Tampere University

David Batet Cardeñoso

TREND EXTRACTION METHODS FOR TIME SERIES ANALYSIS

Faculty of Information Technology and Communication Sciences
Bachelor's thesis
November 2021

Abstract

David Batet Cardeñoso: Trend extraction methods for time series analysis
Bachelor's thesis
Tampere University
Bachelor's Programme in Mathematics and Statistical Data Analysis
November 2021

The need to extract the trend from a given time series is a common problem in a wide variety of fields, such as signal processing or econometrics. Ordinary least squares regression (OLS), locally weighted polynomial regression (LWP), moving average, wavelet decomposition and empirical mode decomposition (EMD) are examples of methods which are used for trend extraction. In the present work, the aforementioned methods are first presented and then showcased on a simulated time series. Some of the properties of the methods are discussed afterwards.

Keywords: trend extraction, trend removal, detrending, drift removal, time series analysis

Contents

1 Introduction

The technology developed during the last decades has allowed the collection and processing of data at an unprecedented scale, leading to a recent growing interest in data analysis. One of the objects of study of data analysis are time series. Broadly speaking, a time series is a sequence of values of a variable taken at distinct time instants, for instance the temperature at a weather station taken every day.

In some cases, time series can have slowly changing fluctuations, and its values can increase or decrease over a large time scale. This variation in the values of the time series, referred to as *trend* or *drift*, might impede its analysis. When this happens, it is useful to be able to identify and extract this underlying trend. In other situations, the trend is the object of interest. An example which has recently captured global attention are the upward trends detected in sea and surface temperature time series. However, the problem of trend extraction occurs naturally in many fields. It can be found in medical imaging, oceanography, econometrics, seismology, and climatology, to name a few.

The present work focuses around five methods used for trend extraction: ordinary least squares regression (OLS), locally weighted polynomial regression (LWP), moving average, wavelet decomposition and empirical mode decomposition (EMD). The structure is as follows. Chapter 2 sets the fundaments for the following chapters. In Chapter 3, the trend extraction methods are presented. The methods are then used on a simulated time series in Chapter 4, and some of its properties analyzed. Finally, Chapter 5 gathers the conclusions of the work and proposes possible paths to be taken in future work on the topic.

It should be noted that there is a wide collection of trend extraction methods, and thus the ones covered here are merely a subset of them, chosen based on how often they appeared in the reviewed literature. To provide an example, some classical methods which have not been included are the Henderson and the Hodrick-Prescott filters, as well as singular spectrum analysis (SSA) [1, 2]. In recent years, EMD has inspired a vast variety of decomposition methods, some of which have also been proposed to extract trends from time series, such as ensemble EMD (EEMD) [3, 4], the synchrosqueezed wavelet transform [5], and variational mode decomposition (VMD) [6, 7]. These methods are also not considered here.

2 Preliminaries

2.1 Time series

Definition 2.1. A *time series* is a set of observations $\{x_t\}_{t\in\mathcal{T}}$, each one associated with a different time instant $t \in \mathcal{T} \subseteq \mathbb{R}$. If the set of times $\mathcal{T}$ corresponding to the observations is continuous, e.g. an interval, then the time series is said to be *continuous-time*. If it is discrete, then the time series is *discrete-time* [8].

Remark. Since the present work revolves around finite discrete-time time series, here they are simply referred to as time series. A common assumption which is also made here is that the values $\{x_t\}$ are real and evenly spaced in time. These assumptions allow the time series to be written as $\{x_t\}_{t=1}^{N}$ where $x_t \in \mathbb{R}$. Note that if $\{y_t\}_{t\in\mathcal{T}}$ is a finite discrete-time time series with N values spaced by T which starts at time $t = t_0$, then one can define the time series $\{x_t\}_{t=1}^{N}$ as $x_t = y_{t_0+(t-1)T}$ in order to obtain the time indexing $t = 1 \ldots N$.

In time series analysis, $\{x_t\}$ is often seen as a realization of a set of random variables $\{X_t\}$. Once a model is chosen for $\{X_t\}$ and its parameters fit to $\{x_t\}$, the model can be used to describe and better understand $\{x_t\}$. The set $\{X_t\}$ is called the *underlying process* of $\{x_t\}$.

Definition 2.2. The underlying process $\{X_t\}_{t\in\mathcal{T}}$ of a time series is *wide-sense stationary* if $\mathrm{E}\left[|X_t|^2\right] < \infty$ for all $t \in \mathcal{T}$, $\mathrm{E}[X_t]$ is constant in $t \in \mathcal{T}$, and $\mathrm{cov}\left(X_{t_1}, X_{t_2}\right)$ depends solely on $t_1 - t_2$ for all $t_1, t_2 \in \mathcal{T}$ [9].

