ARTICLE IN PRESS

Computational Statistics and Data Analysis ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Trend filtering via empirical mode decompositions

Azadeh Moghtaderi [a,b,*], Patrick Flandrin [b], Pierre Borgnat [b]

[a] Department of Mathematics and Statistics, Queen's University, Kingston, Ontario, Canada K7L 3N6

[b] CNRS, École Normale Supérieure de Lyon, Laboratoire de Physique, 46 allée d'Italie 69364 Lyon CEDEX 07, France

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
Empirical mode decomposition
Trend filtering
Adaptive data analysis
Monthly mean carbon dioxide cycle
Seasonality

ABSTRACT

The problem of filtering low-frequency trend from a given time series is considered. In order to solve this problem, a nonparametric technique called empirical mode decomposition trend filtering is developed. A key assumption is that the trend is representable as the sum of intrinsic mode functions produced by the empirical mode decomposition (EMD) of the time series. Based on an empirical analysis of the EMD, an automatic procedure for selecting the requisite intrinsic mode functions is proposed. To illustrate the effectiveness of the technique, it is applied to simulated time series containing different types of trend, as well as real-world data collected from an environmental study (atmospheric carbon dioxide levels at Mauna Loa Observatory) and from a bicycle rental service (rental numbers of Grand Lyon Vélo'v).

1. Introduction

Many real-world time series exhibit a "composite" behavior, in the sense that such a time series can be decomposed into a superposition of two "components". Typically one of these components can be classified as "trend", while the other component is classified as "fluctuation". (The term "residual" is sometimes used instead of "fluctuation". In this paper, however, the term residual has a precise meaning in the context of the empirical mode decomposition; see Section 2.) The problem of effecting such a decomposition, and classifying the resulting components as trend or fluctuation, is called the *trend filtering problem* (or *trend estimation problem*). Solving this problem is desirable, since an analysis of the trend component of a time series can often yield valuable information which can be used, e.g., for prediction. An obvious initial barrier to solving the trend filtering problem is that the term "trend" is highly context-dependent. In general, one must adopt an *ad hoc* definition of trend.

A common *ad hoc* definition of trend is that of a "long-term change in the mean" (Chatfield, 1996; Alexandrov et al., 2008). This leads to regression-based techniques, where the trend component is described, for example, by a low-degree polynomial. Other techniques do not impose such a strict definition. Nonparametric trend filtering assumes that the fluctuation possesses generic stationarity properties, and that the trend can be found by an *ad hoc* smoothing operation, e.g., using the Henderson filter (Henderson, 1916) or Hodrick–Prescott filter (Hodrick and Prescott, 1997; Maravall and del Río, 2007). One can also interpret the trend filtering problem in the frequency-domain sense—that is, one can assume the trend is represented by a particular set of low-frequency oscillations. This turns the trend filtering problem into a *bona fide* filtering problem. Viewed in this way, it may be profitable to use Wiener–Kolmogorov filtering (Pollock, 2006) to solve the trend filtering problem. Yet more parametric and semiparametric approaches have been proposed for modeling trend

* Corresponding author at: Department of Mathematics and Statistics, Queen's University, Kingston, Ontario, Canada K7L 3N6. Tel.: +1 613 533 2425; fax: +1 613 533 2964.

E-mail addresses: azadeh@mast.queensu.ca, azadeh.moghtaderi@gmail.com (A. Moghtaderi), patrick.flandrin@ens-lyon.fr (P. Flandrin), pierre.borgnat@ens-lyon.fr (P. Borgnat).

0167-9473/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2011.05.015

(Beran and Feng, 2002). Finally, it is worthwhile to mention that generalized "trend cycles", defined as "short-term trend [that] generally includes cyclical fluctuations", have also been studied (Alexandrov et al., 2008). Deciding if trend cycles should be considered as trend depends on the application and, of course, the observation scale.

In this paper, we introduce a novel approach to solving the trend filtering problem. This approach, which we call *empirical mode decomposition trend filtering*, is based on the following definition: The trend component of a time series is "slowly varying" in the sense that it is represented by the "slowest" intrinsic mode functions produced by the *empirical mode decomposition* (EMD) (Huang et al., 1998). By examining certain properties of the intrinsic mode functions' energies and zero crossing numbers, we will attempt to answer the question "Which of the intrinsic mode functions should be deemed the slowest?" In particular, we will provide evidence that certain changes in these properties characterize the "tipping point" between trend and fluctuation.

It must be mentioned that the use of the EMD to solve the trend filtering problem has already been proposed in the literature. However, such work has either relied on an *a priori* model for the fluctuation (Flandrin et al., 2004a), or has considered the trend as being the final residual time series produced by the EMD (Wu et al., 2007). In a sense, using the EMD to solve the trend filtering problem shares common features with *singular-spectrum analysis* applied to the same problem (Vautard and Ghil, 1989; Ghil and Vautard, 1992; Vautard et al., 1991). This is because the SSA also effects a decomposition into oscillatory components. Like the EMD-based method proposed by Wu et al. (2007), a possible approach to solving the trend filtering problem using SSA is to identify the trend as the lowest frequency oscillatory component. Other possibilities are to look for oscillatory components with prescribed smoothness or monotonicity properties; see Alexandrov et al. (2008).

The rest of the paper is organized as follows. In Section 2, we briefly review some background material concerning the EMD. In Section 3, we state what trend means in the context of this paper. In Section 4, we describe the details of EMD trend filtering. The performance of EMD trend filtering is demonstrated in Sections 5–7 through analyses of simulated and real-world time series. Finally, we make concluding remarks in Section 8.

Although this paper is intended to be self-contained, due to limited space, we have not included all the simulations supporting the proposed trend filtering method. See Moghtaderi et al. (2011) for additional material.

2. The empirical mode decomposition

The *empirical mode decomposition* (EMD) is an algorithm which decomposes a time series into a finite additive superposition of oscillatory components, each of which is called an *intrinsic mode function* (IMF); see Huang et al. (1998). The EMD does not rely on any technical assumptions concerning the nature of the time series—this includes modeling assumptions. The basic idea is that the IMFs are computed subject to two requirements: First, the number of local extrema and number of zero crossings of each IMF vary by at most one. Second, the mean of the upper and lower envelopes of each IMF should be identically equal to zero, the envelopes being computed by means of a fixed interpolation scheme. (In the numerical results presented in this paper, we have confined ourselves to the use of cubic spline interpolation.) The IMFs are computed by means of an iterative scheme. This scheme, however, depends on a stopping criterion which guarantees that (i) the requirements above are satisfied within a given tolerance and (ii) that each IMF is meaningful in both its amplitude and frequency modulations. We again refer to Huang et al. (1998) for details.

To make this description more precise, let $\boldsymbol{X} = \{X_t\}_{t\geq 0}$ be a (real, discrete-time, stochastic) process, and let $\boldsymbol{\mathcal{X}} = (X_0, X_1, \ldots, X_{N-1})$ be a realization of $\boldsymbol{X}$. (These assumptions illustrate a notational convention that is in force throughout the rest of the paper, namely that time series of length N are written in bold typeface and are regarded as elements of the Euclidean space $\mathbb{R}^N$.) As an initialization step, set $i = 1$ and $\boldsymbol{\rho}^0 = \boldsymbol{\mathcal{X}}$. The EMD computes the IMFs of $\boldsymbol{\mathcal{X}}$ using the following algorithm.

(1) Identify the local maxima and local minima of $\boldsymbol{\rho}^{i-1}$.
(2) Together with the chosen interpolation scheme, use the maxima and minima from step (1) to compute the upper and lower envelopes of $\boldsymbol{\rho}^{i-1}$.
(3) Determine the *local trend*, denoted $\boldsymbol{Q}^i$, as the mean of the upper and lower envelopes from step (2).
(4) Compute the *local fluctuation*, denoted $\boldsymbol{h} = \boldsymbol{\mathcal{X}} - \boldsymbol{Q}^i$.
(5) If $\boldsymbol{h}$ is not an IMF, in the sense that it does not satisfy the two requirements described in the beginning of this section, then go to step (1) with $\boldsymbol{\rho}^{i-1} = \boldsymbol{h}$. (Huang et al. (1998) call this the *sifting process*; it is this process which depends on the stopping criterion.)
(6) If $\boldsymbol{h}$ is an IMF, in the sense that it satisfies the two requirements described at the beginning of this section, then the ith *intrinsic mode function* of $\boldsymbol{\mathcal{X}}$ is $\boldsymbol{\mathcal{M}}^i = \boldsymbol{h}$, and the ith *residual* is $\boldsymbol{\rho}^i = \boldsymbol{\mathcal{X}} - \boldsymbol{\mathcal{M}}^i$. Increment i, and then go to step (1).

The algorithm halts when the ith residual has no further oscillations, in the sense that it has no local maxima or local minima. We denote by $\mathcal{I}$ the largest index for which $\boldsymbol{\mathcal{M}}^i$ is defined. Then

$$\boldsymbol{\mathcal{X}} = \sum_{i=1}^{\mathcal{I}} \boldsymbol{\mathcal{M}}^i + \boldsymbol{\rho}^{\mathcal{I}}. \tag{1}$$

Please cite this article in press as: Moghtaderi, A., et al., Trend filtering via empirical mode decompositions. Computational Statistics and Data Analysis (2011), doi:10.1016/j.csda.2011.05.015

In this decomposition, $\mathcal{M}^1$ through $\mathcal{M}^{\mathcal{I}}$ can be thought of as containing a "spectrum" of local oscillations in $\mathcal{X}$, with the shortest-period (highest frequency) oscillations represented in $\mathcal{M}^1$ and the longest-period (lowest frequency) oscillations represented in $\mathcal{M}^{\mathcal{I}}$. The computational complexity of the algorithm depends on $\mathcal{X}$, the chosen interpolation scheme, and the stopping criterion. However, the algorithm usually halts in a reasonably small number of steps. For example, it is known (Flandrin et al., 2004a) that if $\mathcal{X}$ is a broadband process (i.e., if it includes a relatively wide range (or band) of frequencies, with no dominant peaks), then the decomposition produced by the EMD has an almost dyadic filter-bank structure, typically with $\mathcal{I} \approx \log_2 N$. Moreover, it is known that the sifting process typically halts after some tens of iterations (Huang et al., 1998).

