adjustment on HP and EEMD-ISC trends extracted.



(b)

Fig. 12. (*Continued*)

be applied on any time series with any time scales fluctuations. This is the main advantages of our method compared to other classical approaches ($X12$, HP, and Moving Average).

In the case of monotonic trends, preliminary results [Mhamdi *et al.* (2010)]) indicate that the EMD approach gives similar results as the Hodrick–Prescott filter. In addition, Hodrick–Prescott, which can be considered as a filter [Schlicht (2005)], exhibits a delay adjustment when applied to extract trends with sudden breaks. Let us illustrate this by a simulation: monthly seasonal time series with sudden breaks in the trend component at $t = 200$ (Fig. 12(a)).

From Fig. 12(b), we can remark that better trend adjustment is obtained when applying our algorithm (EEMD-ISC) (Table 10).

Table 10. Statistics of estimated trend extracted errors (d) obtained through EEMD-ISC and Hodrick–Prescott in the case of sudden breaks in trend component.

	Hodrick–Prescott			EEMD-ISC		
σ^2	Min	Mean	Max	Min	MeanMax	
0.2	0.0157	0.0159	0.0161	0.0074	0.0109	0.013
0.6	0.0157	0.0158	0.0161	0.0070	0.0104	0.014
1.0	0.0158	0.0159	0.0160	0.0069	0.0107	0.015
1.2	0.0158	0.0160	0.0160	0.0075	0.0092	0.0143
1.6	0.0159	0.0160	0.0161	0.0074	0.0098	0.0150
2.0	0.0159	0.0159	0.0161	0.0072	0.0101	0.0143

7. Conclusion

EMD appears to be an attractive alternative approach for analysis and trend extraction from time series. This method does not require any optimal tuning parameter thanks to its adaptive nature and also can be applied on a time series with arbitrary unknown time scales fluctuations. In this study, we have show that the use of the EEMD (variant of EMD) ameliorates the accuracy of the EMD results for trend extraction. Based on Imfs seasonality checking criteria, we have also developed an EEMD-based method (EEMD-ISC) dealing with seasonal time series. The results obtained are very close to the optimal Hodrick–Prescott trends.

References

Alexandrov, T., Bianconcini, S., Bee Dagum, E., Maass, P. and Mc Elroy, T. (2009). A review of some modern approaches to the Problem of trend extraction. *in Research Report Series, Statistics—3*, U.S. Census Bureau, Washington.

Flandrin, P., Goncalves, P. and Rilling, G. (2004). Detrending and denoising with empirical mode decomposition. in *EUSIPCO 2004*, September 6–10, Vienna, Austria.

Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., Yen, N., Tung, C. C. and Liu, H. H. (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and nonstationary time series analysis. *Proc. R. Soc. Lond. A.*, **454**: 903–995.

Mhamdi, F., Poggi, J.-M. and Jaidane, M. (2010). Empirical mode decomposition for trend extraction: application to electrical data. *Proceedings of COMPSTAT*, August 22–27, Paris, **454**: 1391–1398.

Misiti, M., Misiti, Y., Oppenheim, G. and Poggi, J.-M. (2007). *Wavelets and Their Applications*, Hermes Lavoisier, ISTE Publishing Knowledge.

Moghtaderi, A., Flandrin, P. and Borgnat, P. (2010). Trend filtering: empirical mode decompositions vs. ℓ_1 and Hodrick–Prescott. Accepted to *Advances in Adaptive Data Analysis*.

Niazy, R. K., Beckmann, C. F., Brady, J. M. and Smith, S. M. (2009). Performance evaluation of the ensemble empirical mode decomposition. *J. Adv. Adapt. Data Anal.*, **1**(2): 231–242.

Ould Mohamed Mahmoud, M., Mhamdi, F. and Jaidane-Saidane, M. (2009). Long term multi-scale analysis of the daily peak load based on the empirical mode decomposition. *IEEE PowerTech*, June 28–July 2, Romania.

Pollock, D. S. G. (2003). Sharp filters for short sequences. *J. Stat. Plann. Infer.*, **113**: 663–683.

Ren, D., Yang, S., Wu, Z. and Yan, G. (2006). Evaluation of the EMD end effect and a window based method to improve EMD. *International Technology and Innovation Conference*, November 6–7, China.

Schlicht, E. (2005). Estimating the smoothing parameter in the so-called Hodrick–Prescott filter. *J. Japan Statist. Soc.*, **35**: 99–119.

Suling, J., Yanqin, G., Qiang, W. and Jian, Z. (2009). Trend extraction and similarity matching of financial time series based on EMD method. *World Congress on Engineering and Computer Science*, October 2009, 20–22, San Francisco.

Wu, Z., Huang, N. E., Long S. R. and Peng, C. K. (2007). On the trend, detrending, and variability of nonlinear and nonstationary time series. *PNAS*, **104**(38): 14889–14894.

Zhou, N., Trudnowski, D., Pierre, J. W., Sarawgi, S. and Bhatt, N. (2008). An algorithm for removing trends from power-system oscillation data. *IEEE PES*, Pittsburgh, PA.

Zhaohua, W. and Huang. N. E. (2004). A study of the characteristics of white noise using the empirical mode decomposition method. *Roy. Soc. Lond.*: 1597–1611