Example 2.3. A classical way to model $\{x_t\}$ is to decompose $\{X_t\}$ as

$$X_t = \tau_t + s_t + Y_t,$$

where τ_t is a slowly changing function called *trend*, s_t is a periodic function called *seasonal component*, and Y_t is the *random noise component* [8, 10]. A commonly-made assumption for this model is that $\{Y_t\}$ is wide-sense stationary. This assumption simplifies the analysis of $\{x_t\}$. One can first obtain $\{y_t\}$ by subtracting $\{\tau_t\}$ and $\{s_t\}$ from $\{x_t\}$, then leverage the theory of wide-sense stationary processes to analyze $\{y_t\}$, and finally extend the analysis naturally to $\{x_t\}$. This approach relies on being able to extract the two deterministic components of the model: the trend and the seasonal

component. Note that if $\{Y_t\}$ is a wide-sense stationary process, then in general $\{X_t\}$ will not be, since $\mathrm{E}[X_t] = \tau_t + s_t + \mathrm{E}[Y_t]$ and both τ_t and s_t depend on t.

2.2 Trends

The simple model considered in Example 2.3 includes a trend, which has been loosely defined as being a slowly changing function. The definition of trend is often taken for granted or given in an informal sense. This is due to the difficulty in finding a definition that is formally rigorous. In fact, there is currently no *precise* definition of trend with a broad consensus [11]. With this in mind, the rather imprecise definition from Example 2.3 is used in the present work.

There are many factors which can influence what is considered a trend in a given time series. The most obvious one is perhaps the choice of model, which in turn depends on the time series itself. In some cases, one might not be interested in considering a seasonal component and a trend separately, and might choose to model the time series as $X_t = \tau_t + Y_t$ instead. On the other hand, in other cases it might make more sense to use a model which decomposes the time series into a product of its components, for example, $X_t = \tau_t s_t Y_t$. In any case, as commonly done, in this thesis the trend is considered as an additive component of the time series.

Remark. Under the assumption that $\tau_t, s_t, Y_t > 0$, the model $X_t = \tau_t s_t Y_t$ can be linearized by taking the logarithm of both sides: $\log X_t = \log \tau_t + \log s_t + \log Y_t$.

The concept of trend is also strongly linked to the time scale of the phenomenon being studied, which determines what is considered a *slow* change. For instance, if a certain phenomenon lasts approximately 24 hours, one might want to interpret a sudden 10 minute sharp increase in a time series related to such phenomenon as the result of randomness rather than as a trend.

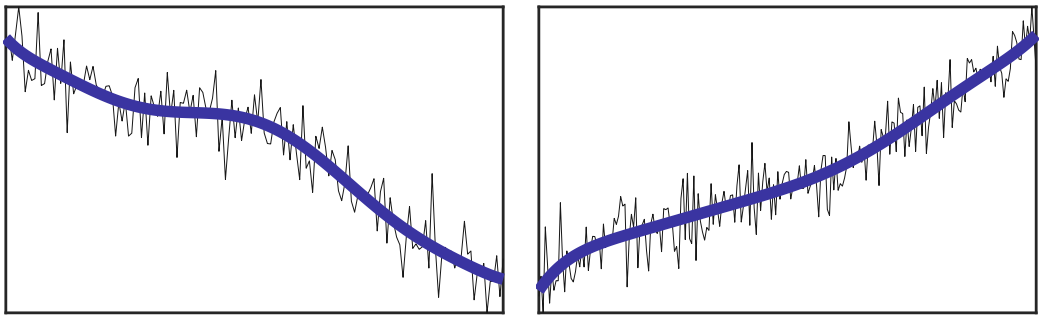

Figure 2.1. Two time series (black) and their corresponding trends (blue).

3 Trend extraction methods

3.1 Ordinary least squares regression

Using *ordinary least squares regression (OLS)* is a relatively simple way to extract the trend from a time series $x = \{x_t\}$ [12–16]. This method relies on choosing a set of predefined functions $\Phi = \{\varphi^1, \dots, \varphi^r\}$ whose linear combination suffices to obtain a good enough estimation of the trend. The trend is estimated as $\hat{\tau} = \{\hat{\tau}_t\}$, where

$$\hat{\tau}_t = \lambda^1 \varphi_t^1 + \cdots + \lambda^r \varphi_t^r$$

for some coefficients $\lambda^1, \dots, \lambda^r \in \mathbb{R}$. The coefficients are chosen so that Euclidean distance between x and the estimated trend $\hat{\tau}$ is minimized. In other words, $\lambda^1, \dots, \lambda^r$ are obtained from solving the optimization problem

$$\min_{\lambda^i \in \mathbb{R}} \|x - \hat{\tau}\|_2^2 \;=\; \min_{\lambda^i \in \mathbb{R}} \sum_{t=1}^{N} (x_t - \hat{\tau}_t)^2,$$

where $\|\cdot\|_2$ denotes the Euclidean norm.

A commonly chosen set of functions are polynomials which span the space of polynomials up to a certain degree, for example, $\Phi = \{1, t, t^2, \dots, t^{r-1}\}$. In this case, the choice of degree depends on the time series. In general, the more complex the trend is, the higher the chosen degree. Although being quite simple, this method requires the choice of predefined functions which have a great impact on the performance of the method [7]. This will be showcased in Chapter 4.