3. Trend in EMD

As discussed in Section 1, the term "trend" is context-dependent. In this section we state what we mean by trend in this paper and in the context of EMD trend filtering. To begin with, let us introduce some notation. Let $\mathcal{Y} = (Y_0, Y_1, \ldots, Y_{N-1})$ be a realization of a process $\boldsymbol{Y} = \{Y_t\}_{t\geq 0}$, and let $\mathcal{C} = (C_0, C_1, \ldots, C_{N-1}) \in \mathbb{R}^n$ be a *trend* component. Assume also that $\boldsymbol{Y}$ is a broadband process with a continuous spectrum.

From $\mathcal{Y}$ and $\mathcal{C}$ we may form two new time series: The first is $\mathcal{Y} + \mathcal{C}$, the *additive mix* of $\mathcal{Y}$ and $\mathcal{C}$; the second is $\mathcal{C}\mathcal{Y}$, the *multiplicative mix* of $\mathcal{Y}$ and $\mathcal{C}$. (Here the multiplication is being performed componentwise.) In either case, we say that $\mathcal{Y}$ is *fluctuation* of the mix. Now let $\mathcal{X}$ be the additive or multiplicative mix of $\mathcal{Y}$ and $\mathcal{C}$. The question we wish to answer is: "Solely given $\mathcal{X}$ as data, under what conditions should it be possible to accurately estimate $\mathcal{C}$ from $\mathcal{X}$"? To do so, we must constrain the trend and fluctuation of the mix in some fashion. We take the following pragmatic approach that is based on properties of EMD.

Recall that in EMD, the successive IMFs are oscillations going from high frequency to low frequency, and that this property is valid locally in time (there is not necessarily a global separation of spectrum of successive IMFs) (Huang et al., 1998). A loose "definition" of an extractable trend in this paper is that $\mathcal{C}$ is locally slowly varying as compared to $\mathcal{Y}$. Hence, a pragmatic way of satisfying this is that the trend should be obtained as the sum of the last few IMFs and the residual extracted from $\mathcal{X}$.

Let us now turn the attention to some properties of fluctuation of the mix which can also define (in contrast) the extractable trend. First, and in agreement with Flandrin et al. (2004b) and Wu and Huang (2004), the mean frequency of the successive IMFs of broadband processes decrease, similarly to constant-Q filter-banks, with a factor near 2. This will be the first criterion studied in Section 4.1 by estimating the mean frequency from the number of zero crossings of IMFs. Second, the finding of Rilling et al. (2005) is that the "energy" of the IMFs of fractional Gaussian noise (fGn) processes decreases as the index of the IMFs increases. In Section 4.2, we will provide evidence that this property holds for broadband processes other than fGn. An explicit assumption in our work is that the fluctuation $\mathcal{Y}$ contaminating the trend $\mathcal{C}$ has such an energy profile. This does not exclude situations with a substantial energy increase downwards low frequencies, as is the case for fGns with Hurst exponent $H > 1/2$. Indeed the decreasing energy condition does not apply directly to the broadband processes, but to their IMFs. In practice, given the previously mentioned dyadic structure for the IMF spectra, processes $\boldsymbol{Y}$ with power spectra diverging as $f^{-\alpha}$ at the zero frequency are admissible provided that $\alpha < 1$.

In the presence of a trend, the prescription used in this paper is that the IMF index which shows a rupture in the two properties described above separates the trend from the fluctuation. It follows from this prescription that a trend in the present work is neither restricted to be monotonic nor to be some polynomial functions. The trend in this work can however contain oscillations while in Wu et al. (2007) only the residual of EMD was deemed a trend, hence constraining it to have no oscillations at all.

In the following two sections, we will describe in detail the properties discussed above and their abilities in separating trend and fluctuation.

4. EMD trend filtering

Let $\mathcal{X}$ be the additive mix of $\mathcal{Y}$ and $\mathcal{C}$, where these entities are given as in the previous section. As described there, our goal is to accurately estimate $\mathcal{C}$ from $\mathcal{X}$. This section is devoted to describing *EMD trend filtering* which can be used to obtain such an estimate. The following notation and terminology will be employed throughout this section. Let $\mathcal{M}^i$ be the IMFs of $\mathcal{X}$, where $1 \leq i \leq \mathcal{I}$, and let i_* be such that

$$\mathcal{C}_{i_*} = \sum_{i=i_*}^{\mathcal{I}} \mathcal{M}^i + \boldsymbol{\rho}^{\mathcal{I}} \tag{2}$$

is the best approximation to $\mathcal{C}$ in the Euclidean metric. We call i_* the *best index* and $\mathcal{C}_{i_*}$ the *best approximation* of $\mathcal{C}$. Estimating $\mathcal{C}$ is equivalent to estimating the best index. If $\hat{i}_*$ is an estimate of i_*, then we denote by $\hat{\mathcal{C}}_{i_*}$ the corresponding estimate of $\mathcal{C}$. If the mix is multiplicative and the elements of $\mathcal{C}$ are positive, then the situation reduces to the additive case. Indeed, one can take logarithms to obtain $\log|\mathcal{X}| = \log \mathcal{C} + \log|\mathcal{Y}|$, where the logarithm and absolute value functions are being applied elementwise.

EMD trend filtering, described over the course of the next three subsections, actually consists of three approaches to estimating i_*. These are called, respectively, the *ratio, energy*, and *energy-ratio* approaches. We note that the simulations

Please cite this article in press as: Moghtaderi, A., et al., Trend filtering via empirical mode decompositions. Computational Statistics and Data Analysis (2011), doi:10.1016/j.csda.2011.05.015

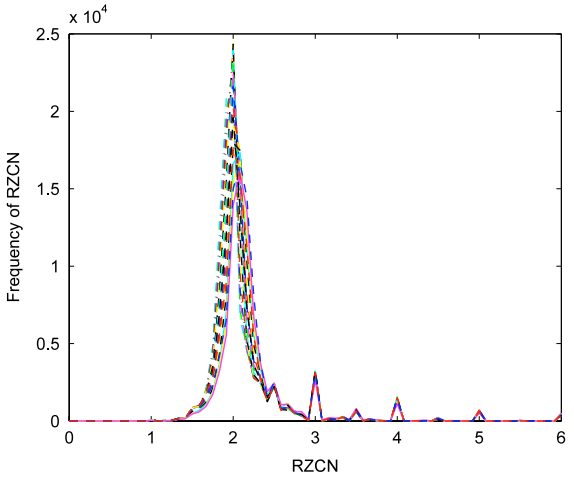

Fig. 1. Empirical distribution of the elements of $\vec{R}$ for broadband data: computed for 10,000 realizations of 20 broadband processes in the collection. Each line of different type associates with a broadband process in the collection.

provided in the next two subsections are for additive mixes only. However, we emphasize that the same outcome applies to the multiplicative mixes but the appropriate simulations can be found in the Appendix in Moghtaderi et al. (2011).

4.1. Ratio approach

In this subsection we describe the first approach to estimate i_*, which is based on an empirical property of the zero crossing numbers of IMFs.

Let us first establish some additional notation. For a given time series, the zero crossing number of its ith IMF is denoted by Z^i, and let us define $R^i = Z^{i-1}/Z^i$ for $i \geq 2$. (This is well-defined since $Z^i \geq 1$; see Section 2.) Of course, R^i depends fundamentally on the given time series; since the particular time series is always clear from context, we suppress this dependence. We call R^i the ith *ratio of the zero crossing numbers* (ith RZCN). It has been observed by Flandrin et al. (2004b) and Wu and Huang (2004) that if the time series under study is a realization of a generic broadband process, the approximation $R^i \approx 2$ holds.

Let us first support this observation. We considered 20 broadband processes of the following types: 17 fGn processes with $H = 0.1, 0.15, 0.2, \ldots, 0.9$, two stationary AR(2) processes, and a nonstationary AR(2) process with time-dependent coefficients. For each process in the collection, we simulated $B = 10{,}000$ realizations of length $N = 2000$, then computed the IMFs of each realization along with their zero crossing numbers. Denoting the ith RZCN of its bth realization by $R^{i,b}$, where $2 \leq i \leq \mathcal{I}^b$, and setting $\vec{R}^b = (R^{2,b}\ R^{3,b}\ \cdots\ R^{\mathcal{I}^b,b})$, we then computed the empirical distribution of the elements of $\vec{R} = (\vec{R}^1\ \vec{R}^2\ \cdots\ \vec{R}^B)$. Fig. 1 displays this empirical distribution, and supports the contention that $R^i \approx 2$. In fact, this distribution is approximately Gaussian with mean 2.

Generically, the approximation $R^i \approx 2$ fails for i near the best index i_*. This observation is supported by the following data. For each broadband process in the collection and using its realizations, we constructed 10,000 additive mixes, using $\boldsymbol{\mathcal{C}}^3$ (displayed in Fig. 4) as a trend. We then computed the IMFs of each mix along with their RZCNs and set $\vec{R}^b$ and $\vec{R}$ as described earlier. The left-hand plot in Fig. 2 displays the empirical distribution of the elements of $\vec{R}$ for additive mixes, and supports the contention that $R^i \approx 2$ fails. In fact this empirical distribution is non-Gaussian as its side peaks grow taller in comparison with the distribution shown in Fig. 1. The problem however is that it is not yet clear whether $R^i \approx 2$ fails around i_*. To clarify this, we proceeded with further simulations. For each broadband process in the collection, we used the IMFs obtained for each mix and used the knowledge of $\boldsymbol{\mathcal{C}}^3$ to evaluate the best index i_* (see Section 5 for details). For each mix, we then computed the *best approximation of the fluctuation* by eliminating those IMFs whose indices are greater than or equal to i_* (we call this *detrending the mix*). We set $\vec{R}^b$ and $\vec{R}$ for the remaining IMFs and computed the empirical distribution of the elements of $\vec{R}$. This distribution, shown in the right-hand plot in Fig. 2, is Gaussian with mean 2 as was the case in Fig. 1. We therefore conclude that $R^i \approx 2$ fails around i_*.

Based on what we have described above, we propose to estimate i_* by choosing $\hat{i}_*$ to be the smallest index i for which R^i is "significantly different from 2". We refer to this as the *ratio approach*. The results of our simulations for broadband processes suggest that in order to conclude whether or not R^i is "significantly different from 2", a common threshold test can be used. For $0 \leq p \leq 100$, we therefore compute p% and $(100 - p)$% significance level of the empirical distribution shown in Fig. 1 as the *left threshold* and the *right threshold* respectively. At the end, any RZCN outside of the appropriate right and left thresholds is considered *significantly different from* 2. A weakness of the ratio approach is that, since selection of the left and right thresholds is based on empirical results, it is always possible that for a given p, the smallest i for which R^i appears significantly different from 2 is a false detection. In Section 5, we will discuss how to select an "optimum" p.

Please cite this article in press as: Moghtaderi, A., et al., Trend filtering via empirical mode decompositions. Computational Statistics and Data Analysis (2011), doi:10.1016/j.csda.2011.05.015

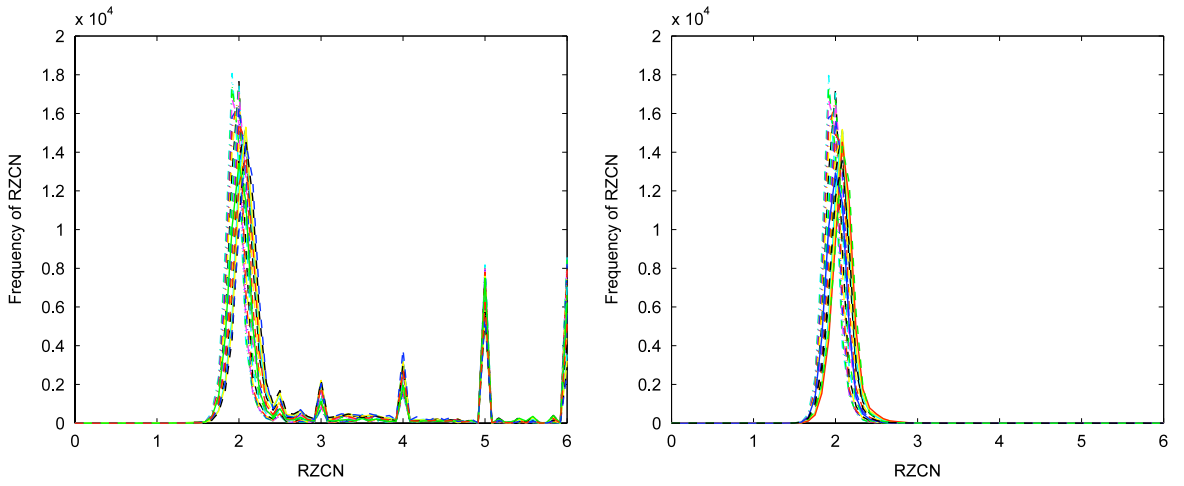

Fig. 2. Empirical distribution of the elements of $\vec{R}$ for additive mixes: left: computed for additive mixes obtained by adding $\boldsymbol{e}^3$ from Fig. 4 and realizations of broadband processes in the collection. Right: computed for detrended additive mixes.

4.2. Energy approach

In this subsection we describe the second approach to estimate i_*, which is based on an empirical property of the so-called "energy" of the IMFs. To describe this property, we need to establish some additional notation. Let $\{Z_t\}_{t\geq 0}$ be an arbitrary process. For a given time series which is a realization of $\{Z_t\}$, we define the *energy of its* ith *IMF*, denoted G^i, by

$$G^i \triangleq \sum_{t=0}^{N-1} |\mathcal{M}_t^i|^2, \quad 1 \leq i \leq \mathcal{I}.$$

Assume now that we have B different time series obtained from $\{Z_t\}$. Given the bth time series, $1 \leq b \leq B$, if $G^{i,b}$ denotes the energy of its ith IMF, the *averaged energy of its* ith *IMF* is defined by $\overline{G^i} \triangleq \frac{1}{B}\sum_{b=1}^{B} G^{i,b}$. It is shown in Rilling et al. (2005) that if the time series under study are realizations of a generic broadband process, then $\overline{G^i}$ is a decreasing sequence in i. This results were obtained by studying fGn processes. For additional simulations, see Moghtaderi et al. (2011).

Our key observation is that, generically, $\overline{G^i}$ increases for i near the best index i_*. This observation is supported by the following data. Recall the additive mixes obtained in Section 4.1. We computed the IMFs of each mix along with $G^{i,b}$ and for each broadband process in the collection, we computed $\overline{G^i}$. The left-hand plot in Fig. 3 displays $\log_2 \overline{G^i}$ computed for additive mixes. For each broadband process in the collection, we observe that $\overline{G^i}$ increases at some i but we cannot yet determine whether or not it has occurred around i_*. To clarify this, we detrended each mix as described in Section 4.1 and then recomputed $G^{i,b}$ using the remaining IMFs. For each broadband process in the collection, we then computed $\overline{G^i}$ and observed that $\overline{G^i}$ increases at the best index i_*. The right-hand plot in Fig. 3 displays $\log_2 \overline{G^i}$ computed for detrended additive mixes only up to $i = 5$. This is because for some examples, $i_* > 5$ but for the majority $i_* = 5$. Based on the above discussion, identifying the smallest index $i \geq 2$ such that $\overline{G^i} > \overline{G^{i-1}}$ evaluates $\hat{i}_*$. This approach is called the *energy approach*.

As for the ratio approach, one could think of looking for significant increases which would be based on some statistical information about the dispersion of energy of each IMF. This viewpoint has been considered first for white Gaussian noise in Huang et al. (2003) and further generalized in Flandrin and Gonçalves (2004) and Flandrin et al. (2004a), even in a detrending perspective. The limitation however is that the associated confidence intervals depend strongly on some prior knowledge about the spectra of broadband processes. This is the main reason that we do not follow such direction, as we are interested in a procedure which is not model-dependent. A limitation with the energy approach is that one is often given a single time series to use for trend filtering. Computation of energy based on only one time series may cause an increase in G^i when $i \neq i_*$.

4.3. Energy-ratio approach

To overcome limitations of the previous approaches, we introduce the last and most important approach to estimate i_*. As described, the energy and ratio approaches are confronted with possible false detections of the smallest index which does not associate with the trend. Since the criteria proposed by the energy and ratio approaches to evaluate $\hat{i}_*$ are independent, the number of false detections can be reduced by combining these two approaches.

To be more precise, for each $2 \leq i \leq \mathcal{I}$, we compute each index i such that $G^i > G^{i-1}$. For a fixed p, we also evaluate every index i where R^i is significantly different from 2. We then evaluate $\hat{i}_*$ to be the smallest common index in both approaches. This approach is called the *energy-ratio approach*.

Please cite this article in press as: Moghtaderi, A., et al., Trend filtering via empirical mode decompositions. Computational Statistics and Data Analysis (2011), doi:10.1016/j.csda.2011.05.015

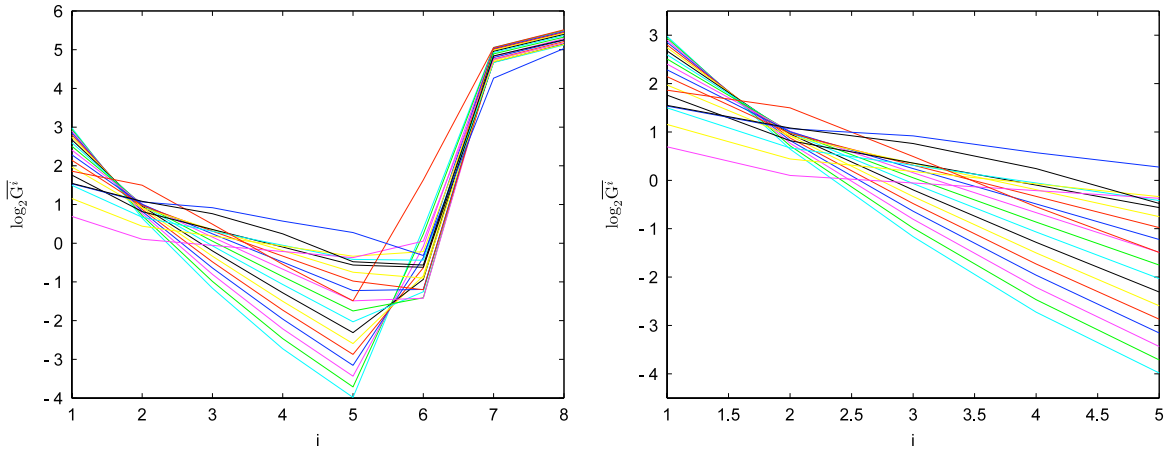

Fig. 3. $\mathbf{log_2}\overline{\boldsymbol{G}^{\boldsymbol{i}}}$: Left: computed for additive mixes. Right: computed for detrended additive mixes and displayed only up to $i = 5$.