3.2 Locally weighted polynomial regression

Locally weighted polynomial regression (LWP) is another possible method which can be used to extract the trend from a time series $x = \{x_t\}$ [1]. This method is also based on minimizing the distance between the time series and the estimated trend. However, in this case the values of the estimated trend are computed locally. For each time t, a local trend $\upsilon = \{\upsilon_u\}_{u=1}^{N}$ around t is first found, and then used to define the estimated trend at time t, as $\hat{\tau}_t = \upsilon_t$.

Each local trend υ is given by

$$\upsilon_u = \lambda^1 \varphi_u^1 + \cdots + \lambda^r \varphi_u^r$$

for some coefficients $\lambda^1, \ldots, \lambda^r \in \mathbb{R}$, and a predefined polynomial basis $\Phi = \{\varphi^1, \ldots, \varphi^r\}$. The coefficients are chosen so that a predefined locally weighted distance between x and υ is minimized. In this case, $\lambda^1, \ldots, \lambda^r$ are obtained by solving the optimization problem

$$\min_{\lambda^i \in \mathbb{R}} \|x - \upsilon\|_w^2 = \min_{\lambda^i \in \mathbb{R}} \sum_{u=1}^{N} w_u (x_u - \upsilon_u)^2, \tag{3.1}$$

where $\|\cdot\|_w$ is a predefined locally weighted seminorm and $w_1, \ldots, w_N \geq 0$ its corresponding weights.

Remark. Some observations need to be made:

1. Although υ, Φ, $\lambda^1, \ldots, \lambda^r$, $\|\cdot\|_w$, and $w_1, \ldots, w_N$ depend on t, this is not included in the notation to avoid overcomplicating it.
2. In general, the seminorm $\|\cdot\|_w$ is not a norm. Note that for any z such that $z_u = 0$ if and only if $w_u > 0$, the equality $\|z\|_w = 0$ holds.
3. The values x_u for which $w_u = 0$ have no effect in the optimization problem (3.1), whereas those for which w_u is greater have a greater influence.

The weights are defined with the last observation in mind. Only the times neighboring t are assigned a nonzero weight, with those closer to t having a greater weight. In particular, a predefined integer k is chosen and the weights are given by

$$w_u = W\left(\frac{u - t}{d_t}\right),$$

where $W(\eta)$ is a predefined *weight function*, and d_t is the distance between t and its k-th nearest neighbor in time. $W(\eta)$ is often chosen so that the following holds [17]:

1. $W(\eta) = W(-\eta)$ for all $\eta \in \mathbb{R}$.
2. $W(\eta)$ is strictly positive and nonincreasing for $0 \leq \eta < 1$.
3. $W(\eta) = 0$ for $\eta \geq 1$.

Therefore, the value of the estimated trend $\hat{\tau}_t$ at each point t depends at most on the $k - 1$ nearest values from x_t in time (including x_t).

Example 3.1. An example of weight function is the *tricube* weight function, defined as $W(\eta) = (1 - |\eta|^3)^3$ for $|\eta| \leq 1$ and $W(\eta) = 0$ everywhere else [17].

This method requires the choice of several parameters. It is necessary to choose k, the weight function W, and the degree of the polynomials. Usually, polynomials of degrees 1 or 2 are used [1, 17, 18]. However, there is a classical trend extraction method called *moving average* which can be seen as a particular case of LWP using polynomials of degree 0. The method is presented below.

3.3 Moving average

Moving average is a very simple trend extraction method based on locally averaging the values of the given time series $\{x_t\}$. In this method, the estimated trend $\{\hat{\tau}_t\}$ is given by

$$\hat{\tau}_t = \frac{1}{k} \sum_{i=-q}^{q} x_{t+i}, \tag{3.2}$$

where $k = 2q + 1$ is a predefined positive odd integer. Note that the expression above is defined only for $q + 1 \leq t \leq N - q$, as x_t does not exist for $t < 1$ or $t > N$ [8]. This means that the values on the boundaries need to be treated differently. One possibility is to extend $\{x_t\}$ by mirroring it on its boundaries, so that Equation (3.2) is defined for every time t. Namely, to define $x_t = x_{1-t}$ for $t < 1$ and $x_t = x_{2N-t+1}$ for $t > N$.

Remark. As mentioned in Section 3.2, this method can be seen as a particular case of LWP. To recover the moving average method from LWP, one can choose $\Phi = \{\varphi^1 = 1\}$, $k' = k + 2$, and the *trivial* weight function, defined as $W(\eta) = 1$ for $|\eta| < 1$ and $W(\eta) = 0$ everywhere else. After these choices, the optimization problem (3.1) becomes

$$\min_{\lambda^1 \in \mathbb{R}} \|x - v\|_w^2 = \min_{\lambda^1 \in \mathbb{R}} \sum_{i=-q}^{q} (x_{t+i} - \lambda^1)^2 \tag{3.3}$$

for $q' + 1 \leq t \leq N - q'$, where $q' = (k' - 1)/2$. The moving average method then results from the fact that the value λ^1 which solves the optimization problem (3.3) is

$$\lambda^1 = \frac{1}{k} \sum_{i=-q}^{q} x_{t+i},$$

giving $\hat{\tau}_t = v_t = \lambda^1 1 = \lambda^1$. Once again, the values on the boundaries require additional attention. In this case, LWP yields the values $\hat{\tau}_t = (k\tau_{q'+1} + x_1)/(k + 1)$ for $t < q' + 1$,

and $\hat{\tau}_t = (k\hat{\tau}_{N-q'} + x_N)/(k+1)$ for $t > N - q'$. The reason why the values $\hat{\tau}_t$ on the left boundary are the same is that the optimization problems which give these values are also the same. The same happens on the right boundary.

The main drawback of this method, as will be showcased in Chapter 4, is that it is very sensitive to the choice of the parameter k. However, an interesting property of the method is that it can be interpreted as a linear time-invariant (LTI) filter, in particular, as a low-pass filter [7]. As a consequence, the theory of LTI filters can be leveraged to better understand the behavior of the method. In fact, it is possible to find examples of other LTI filters which have been used to extract trends, such as Butterworth filters [19, 20] or exponentially weighted moving average [21].

3.4 Wavelet decomposition

The *wavelet decomposition* of the given time series $\{x_t\}$ is sometimes used to obtain an estimation of its underlying trend. The time series is decomposed into several components, each corresponding to a different time scale, and the estimated trend is obtained from the largest-scale component [13].

The wavelet decomposition is the result of successively applying high-pass (HP) and low-pass (LP) filters. In the first stage of the decomposition, $\{x_t\}$ is taken as an input for the initial HP and LP filters, and the outputs are downsampled by a factor of two. The results d^1 and a^1 are called first-level *detail* and *approximation coefficients*, respectively. The coefficients d^i and a^i of the next levels are computed analogously, taking a^{i-1} as an input. The *n-th level decomposition* of $\{x_t\}$ is given by $d^1, \ldots, d^n, a^n$, where a^n is corresponds to the largest scale of the decomposition.

The HP and LP filters are determined by the choice of two N-dimensional vectors called *mother* and *father wavelets*, respectively. Although some properties are required on the mother and father wavelets, these are not presented here.

Remark. The wavelet decomposition is only loosely presented above, as a rigorous explanation would require introducing an extensive background, which is beyond the scope of this work. The interested reader is referred to Frazier's book on wavelets, where their theory is explored in a well-structured and detailed way [22].

An important detail to note about this method is that different choices of wavelet lead to different wavelet decompositions. Thus, it is necessary to choose a wavelet whose properties are considered suitable for the time series. It is possible to find many examples wavelets which have been used for trend extraction in the literature, such

as the 16th order symlet [23], the 4th order Daubechies wavelet [7, 13], the Morlet wavelet [19], and the 4th order coiflet [1]. Nevertheless, the most important parameter to be selected is the number of levels in the decomposition, as it determines the time scale of the estimated trend. The greater the number of levels, the greater the time scale of the estimated trend.

3.5 Empirical mode decomposition

Empirical mode decomposition (EMD) is another decomposition method that can be used to extract the trend from a time series $\{x_t\}$. As in wavelet decomposition, $\{x_t\}$ is decomposed into several additive components, each one associated with a different time scale. The trend is estimated by adding the largest-scale components [11].

The EMD is obtained by successively extracting the current smallest-scale component until no more such components can be extracted. The method is based on the following heuristic. For every pair of consecutive extrema of $\{x_t\}$ occurring at times t_- and t_+ (e.g. two minima), one can consider $\{x_t\}$ to locally be the sum of a high-frequency part $\{d_t\}_{t=t_-}^{t_+}$ and a low-frequency part $\{m_t\}_{t=t_-}^{t_+}$

$$x_t = m_t + d_t, \qquad t_- \leq t \leq t_+,$$

with $\{d_t\}$ having an extremum at t_- and t_+, and another one in between [24]. Moreover, one can assume that the high-frequency part is centered around 0. With this idea in mind, EMD extracts the smallest-scale component $\{d_t\}$ with the following algorithm:

Algorithm 3.2.

1. The upper $\{u_t\}$ and lower $\{l_t\}$ envelopes of $\{x_t\}$ are found by interpolating between the maxima and the minima, respectively.
2. The global low-frequency part $\{m_t\}$ is defined as the average of the upper and lower envelopes: $m_t = (u_t + l_t)/2$.
3. The smallest-scale component is extracted: $d_t = x_t - m_t$.
4. The *residual* is $r_t = x_t - d_t$

In order to continue extracting the next smallest-scale components, this algorithm is repeated until a certain stopping criterion is met, by taking the previous $\{r_t\}$ instead

of $\{x_t\}$ at each iteration. After n iterations, the decomposition is given by

$$x_t = \sum_{i=1}^{n} d_t^i + r_t^n, \tag{3.4}$$

where $\left\{d_t^i\right\}$ is the smallest-scale component obtained at the i-th iteration, and $\left\{r_t^n\right\}$ is the residual of the n-th iteration.