5. Performance evaluation of the EMD trend filtering; evaluation of an "optimum" *p*

We follow two main goals in this section. The first goal is to evaluate the overall performance of the EMD trend filtering. The second goal is to empirically evaluate an "optimum" p which can improve the performance of the energy-ratio approach in comparison with the energy and ratio approaches. In order to do so, we use 6 simulated examples including 3 additive and 3 multiplicative mixes such that

$$\mathcal{X}^k = \begin{cases} \mathcal{C}^k + \mathcal{Y}^k, & 1 \le k \le 3 \\ \mathcal{C}^k \mathcal{Y}^k, & 4 \le k \le 6. \end{cases} \tag{3}$$

In order to construct the above mixes, we use the following.

Let $\boldsymbol{Y}^k = \{Y_t^k\}_{t\ge 0}$, $1 \le k \le 6$, be 6 generic broadband processes such that for $1 \le k \le 2$, we have

$$Y_t^1 = 0.8Y_{t-1}^1 - 0.4Y_{t-2}^1 + \zeta_t, \quad \text{and}$$
$$Y_t^2 = 0.2Y_{t-1}^2 + 0.5Y_{t-2}^2 + \xi_t,$$

where $\{\zeta_t\}$ and $\{\xi_t\}$ are two independent white noise processes with variance 10^4, and for $3 \le k \le 6$, we have 4 fGn processes with Hurst exponents 0.7, 0.5, 0.15, and 0.75 respectively. Let $\mathcal{Y}^k = (Y_0^k, Y_1^k, \ldots, Y_{N-1}^k)$ be a realization of $\boldsymbol{Y}^k$. Now, let us assume that $\mathcal{C}^k = (C_0^k, C_1^k, \ldots, C_{N-1}^k)$, $1 \le k \le 6$, are 6 trends where for $1 \le k \le 4$, we have 4 randomly constructed trends using piecewise linear and cubic spline techniques and for $5 \le k \le 6$, we have

$$C_t^5 = 2 - \mathrm{e}^{\frac{-(t-1000)^2}{2\times 400^2}}, \quad \text{and}$$
$$C_t^6 = 1.5 + \cos(2\pi f_s t), \quad f_s = 0.002.$$

Fig. 4 displays $\mathcal{C}^k$ for $1 \le k \le 6$ when $N = 2000$. For each k, we created $B = 10{,}000$ realizations of length $N = 2000$ of $\boldsymbol{Y}^k$ and constructed the mixes for each realization following Eq. (3). We denote the bth realization of the kth example by b_k.

In order to achieve the goals described earlier in this section, we started with the following computations. We applied EMD to $\mathcal{X}^{b_k}$ (or $\log|\mathcal{X}^{b_k}|$ for multiplicative mixes) in order to extract its IMFs. Denote $\mathcal{I}$, $\mathcal{M}^i$, and $\boldsymbol{\rho}^{\mathcal{I}}$ by $\mathcal{I}^{b_k}$, $\mathcal{M}^{i^{b_k}}$, and $\boldsymbol{\rho}^{\mathcal{I}^{b_k}}$ respectively. For each $i_\dagger \in \{1, 2, \ldots, \mathcal{I}^{b_k}\}$, we computed

$$\mathcal{C}_{i_\dagger}^{b_k} = \sum_{i=i_\dagger}^{\mathcal{I}^{b_k}} \mathcal{M}^{i^{b_k}} + \boldsymbol{\rho}^{\mathcal{I}^{b_k}}, \tag{4}$$

and the Euclidean distance (ED) between $\mathcal{C}^k$ and $\mathcal{C}_{i_\dagger}^{b_k}$, denoted $\mathrm{E}_{i_\dagger}^{b_k}$. The best index i_* is that $i_\dagger$ which results in minimum $\mathrm{E}_{i_\dagger}^{b_k}$, denoted $\mathrm{E}_{i_*}^{b_k}$. Clearly, $\mathcal{C}_{i_*}^{b_k}$ is the best approximation of $\mathcal{C}^k$. Let $\mathcal{Y}_{i_*}^{b_k} = \mathcal{X}^{b_k} - \mathcal{C}_{i_*}^{b_k}$. Here $\mathcal{Y}_{i_*}^{b_k}$ is the best approximation of the fluctuation $\mathcal{Y}^{b_k}$. We computed the Euclidean norm (EN) of $\mathcal{Y}_{i_*}^{b_k}$ and $\mathcal{Y}^{b_k}$ and denoted them by $\mathrm{E}_{i_*}^{\mathcal{Y}^{b_k}}$ and $\mathrm{E}^{\mathcal{Y}^{b_k}}$ respectively.

We then estimated $\mathcal{C}^k$ using three different trend filtering methods. The methods we used are the Hodrick–Prescott (HP) filter (Hodrick and Prescott, 1997), the singular-spectrum analysis (SSA) (Vautard et al., 1991) and the EMD trend filtering using ratio, energy and energy-ratio approaches. We denoted the trend estimates obtained above by $\hat{\mathcal{C}}_{\mathrm{m}}^{b_k}$ where the letter m indicates the type of trend filtering. For simplicity, we selected m to be r, g, and gr to refer to the ratio, energy, and energy-ratio approaches respectively. Since the ratio and energy-ratio approaches are dependent on p, we denoted the estimates

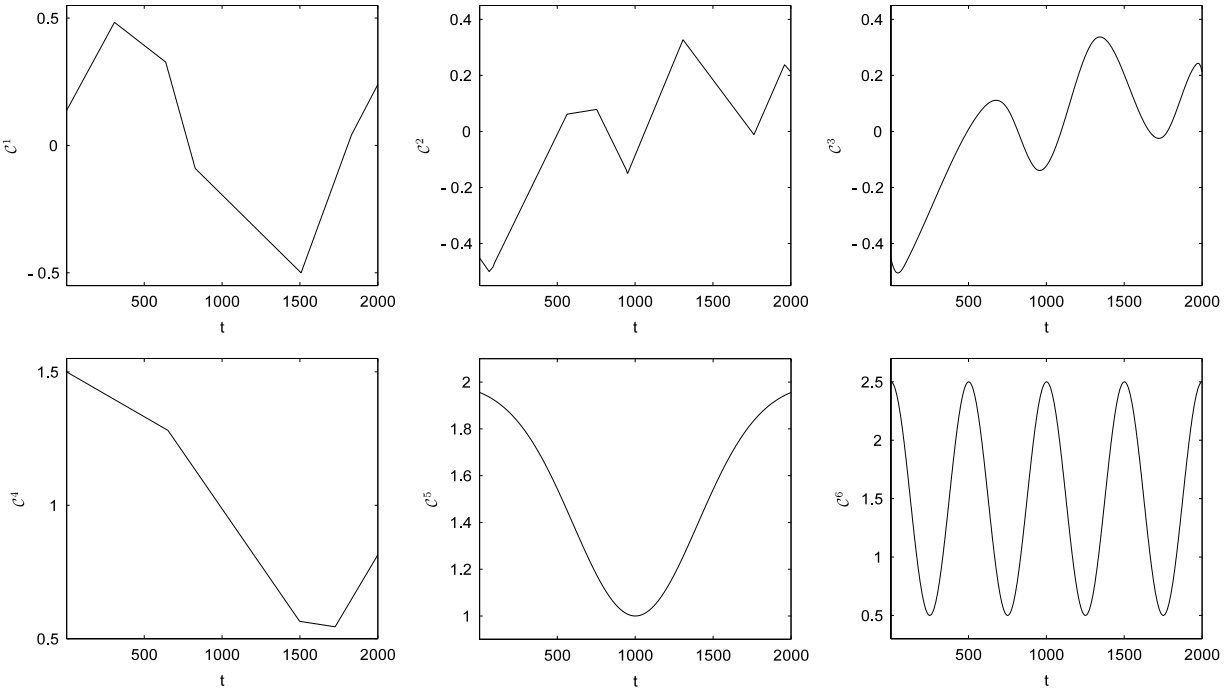

Fig. 4. Trends used in simulated examples: $\mathcal{C}^k$ for $1 \leq k \leq 6$.

Table 1
Performance evaluation of the EMD trend filtering: for $1 \leq k \leq 6$, the second to sixth columns report the average over B of the EDs between C^k and respectively $C_{i_*}^{b_k}$, $C_{\mathrm{hp}}^{b_k}$ with free parameter 10^5, $C_{\mathrm{hp}}^{b_k}$ with free parameter 5×10^5, $C_{\mathrm{ssa}}^{b_k}$ with window length 100, and finally $C_{\mathrm{ssa}}^{b_k}$ with window length 200. The last two columns are the average over B of the ENs of $\mathbf{y}_{i_*}^{b_k}$ and $\mathbf{y}^{b_k}$ respectively.

k	$\overline{\mathrm{E}}_{i_*}^k$	$\overline{\mathrm{E}}_{\mathrm{hp}}^k$	$\overline{\mathrm{E}}_{\mathrm{hp}}^k$	$\overline{\mathrm{E}}_{\mathrm{ssa}}^k$	$\overline{\mathrm{E}}_{\mathrm{ssa}}^k$	$\overline{\mathrm{E}}_{i_*}^{\mathbf{y}^k}$	$\overline{\mathrm{E}}^{\mathbf{y}^k}$
1	0.822	0.840	0.697	0.818	0.627	5.475	5.493
2	0.887	0.930	0.785	0.920	0.729	2.854	2.922
3	0.642	0.655	0.581	0.652	0.542	2.334	2.398
4	4.808	6.014	4.898	3.924	3.070	49.73	49.63
5	4.803	6.396	5.212	4.144	2.863	49.75	49.65
6	8.513	7.315	6.123	6.265	13.75	49.45	49.61

for these methods by $\hat{\mathcal{C}}_{\mathrm{m}}^{b_k,p}$. After all, we computed the ED between $\mathcal{C}^k$ and each trend estimate and denoted them by $\mathrm{E}_{\mathrm{m}}^{b_k}$ (or $\mathrm{E}_{\mathrm{m}}^{b_k,p}$).