In practice, the smallest-scale component $\{d_t\}$ obtained in Algorithm 3.2 requires further refinement. This refinement is achieved with the *sifting procedure*, which consists on repeating steps 1 to 3 by taking the $\{d_t\}$ of the previous repetition instead of $\{x_t\}$, until a certain criterion is met. After doing so, Equation (3.4) still holds. The objective of the sifting procedure is that the smallest-scale component satisfies the definition of intrinsic mode function (IMF), which is given below. For this reason, each of these components are referred to as IMFs from now on.

Definition 3.3 (Huang et al., 1998)**.** An *intrinsic mode function (IMF)* is a function that satisfies two conditions:

1. in the whole data set, the number of extrema and the number of zero crossings must either equal or differ at most by one, and
2. at any point, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero [9].

Definition 3.3 is motivated by the definition of instantaneous frequency proposed by Huang et al. [9], which requires that the IMFs are symmetric with respect to their local mean and that they have the same amount of zero crossings and extrema [25]. Nevertheless, this topic is not further discussed in the present work.

There are a few details of the method which need additional attention. The first one is the definition of the upper and lower envelopes from Algorithm 3.2. Originally, Huang et al. propose to obtain them by interpolating the maxima and the minima using natural cubic splines, as shown in Figure 3.1 [9]. However, this choice is arbitrary and other options have also been explored, such as the usage of B-splines [26].

Additionally, the envelopes can get distorted at the boundaries if one simply chooses x_1 and x_N as knots for the spline interpolation regardless of whether they are actually maxima or minima. A possible solution is to extend $\{x_t\}$ on both ends using the frequency and the amplitude inferred from the first two and the last two extrema, respectively [9]. Once again, there have been other alternatives proposed, such as the one considered by Zeng and He [27].

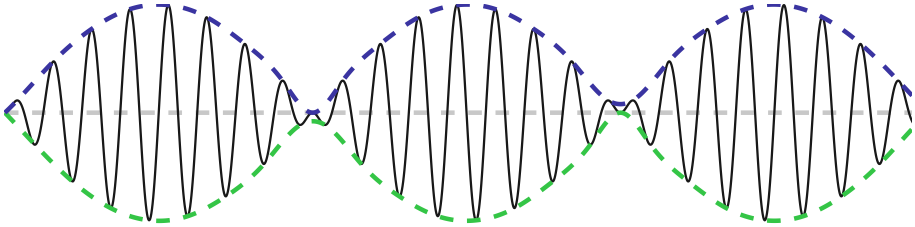

Figure 3.1. Time series (black) with its upper and lower envelopes (blue and green), obtained using natural cubic splines.

Another important detail are the stopping criteria. The overall stopping criterion follows naturally from the initial reasoning behind EMD. The procedure is stopped when the residual is monotonic or zero according to a predefined threshold [9]. For the sifting stopping criterion, Huang et al. [9] originally propose a stopping criterion based on how close the current IMF candidate $\{d_t^{i,j}\}$ is to the previous IMF candidate $\{d_t^{i,j-1}\}$. They define the coefficient

$$SD = \sum_{t=1}^{N} \left(\frac{d_t^{i,j} - d_t^{i,j-1}}{d_t^{i,j-1}} \right)^2,$$

and accept the candidate as an IMF if the coefficient is lower than a certain threshold. An alternative approach is given by Huang et al. [28], which suggest choosing an integer S and stopping the sifting procedure when the number of extrema and zero crossings of the IMF candidate $\{d_t^{i,j}\}$ is the same for S consecutive steps. Unlike the previous criterion, this one relies on the definition of IMF.

Finally, a way to determine the components from which to obtain the estimated trend is needed. Although this can be done by inspection, there exist approaches which allow to do so automatically, such as the *ratio approach*, the *energy approach* and the *energy-ratio approach*, proposed by Moghtaderi et al. [2]. The estimated trend $\{\hat{\tau}_t\}$ is given by

$$\hat{\tau}_t = \sum_{i=q}^{n} d_t^i + r_t^n,$$

for a chosen q. It should be noted that it is possible to stop the decomposition procedure once the residual already corresponds to the trend, as performing more iterations will only decompose the estimated trend. This will be used in Chapter 4.

4 Trend extraction on a simulated time series

In this chapter, the previously introduced trend extraction methods are showcased using a simulated time series. The intention of this chapter is not to provide an exhaustive analysis of each of the methods, but rather to show how they perform on a particular example and to describe some of their properties.

4.1 The simulated time series

The chosen underlying process for the example is $X_t = \tau_t + Y_t$, where τ_t is a continuously differentiable trend and $Y_t \sim \mathcal{N}\left(0, \sigma^2\right)$ for an arbitrary fixed $\sigma^2 = 24^2$, where the Y_t are assumed to be uncorrelated. The underlying process is intentionally simple, so as to not obfuscate the results unnecessarily.

The trend is predefined. This allows an easy way to compare the trend estimated by each of the methods to the actual trend. The chosen trend is defined by sampling from a known arbitrary piecewise infinitely differentiable function. The idea behind this definition is to have a function which is complex enough so that it is not trivial for any of the methods, but at the same time does not have any abrupt changes. In particular, τ_t is given by $\tau_t = f((t-1)/2)$ for $\alpha_0 \leq (t-1)/2 < \alpha_4$, where

$$f(s) = \begin{cases} f_0(s) = -\beta_0(s-\alpha_0)(s-\alpha_1) + \beta_1 \frac{\alpha_1 - s}{\alpha_1 - \alpha_0}, & \alpha_0 \leq s \leq \alpha_1 \\ f_1(s) = g_1(s) \sin h_1(s), & \alpha_1 \leq s \leq \alpha_2 \\ f_2(s) = \beta_4 g_2(s) + \beta_5 \sin h_2(s) & \alpha_2 \leq s \leq \alpha_3 \\ f_3(s) = \beta_6 + \beta_7 (s-\alpha_3)^2, & \alpha_3 \leq s < \alpha_4 \end{cases}.$$

The functions are $g_1(s) = \beta_2 + \beta_3(s-\alpha_1)$, $h_1(s) = 5\pi(s-\alpha_1)/(2(\alpha_2-\alpha_1)) + \pi$, $g_2(s) = s^2(s-\alpha_2)(s-\alpha_3) - \alpha_2^2(s-\alpha_3)^2/2 - \alpha_3^2(s-\alpha_2)^2/2$, and $h_2(s) = (s-\alpha_2)^2/(\alpha_3-\alpha_2)^2\pi + 3\pi/2$, whereas the parameters β_0, β_4 and β_6 are

$$\begin{aligned} \beta_0 &= \frac{1}{\alpha_1 - \alpha_0}\left(\frac{5\pi}{2}\frac{\beta_2}{\alpha_2 - \alpha_1} - \frac{\beta_1}{\alpha_1 - \alpha_0}\right), \\ \beta_4 &= \frac{2}{\alpha_2^2(\alpha_2 - \alpha_3)^2}\left(\beta_2 + \beta_3(\alpha_2 - \alpha_1) - \beta_5\right), \\ \beta_6 &= \beta_5 - \beta_4 \frac{\alpha_3^2}{2}(\alpha_3 - \alpha_2)^2, \end{aligned}$$

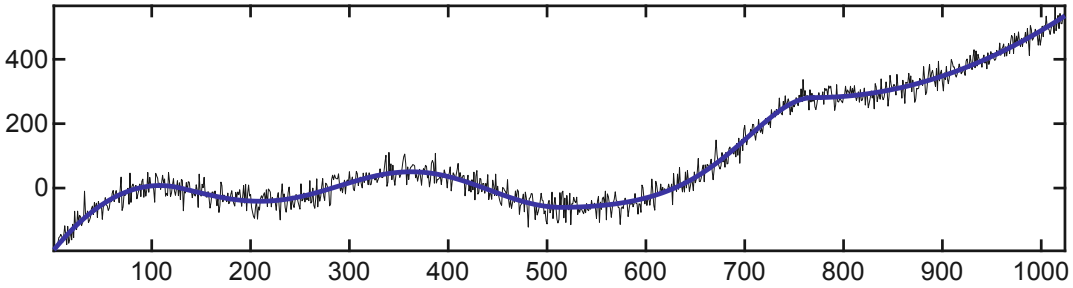

Figure 4.1. Time series generated from X_t (black) and its trend τ_t (blue).

and the values for the remaining parameters are $\alpha_0 = 0$, $\alpha_1 = 64$, $\alpha_2 = 256$, $\alpha_3 = 384$, $\alpha_4 = 512$, $\beta_1 = -192$, $\beta_2 = 36$, $\beta_3 = 1/8$, $\beta_5 = 128$ and $\beta_7 = 1/64$. The trend is shown in Figure 4.1.

4.2 Details on the methods used

Each of the previously presented methods can have several small modifications. For example, in wavelet decomposition, one can choose among many families of wavelets. In practice, however, it is only possible to limit oneself to a subset of these variations. The different options that are considered here to extract the trend from the simulated time series are presented below. For convenience, in this section the number of values of the simulated time series is denote by $N = 1024$.

4.2.1 Ordinary least squares regression

Only polynomials are considered, with degrees ranging from 0 to $N/4$.

4.2.2 Locally weighted polynomial regression

In this case, polynomials of degrees 0, 1 and 2 are considered, and odd values for k ranging from 5 to $N/2 - 1$. The chosen weight functions are the tricube and the trivial weight functions, presented in Sections 3.2 and 3.3.

4.2.3 Moving average

For the moving average method, odd values for k ranging from 1 to $N/2 - 1$ are considered. The boundaries are treated in the two ways covered in Section 3.3. Namely, by mirroring $\{x_t\}$ on the boundaries, and also by extending the trend as a constant.

4.2.4 Wavelet decomposition

The used wavelets are the 4th order Daubechies wavelet, the 4th order coiflet, and the 16th order symlet, as chosen by Homborg et al. [13], Mushini et al. [19] and Alexandrov et al. [1], respectively. The decomposition is considered up to several levels: from 1 to 8. The boundaries are again treated in two different ways: by constant extension, that is, defining $x_t = x_1$ for $t < 1$ and $x_t = x_N$ for $t > N$, and by mirroring.