For each k, we then averaged all the ENs and EDs computed above over B realizations and denoted them by $\overline{\mathrm{E}}_{\mathrm{m}}^k$ (or $\overline{\mathrm{E}}_{\mathrm{m}}^{k,p}$). In this paper, in order to obtain $\overline{\mathrm{E}}_{\mathrm{hp}}^k$, we used two free parameters of 10^5 and 5×10^5, for $\overline{\mathrm{E}}_{\mathrm{ssa}}^k$, we used the window lengths of 100 and 200, and for $\overline{\mathrm{E}}_{\mathrm{r}}^{k,p}$ and $\overline{\mathrm{E}}_{\mathrm{gr}}^{k,p}$, we used 16 fixed p where $1 \leq p \leq 40$. Tables 1–4 report all the averaged EDs and ENs computed using these parameters.

To evaluate the performance of the EMD trend filtering which was the first goal in this section, we make two attempts. The first attempt is to compare the best approximation of $\mathcal{C}^k$ obtained from the EMD trend filtering with estimates obtained from the HP filter and the SSA. In order to do so, for each k, we compared the reported EDs from the second column of Table 1 with those from the third to sixth columns. Since these EDs are comparable, we conclude that the EMD trend filtering performs similarly to the HP filter and the SSA. Note that since both HP filter and the SSA are dependent on free parameters, the quality of their performance can vary in comparison with the EMD trend filtering. This is clear from the reported EDs in Table 1. The second but also necessary attempt we make is to compare the fluctuation of each mix with the best approximation of the fluctuation. This is done by comparing the averaged ENs reported in the last two columns of Table 1. The fact that these two columns are comparable is an indication that the EMD trend filtering is a well-performed method in estimating the trend.

Recall the second goal in this section which is to empirically evaluate an "optimum" p which makes the energy-ratio approach to perform better than the energy and ratio approaches. We should note that by using the term "optimum" in this paper, we mean for the given examples. In order to obtain such p, we used the averaged EDs reported in Tables 2–4.

Please cite this article in press as: Moghtaderi, A., et al., Trend filtering via empirical mode decompositions. Computational Statistics and Data Analysis (2011), doi:10.1016/j.csda.2011.05.015

Table 2
Averaged EDs for ratio approach: for each k and for 16 selected fixed $1 \leq p \leq 40$, this table reports the average over B of the EDs between $\boldsymbol{c}^k$ and $\hat{\boldsymbol{c}}_{\mathrm{r}}^{b_k,p}$.

p	$\overline{\mathrm{E}}_{\mathrm{r}}^{1,p}$	$\overline{\mathrm{E}}_{\mathrm{r}}^{2,p}$	$\overline{\mathrm{E}}_{\mathrm{r}}^{3,p}$	$\overline{\mathrm{E}}_{\mathrm{r}}^{4,p}$	$\overline{\mathrm{E}}_{\mathrm{r}}^{5,p}$	$\overline{\mathrm{E}}_{\mathrm{r}}^{6,p}$
1	5.3465	5.8195	3.4415	5.4210	6.3503	21.6160
3	2.9870	3.3997	1.8074	5.4574	5.8766	17.4428
5	1.7086	1.6177	0.9166	6.0374	6.1150	13.1966
9	1.3786	1.1369	0.7699	9.1423	8.8052	12.2892
11	1.3362	1.0648	0.7504	10.9357	10.6262	12.9937
15	1.5126	1.0907	0.7971	14.7207	14.5816	15.9449
16	1.5742	1.1121	0.8143	15.4459	15.3537	16.5434
18	1.7058	1.2109	0.8583	17.2418	17.1439	18.0800
22	2.0115	1.5804	0.9826	20.3119	20.4280	21.2469
24	2.1741	1.7626	1.0589	21.9070	22.0858	22.9679
26	2.3748	1.9432	1.1559	23.4831	23.6743	31.5412
28	2.5967	2.0763	1.2576	24.9817	25.3282	26.1462
30	2.8119	2.1778	1.3505	26.4992	26.9753	27.7565
32	3.0118	2.2456	1.4363	27.8978	28.4667	29.3045
35	3.2997	2.3148	1.5463	29.7126	30.4567	31.2734
40	3.6944	2.3840	1.6949	32.7339	33.7013	34.4058

Table 3
Averaged EDs for energy-ratio approach: for each k and for 16 selected fixed $1 \leq p \leq 40$, this table reports the average over B of the EDs between $\boldsymbol{c}^k$ and $\hat{\boldsymbol{c}}_{\mathrm{gr}}^{b_k,p}$. For each k, the bold averaged EDs associate with the smallest p where $\overline{\mathrm{E}}_{\mathrm{gr}}^{k,p} < \overline{\mathrm{E}}_{\mathrm{g}}^{k}$ and the averaged EDs marked with $*$ are the minimum EDs.

p	$\overline{\mathrm{E}}_{\mathrm{gr}}^{1,p}$	$\overline{\mathrm{E}}_{\mathrm{gr}}^{2,p}$	$\overline{\mathrm{E}}_{\mathrm{gr}}^{3,p}$	$\overline{\mathrm{E}}_{\mathrm{gr}}^{4,p}$	$\overline{\mathrm{E}}_{\mathrm{gr}}^{5,p}$	$\overline{\mathrm{E}}_{\mathrm{gr}}^{6,p}$
1	3.3775	4.3465	2.5001	**5.4290**	6.3591	21.5950
3	2.2746	3.1742	1.7289	5.3000	**5.7286**	17.4446
5	1.4834	**1.5838**	0.9207	5.2036	5.2293	12.9886
9	1.2092	1.1374	0.7512	5.2021*	5.1026*	10.1622
11	**1.0713**	1.0466	**0.7022**	5.2502	5.1730	**9.5106**
15	0.9756	1.0196*	0.6815	5.3481	5.2999	9.1719
16	0.9554	1.0260	0.6796	5.3695	5.3170	9.0538
18	0.9255	1.0850	0.6769*	5.4136	5.3889	8.9986
22	0.9063	1.3705	0.6814	5.4913	5.4842	8.9490
24	0.8989*	1.5157	0.6854	5.5227	5.5367	8.9196*
26	0.9045	1.6683	0.6900	5.5499	5.5780	9.0128
28	0.9202	1.7897	0.6947	5.5840	5.6196	8.9218
30	0.9331	1.8801	0.6984	5.6165	5.6661	8.9366
32	0.9478	1.9572	0.7018	5.6320	5.6843	8.9438
35	0.9741	2.0357	0.7087	5.6720	5.7363	8.9476
40	1.0094	2.1245	0.7182	5.7155	5.7926	8.9678

Table 4
Averaged EDs for energy approach: for each k, this table reports the average over B of the EDs between $\boldsymbol{c}^k$ and $\hat{\boldsymbol{c}}_{\mathrm{g}}^{b_k}$.

	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$
$\overline{\mathrm{E}}_{\mathrm{g}}^{k}$	1.0724	2.2197	0.7345	5.8131	5.9060	9.5901

Comparing the values reported in Tables 2 and 3 shows that for majority of p and k, the averaged EDs associated with the energy-ratio approach are smaller than those for the ratio approach. This means that the energy-ratio approach performs better than the ratio approach regardless of p. As a result, selection of p should only depend on how the energy-ratio approach compares with the energy approach. We therefore compare the averaged EDs reported in Table 3 with those reported in Table 4. For each k, we select the smallest p in Table 3 such that $\overline{\mathrm{E}}_{\mathrm{gr}}^{k,p} < \overline{\mathrm{E}}_{\mathrm{g}}^{k}$ and denote it by p_1^k. We display $\overline{\mathrm{E}}_{\mathrm{gr}}^{k,p_1^k}$ in bold in Table 3 and we have $p_1^k \in \{11, 5, 11, 1, 3, 11\}$. For each k, we observe that for $p > p_1^k$, $\overline{\mathrm{E}}_{\mathrm{gr}}^{k,p}$ decreases until it reaches its minimum (denoted p_2^k and marked with a star in Table 3) and increases again but it does not exceed $\overline{\mathrm{E}}_{\mathrm{g}}^{k}$ (at least for maximum $p = 40$). This observation indicates that first of all, an "optimum" p is not unique as it depends strongly on the type of example. Second, there is a wide range of p values which make the energy-ratio approach to perform better that the energy approach. We therefore select an "optimum" p, denoted p_* to be such that $p_* > \max_k p_1^k$ and also $p_* < \max_k p_2^k$. We

Table 5
Search for i_* in examples 1 and 2: the first two rows are associated with $\mathfrak{X}^1$, where $1 \le i_\dagger \le 10$. The last two rows are associated with $\mathfrak{X}^5$, where $1 \le i_\dagger \le 11$.

$i_\dagger$	1	2	3	4	5	6	7	8	9	10	11
$\mathrm{E}^1_{i_\dagger}$	5.49	4.38	3.21	2.19	1.57	1.2	0.69	0.63	3.80	9.75	–
$\mathrm{E}^{\mathcal{Y}^1}_{i_\dagger}$	2.1e−15	3.32	4.64	5.17	5.32	5.38	5.48	5.50	6.61	11.2	–
$\mathrm{E}^5_{i_\dagger}$	47.4	36.6	27.9	19.7	14.2	10.4	8.25	5.05	3.68	2.49	10.4
$\mathrm{E}^{\mathcal{Y}^5}_{i_\dagger}$	7.5e−14	34.8	40.3	43.9	45.5	46.4	46.9	47.2	47.2	47.2	48.7

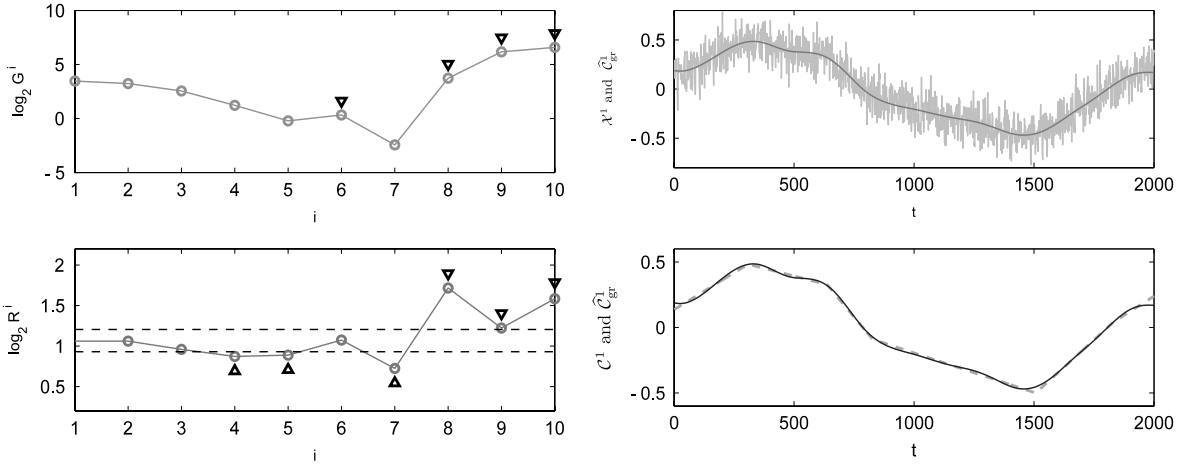

Fig. 5. EMD trend filtering for simulated example 1: top left: the energy approach. The small circles are $\log_2 G^i$ for $1 \le i \le 10$ and the small triangles mark those indices $i \ge 2$ where $G^i > G^{i-1}$. Bottom left: the ratio approach. The small circles are $\log_2 R^i$ for $2 \le i \le 10$ and the small triangles mark those indices i where R^i is significantly different from 2. The dashed lines are the averaged left and right thresholds of the empirical distribution shown in Fig. 1 when $p_* = 18$. Top right: $\mathfrak{X}^1$ vs. $\hat{\mathfrak{C}}^1_{\mathrm{gr}}$. Bottom right: $\mathfrak{C}^1$ (dashed line) vs. $\hat{\mathfrak{C}}^1_{\mathrm{gr}}$ (solid line).

therefore can select any $11 < p_* < 24$. In this paper, we have used $p_* = 18$. For a larger number of simulated examples and a larger selection of p values in achieving the goals described in this section see Moghtaderi et al. (2011).

6. Simulated examples

In this section, we demonstrate the performance of the energy-ratio approach in estimating i_* via two simulated examples. The examples we use here are the additive mix $\mathfrak{X}^1$ and the multiplicative mix $\mathfrak{X}^5$ introduced in Section 5 for further analysis. The notation used in this section is exactly the same as in Section 5 except that since we only work with one time series of each mix, we replace b_k in the notation with k. For the ratio and energy-ratio approaches, we use $p_* = 18$.

6.1. Simulated example 1

Recall $\boldsymbol{Y}^1$ and $\mathfrak{C}^1$ from Section 5. Let $\mathcal{Y}^1 = \{Y^1_0, Y^1_1, \ldots, Y^1_{N-1}\}$ be a realization of $\boldsymbol{Y}^1$. Set the additive mix $\mathfrak{X}^1 = \mathcal{Y}^1 + \mathfrak{C}^1$ for $N = 2000$. We apply EMD to $\mathfrak{X}^1$ and extract its IMFs and obtain $\mathcal{l} = 10$. Using the IMFs obtained for $\mathfrak{X}^1$, we first compute $\mathfrak{C}^1_{i_\dagger}$ for $1 \le i_\dagger \le 10$ as in Eq. (4). We then compute the EDs between $\mathfrak{C}^1$ and $\mathfrak{C}^1_{i_\dagger}$, denoted $\mathrm{E}^1_{i_\dagger}$, and the EDs between $\mathfrak{X}^1$ and $\mathfrak{C}^1_{i_\dagger}$, denoted $\mathrm{E}^{\mathcal{Y}^1}_{i_\dagger}$. These are reported in the first two rows of Table 5. Based on these reported values, we can see that since $i_\dagger = 8$ results in minimum $\mathrm{E}^1_{i_\dagger}$, we conclude that $i_* = 8$. An additional support for this selection is that $\mathrm{E}^{\mathcal{Y}^1}_8$ is the closest value to the EN of $\mathcal{Y}^1$ which is 5.493. We now want to compare the performance of the ratio, energy and energy-ratio approaches in estimating i_*.

Looking at the energy of the IMFs of $\mathfrak{X}^1$, we observe that the IMF indices for which $G^i > G^{i-1}$ are $i = \{6, 8, 9, 10\}$. Based on the energy approach, we evaluate $\hat{i}_* = 6$ which is the smallest observed index in this case. Looking at the RZCN of each IMF on the other hand, we observe that the IMF indices for which R^i is significantly different from 2 are $i = \{4, 5, 7, 8, 9, 10\}$. Based on the ratio approach, we evaluate $\hat{i}_* = 4$. Finally, the energy-ratio approach evaluates $\hat{i}_* = 8$ as the smallest common IMF index between the energy and ratio approaches. It is clear from above that the energy-ratio approach has performed excellently in estimating i_* by eliminating the false detections in the ratio and energy approaches. Fig. 5 displays the energy and ratio approaches together with the estimated trend using $\hat{i}_*$.

Please cite this article in press as: Moghtaderi, A., et al., Trend filtering via empirical mode decompositions. Computational Statistics and Data Analysis (2011), doi:10.1016/j.csda.2011.05.015

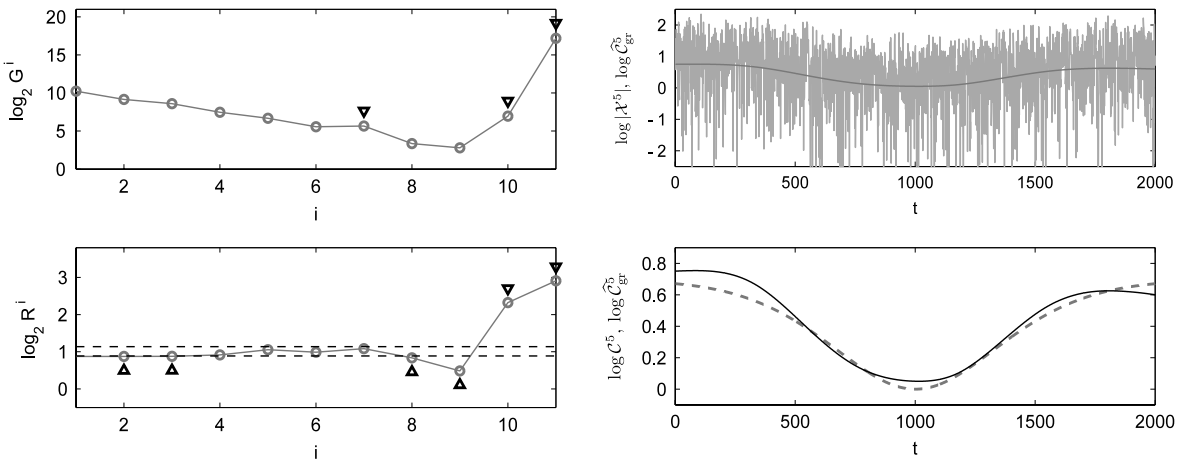

Fig. 6. EMD trend filtering for simulated example 2: top left: the energy approach. Bottom left: the ratio approach. The dashed lines are the averaged left and right thresholds of the empirical distribution of the elements of $\vec{R}$ computed for the log-transformed broadband data when $p_* = 18$. Top right: $\log|\mathcal{X}^5|$ vs. $\log\hat{\mathcal{C}}^5$. Bottom right: $\log\mathcal{C}^5$ (dashed line) vs. $\log\hat{\mathcal{C}}^5$ (solid line).

6.2. Simulated example 2

Recall $\boldsymbol{Y}^5$ and $\mathcal{C}^5$ from Section 5. Let $\mathcal{Y}^5 = \{Y_0^5, Y_1^5, \ldots, Y_{N-1}^5\}$ be a realization of $\boldsymbol{Y}^5$. Set the multiplicative mix $\mathcal{X}^5 = \mathcal{C}^5\mathcal{Y}^5$ for $N = 2000$. We apply EMD to $\log|\mathcal{X}^5|$ and extract its IMFs and obtain $\mathcal{I} = 11$. Similarly to the previous example, we use the IMFs obtained for $\log|\mathcal{X}^5|$ to first compute $\log\mathcal{C}_{i_\dagger}^5$ for $1 \le i_\dagger \le 11$. We then compute the EDs between $\log\mathcal{C}^5$ and $\log\mathcal{C}_{i_\dagger}^5$, denoted $\mathrm{E}_{i_\dagger}^5$, and the EDs between $\log|\mathcal{X}^5|$ and $\log\mathcal{C}_{i_\dagger}^5$, denoted $\mathrm{E}_{i_\dagger}^{\mathcal{Y}^5}$. These are reported in the last two rows of Table 5. Based on these reported values, we can see that since $i_\dagger = 10$ results in minimum $\mathrm{E}_{i_\dagger}^5$, we conclude that $i_* = 10$. An additional support for this selection is that $\mathrm{E}_{10}^{\mathcal{Y}^5}$ is the closest to the EN of $\log|\mathcal{Y}^5|$ which is 47.38. We now compare the performance of the ratio, energy and energy-ratio approaches in obtaining $\hat{i}_*$.