4.2.5 Empirical mode decomposition

The chosen sifting procedure is stopping after S consecutive IMF candidates have equal number of extrema and zero-crossings. Values for S ranging from 1 to 10 are considered, which includes the values mentioned by Huang et al. [29] as being typically successful as the default value ($3 \lesssim S \lesssim 5$). Since there is no guarantee that the sifting procedure will eventually stop [29], an arbitrary maximum of 200 sifting iterations is allowed. The envelopes are here computed using natural cubic splines after extending the maxima and the minima on both boundaries. Finally, different values for the maximum number of IMFs are considered: from 1 to 8. The residual is taken as the trend, as discussed in Section 3.5.

4.3 General comparison of the methods

4.3.1 A brute-force approach

In this subsection, the best potential performance of each of the methods is compared. This is done by generating 100 different time series and then running all the methods with each of the variations considered in Section 4.2. For each of the generated time series and method, the variation which minimizes the *root mean squared error* between the actual trend $\{\tau_t\}$ and the estimated trend $\{\hat{\tau}_t\}$ is chosen as the optimal. The root mean squared error is defined as

$$\mathrm{RMSE}(\hat{\tau}) = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(\tau_t - \hat{\tau}_t)^2}.$$

The fact that only 100 different time series are used is mainly justified by the computational cost of performing this brute-force optimization. The results of the optimization, presented in Figure 4.3 and Table 4.1, show that OLS and LWP are

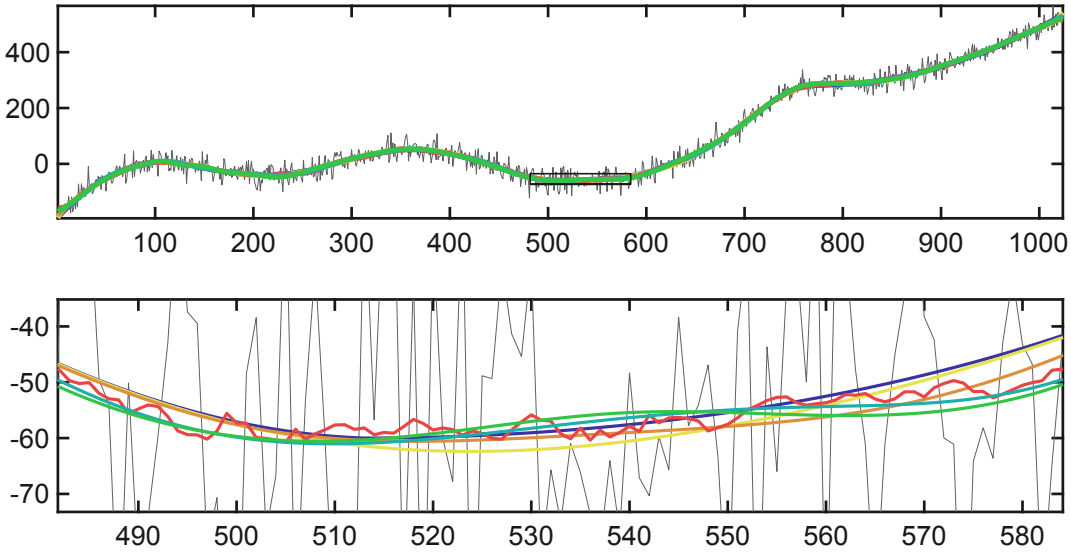

Figure 4.2. In gray, time series generated from X_t. In blue, its trend τ_t. The remaining trends are the result of the brute-force optimization on this time series for each method: OLS (yellow), LWP (orange), moving average (red), wavelet decomposition (cyan), EMD (green).

consistently the two best options, with LWP being the best one. On the other hand, EMD performs slightly better than wavelet decomposition, which in turn performs better than moving average. Figure 4.2 shows the result of the optimization on a single time series. Overall, all the methods perform reasonably well when its parameters are optimally chosen. Statistics of the optimal parameters are shown in Table 4.2.

	Error			**Rank**		
	MED.	MEAN	STD. DEV.	MED.	MEAN	STD. DEV.
OLS regression	3.64	3.68	0.47	2.0	2.09	0.53
LWP regression	3.27	3.23	0.52	1.0	1.11	0.31
Moving average	4.63	4.60	0.47	5.0	4.46	0.74
Wavelet decomp.	4.52	4.47	0.54	4.0	4.01	0.76
EMD	4.19	4.20	0.62	3.0	3.33	0.92

Table 4.1. Per each method, statistics of the optimal error and the rank from all the time series considered. The rank of a method for a given time series is defined based on its optimal error compared to the other methods: from 1 (best method) to 5 (worst method).