Looking at the energy of the IMFs of $\log|\mathcal{X}^5|$, we observe that $i = \{7, 10, 11\}$. Based on the energy approach, we evaluate $\hat{i}_* = 7$. Looking at the RZCN of each IMF on the other hand, we observe that $i = \{2, 3, 8, 9, 10, 11\}$. Based on the ratio approach, we evaluate $\hat{i}_* = 2$. Finally, the energy-ratio approach evaluates $\hat{i}_* = 10$ as the smallest common IMF index between the energy and ratio approaches. It is clear from above that the energy-ratio approach has performed excellently in estimating i_* by eliminating the false detections in the ratio and energy approaches. Fig. 6 displays the energy and ratio approaches together with the estimated log-trend using $\hat{i}_*$.

7. Real-world examples

In this section we demonstrate the performance of the energy-ratio approach via two real-world examples. The first example is the monthly mean carbon dioxide (CO_2) data from Mauna Loa and the second example is the Grand Lyon Vélo'v bicycle rental data from the city of Lyon in France.

7.1. Monthly mean CO_2 at Mauna Loa

In this section, we analyze the monthly mean CO_2 data collected from March 1958 to March 2010 and measured at Mauna Loa observatory in Hawaii (Available via FTP: ftp://ftp.cmdl.noaa.gov/ccg/co2/trends/co2_mm_mlo.txt. The authors have received permission from Dr. Pieter Tans in order to use this data.) The left-hand plot in Fig. 7 displays the monthly mean CO_2 data at Mauna Loa. After removing the averaged seasonal cycle expected in this data, a trend is obtained. This trend is given at the URL together with the data, and it will serve as a reference for a comparison with the result from EMD trend filtering. For more information on the known seasonal cycle and trend calculation see the URL provided above. The right-hand plot in Fig. 7 displays the one year cycles of the monthly mean CO_2 data after removing the expected trend together with their average. We call this average the *expected annual cycle*.

We now use EMD trend filtering for monthly mean CO_2 data in order to estimate its underlying trend. Applying EMD to this data, we obtain $\mathcal{I} = 3$ and following the energy-ratio approach, we evaluate $\hat{i}_* = 3$. The left-hand plot in Fig. 8 displays the estimated trend plotted together with the expected trend obtained from removing the seasonal cycle. Since these two trends look very similar, the smaller plot is made to display only a small portion of these trends. It is clear that the estimated trend from the EMD trend filtering is only a smoother version of the expected trend.

Please cite this article in press as: Moghtaderi, A., et al., Trend filtering via empirical mode decompositions. Computational Statistics and Data Analysis (2011), doi:10.1016/j.csda.2011.05.015

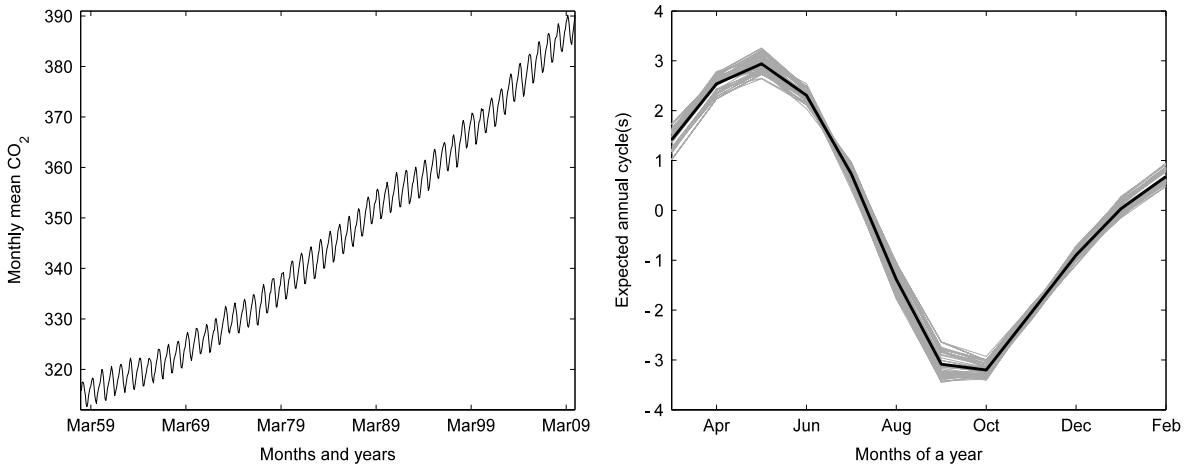

Fig. 7. Monthly mean CO_2 data and the expected annual cycle: left: monthly mean CO_2 data from March 1958 to March 2010. Right: yearly cycles of the detrended data using the expected trend together with their average (dark black line).

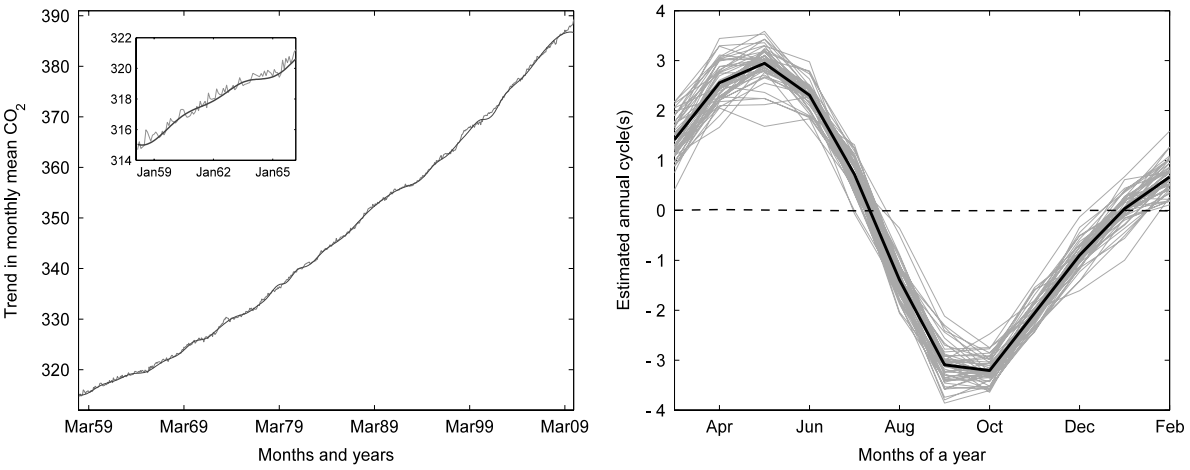

Fig. 8. Estimated trend and annual cycle for the monthly mean CO_2 data: left: expected trend together with the estimated trend using EMD trend filtering. Since these two trends look very similar, the smaller plot is made to display only a small portion of these trends. Right: yearly cycles of the detrended monthly mean CO_2 data using the estimated trend together with their average (dark black line). The dashed line displays the difference between the expected and estimated annual cycles.

After subtracting the estimated trend from the data, we divide the detrended data into one year cycles and then average over all cycles to obtain the *estimated annual cycle*. The right-hand plot in Fig. 8 displays all the one year cycles of the monthly mean CO_2 data after removing the estimated trend together with the estimated annual cycle. The dashed line in Fig. 8 displays the difference between the expected and estimated annual cycles. This difference confirms the strong similarities between the two annual cycles.

7.2. Grand Lyon Vélo'v

In this section we analyze the data from Vélo'v, the community shared bicycle program that started in Lyon in May 2005 (for more information, see http://www.velov.grandlyon.com). The program Vélo'v is a major initiative in public transportation, in which bicycles are proposed to rental by anyone at fully automated stations in many places all over the city, to be returned at any other station. Such a community shared system offers both a new and versatile option of public transportation, and a way to look into the movements of people across the city. In order to understand the dynamics of this system, a question is to estimate and model the evolution in time of the number of rentals made throughout the city (Borgnat et al., 2009). The left-hand plot in Fig. 9 displays the raw data which is the number of hourly rentals for two years of activity of the Vélo'v system from December 2005 to December 2007. (The authors would like to thank JCDecaux for providing an access to this data.)

The number of rentals contains cyclical patterns over the day (e.g., more activities during the day, mainly at specific rush hours, than during the night) and the week (e.g., more activities during week-days than week-ends). It also contains

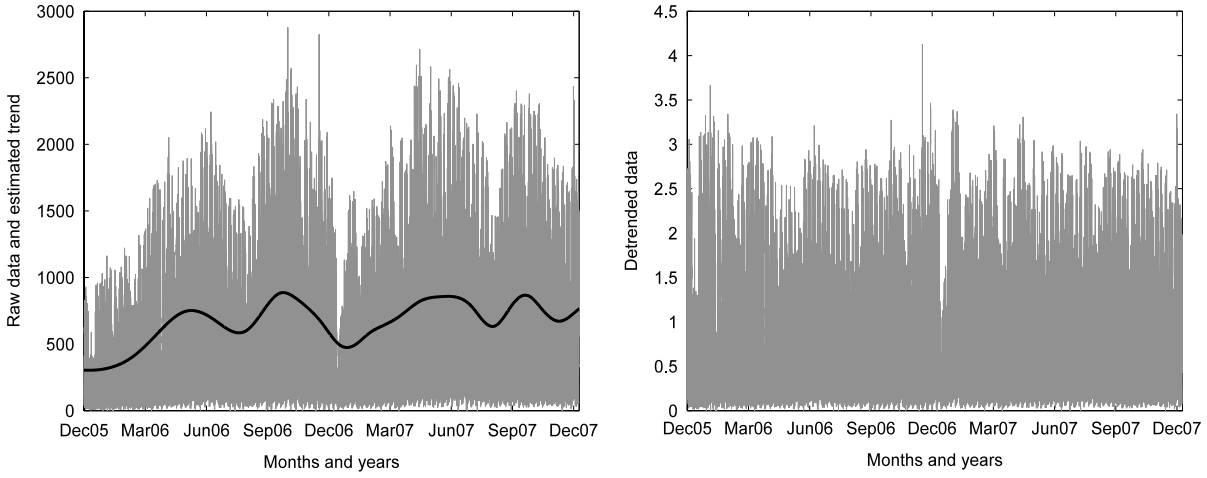

Fig. 9. Vélo'v raw and detrended data: left: the raw Vélo'v data together with the estimated trend using EMD trend filtering. Right: detrended Vélo'v data.

superimposed fluctuation due to external contingencies (e.g., rain or holidays) and a general multiplicative trend over the months (Borgnat et al., 2009). We apply EMD trend filtering to this data in order to estimate the underlying multiplicative trend. We obtain $\imath = 12$ and using the energy-ratio approach we evaluate $\hat{i}_* = 10$ which we use to estimate the trend. The estimated trend is displayed in the left-hand plot in Fig. 9 where superimposed over the data.

This trend is meaningful for the data, and can be related to, and explained by, two effects: (i) the system was expanded in 2005 and 2006 at the same time it was already in exploitation, hence, there is a long-term increase of the hourly rentals over the two years of data, (ii) because of seasonal effects, the use of Vélo'v is smaller during winter, and also during the main summer holidays; this causes several drops of the trend, during winter and also summer holidays.

Using the estimated trend, detrended Vélo'v data are obtained by dividing the number of hourly locations by the estimated trend. This is displayed in the right-hand plot in Fig. 9. The detrended data is, visually, more stationary than the raw data. This allows a good estimation of the cyclic pattern over the week of the number of hourly rentals. Fig. 10 displays the weekly cycles of the Vélo'v data after removing the estimated trend, and the average over all the weeks. The estimate of the average usage of the Vélo'v bicycles as a function of time in the week, is meaningful in that it reveals the main features of the Vélo'v activity: during week-days, there are three sharp peaks of rentals in the morning, noon and the end of the afternoon; during the week-ends, there are small peaks at noon, and smooth and large peaks during the afternoon.

Finally, let us note that here the multiplicative trend filtering procedure was applied to a case where the underlying process that the trend multiplies to is not actually a broadband process: it is more specifically a periodic process (with clear periods of one week and one day) with added fluctuation. Nevertheless, the procedure is able to find the relevant multiplicative trend describing the evolutions at the scale of the seasons, and that is used to detrend the data. This is believed to be due mostly to the fact that the fluctuation has typical periodic scales (one day or one week) which are much smaller than the typical scale of evolution (several months) of the trend, making of this scale separation a prerequisite that might be more important than the existence of a broadband spectrum in a stricter sense.

8. Conclusion

This paper proposed a new procedure, called EMD trend filtering, to solve the trend filtering problem. The essential assumption is that the trend is described by a set of low-frequency IMFs. To decide which IMFs comprise the trend, we used the observations that the trend causes (i) a change in ratios of zero crossing numbers and (ii) an increase in the IMFs' energy compared to the expected behavior of broadband processes. EMD trend filtering was shown to perform well on several time series with additive or multiplicative trends. We emphasize that – being based on the EMD – the procedure is fully data-driven and depends only on the choice of significance level.

The effectiveness of EMD trend filtering was demonstrated via its application to simulated and real-world time series. The real-world studies examined CO_2 data (where the trend entered additively) and bicycle rental data (where the trend entered multiplicatively). In either case, trend filtering allowed us to estimate the periods of cyclical components in the data. Interestingly, the fluctuation of the bicycle rental data did not seem to meet the regularity assumptions imposed in the paper. In particular, the fluctuation seemed to display greater oscillatory behavior than the broadband assumption would permit. This indicates that EMD trend filtering may be robust to departures from the broadband assumption.

As future work, one may be able to go beyond trend filtering and use the same type of approach to group the IMFs describing the trend, then the major cycles, and finally the fluctuation. This would permit automatic and model-free decompositions of time series.

Please cite this article in press as: Moghtaderi, A., et al., Trend filtering via empirical mode decompositions. Computational Statistics and Data Analysis (2011), doi:10.1016/j.csda.2011.05.015

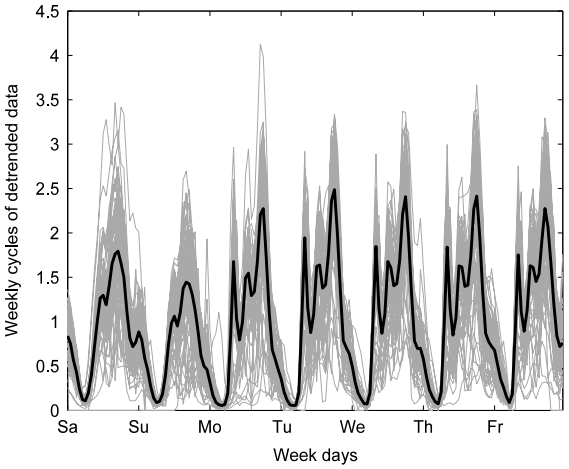

Fig. 10. Weekly detrended Vélo'v data: the weekly cycles of the detrended Vélo'v data and their average (dark black line).

Acknowledgments

The authors would like to thank the anonymous reviewers and the Associate Editor for their helpful comments and suggestions. Most of the work reported here was completed during the postdoctoral stay of Azadeh Moghtaderi at ENS Lyon, which was supported by ANR Grant ANR-07-BLAN-0191-01 StaRAC.

References

Alexandrov, T., Bianconcini, S., Dagum, E.B., Maass, P., McElroy, T., Mar. 2008. A review of some modern approaches to the problem of trend extraction. Tech. Rep. RRS2008/03, U.S. Census Bureau, Washington, DC.

Beran, J., Feng, Y., 2002. SEMIFAR models—a semiparametric approach to modeling trends, long-range dependence and nonstationarity. Computational Statistics and Data Analysis 40, 393–419.

Borgnat, P., Abry, P., Flandrin, P., Rouquier, J.-B., 2009. Studying Lyon's Vélo'v: a statistical cyclic model. In: Proceedings of ECCS'09 (European Conference of Complex Systems), Warwick, United Kingdom.

Chatfield, C., 1996. The Analysis of Time Series: An Introduction. Chapman and Hall/CRC, London, UK.

Flandrin, P., Gonçalves, P., 2004. Empirical mode decompositions as data-driven wavelet-like expansions. International Journal of Wavelets, Multiresolution and Information Processing 2 (4), 477–496.

Flandrin, P., Gonçalves, P., Rilling, G., 2004a. Detrending and denoising with empirical mode decompositions. In: Proceedings of EUSIPCO-04, Vienna, Austria, pp. 1581–1584.

Flandrin, P., Rilling, G., Gonçalves, P., 2004b. Empirical mode decomposition as a filter bank. IEEE Signal Processing Letters 11 (2), 112–114.

Ghil, M., Vautard, R., 1992. Interdecadal oscillations and the warming trend in global temperature time series. Nature 58, 95–126.

Henderson, R., 1916. Note on graduation by adjusted average. Transactions on the Actuarial Society of America 17, 43–48.

Hodrick, R.J., Prescott, E.C., 1997. Postwar US business cycles: an empirical investigation. Journal of Money, Credit, and Banking 29 (1), 1–16.

Huang, N.E., Shen, Z., Long, S.R., Wu, M.L., Shih, H.H., Zheng, Q., Yen, N.C., Tung, C.C., Liu, H.H., 1998. The empirical mode decomposition and Hilbert spectrum for nonlinear and non-stationary time series analysis. Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences 454, 903–995.

Huang, N.E., Wu, M.-L.C., Long, S.R., Shen, S.S.P., Qu, W., Gloersen, P., Fan, K.L., 2003. A confidence limit for the empirical mode decomposition and Hilbert spectral analysis. Proceedings of the Royal Society of London A 459 (2037), 2317–2345.

Maravall, A., del Río, A., 2007. Temporal aggregation, systematic sampling, and the Hodrick–Prescott filter. Computational Statistics and Data Analysis 52, 975–998.

Moghtaderi, A., Flandrin, P., Borgnat, P., Feb. 2011. Trend filtering via empirical mode decompositions, Tech. Rep. ensl-00565293, version 1, École Normale Supérieure de Lyon. URL http://prunel.ccsd.cnrs.fr/ensl-00565293.

Pollock, D.S.G., 2006. Wiener–Kolmogorov filtering, frequency-selective filtering and polynomial regression. Econometric Theory 23, 71–83.

Rilling, G., Flandrin, P., Gonçalves, P., 2005. Empirical mode decomposition, fractional Gaussian noise, and Hurst exponent estimation. IEEE International Conference on Acoustics, Speech, and Signal Processing 489–492.

Vautard, R., Ghil, M., 1989. Singular-spectrum analysis in nonlinear dynamics with applications to paleoclimatic time series. Physica D 35, 395–424.

Vautard, R., Yiou, P., Ghil, M., 1991. Singular-spectrum analysis: a toolkit for short, noisy chaotic signals. Physica D 350, 324–327.

Wu, Z., Huang, N.E., 2004. A study of the characteristics of white noise using the empirical mode decomposition method. Proceedings of the Royal Society of London A 460, 1597–1611.

Wu, Z., Huang, N.E., Long, S.R., Peng, C.-K., 2007. On the trend, detrending, and variability of nonlinear and nonstationary time series. Proceedings of the National Academy of Sciences 4 (38), 14889–14894.