4.3.2 Error with respect to a parameter of interest

Although interesting as an initial approach, when thinking about real-world data, the previous comparison is only partially useful. Using a brute-force optimization, it has been possible to detect what is the best error that can be obtained for each method out

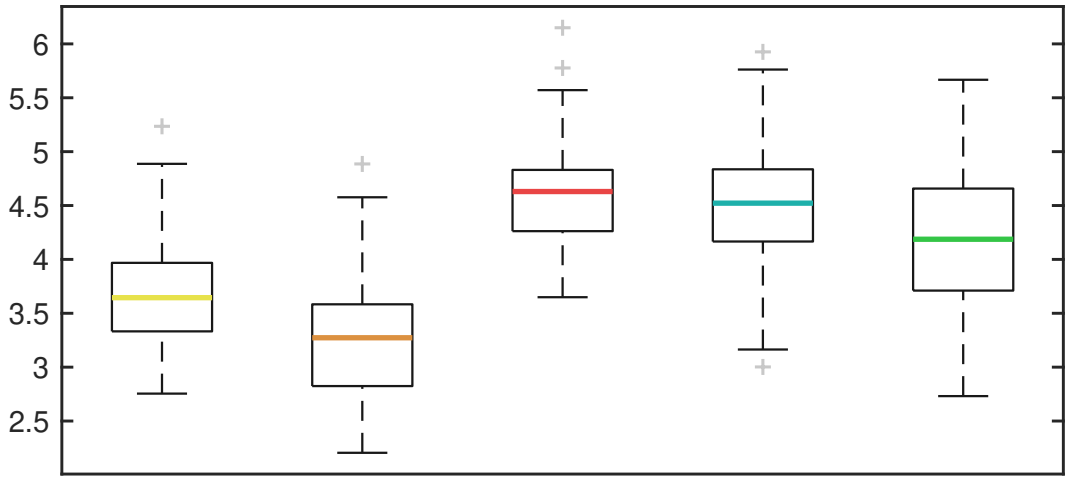

Figure 4.3. Per each of the methods, boxplot of the optimal errors from all the time series considered. Left to right: OLS (yellow), LWP (orange), moving average (red), wavelet decomposition (cyan), EMD (green).

Ordinary least squares regression					
	VALUE	FREQ.	MEDIAN	MEAN	STD. DEV.
POLYNOMIAL DEGREE			21.5	21.79	2.18
Locally weighted polynomial regression					
	VALUE	FREQ.	MEDIAN	MEAN	STD. DEV.
POLYNOMIAL DEGREE			2.0	2.00	0.00
k			180.0	179.60	16.54
WEIGHT FUNCTION	Tricube	94			
	Trivial	6			

Table 4.2. Statistics of the brute-force optimization. *(Continued on next page)*

of all the variations considered. However, in most cases one cannot know which is the optimal variation of each of the methods. Moreover, it could be possible that a particular variation of a method gives the best result compared to any method, but any other variation of the method performs very poorly. With this in mind, it is perhaps also relevant to see how the root mean squared error changes along a parameter of interest, for example the amount of levels in the decomposition in the wavelet decomposition method.

For each of the methods, a parameter among those which determine the variations is chosen, while the other parameters are fixed. The choice of fixed parameters is based on Table 4.2. For OLS there is only one parameter, which is the degree of the polynomial, so there is no alternative choice. The parameter of interest chosen for the LWP is k, whereas the weight function is fixed to the tricube function and the

Moving average					
	VALUE	FREQ.	MEDIAN	MEAN	STD. DEV.
k			35.0	35.86	2.69
BOUNDARY	Constant	0			
	Mirror	100			

Wavelet decomposition					
	VALUE	FREQ.	MEDIAN	MEAN	STD. DEV.
WAVELET	Daub. 4th	19			
	Coif. 4th	28			
	Sym. 16th	53			
LEVELS			5.0	5.16	0.37
BOUNDARY	Constant	72			
	Mirror	28			

Empirical mode decomposition					
	VALUE	FREQ.	MEDIAN	MEAN	STD. DEV.
S			7.0	6.70	2.80
MAX. IMFS			6.0	6.25	0.58

Table 4.2. *(Continued)* Statistics of the brute-force optimization.

degree of the polynomials to 2. For moving average, the chosen parameter is again k, and the boundary treatment is fixed to mirroring the original time series. For the wavelet decomposition method, the chosen parameter of interest is the number of levels of the decomposition. The wavelet is fixed to the 16th order symlet and the boundary is treated by mirroring. Finally, in EMD, the value S is fixed to 7 and the chosen parameter of interest is the maximum amount of IMFs.

The root mean squared errors of the corresponding estimations of the trend are shown in Figure 4.4 using a continuous box plot. It is clear from the graphs that in this case it is possible to approximately choose an optimal default value for each of the parameters of interest. After noticing this, the next natural question is what happens when one deviates from this value. When using moving average, deviating slightly from the optimal value for k can lead to very poor results. This property, which is evident in Figure 4.4, is often mentioned as the main disadvantage of moving average. In comparison, doing the same when extracting the trend with a LWP has a much less drastic effect, making the latter method preferable in this sense. For the OLS method, the effect is less pronounced than in moving average, although one still needs to be careful when selecting the polynomial degree. Choosing a degree which is too high

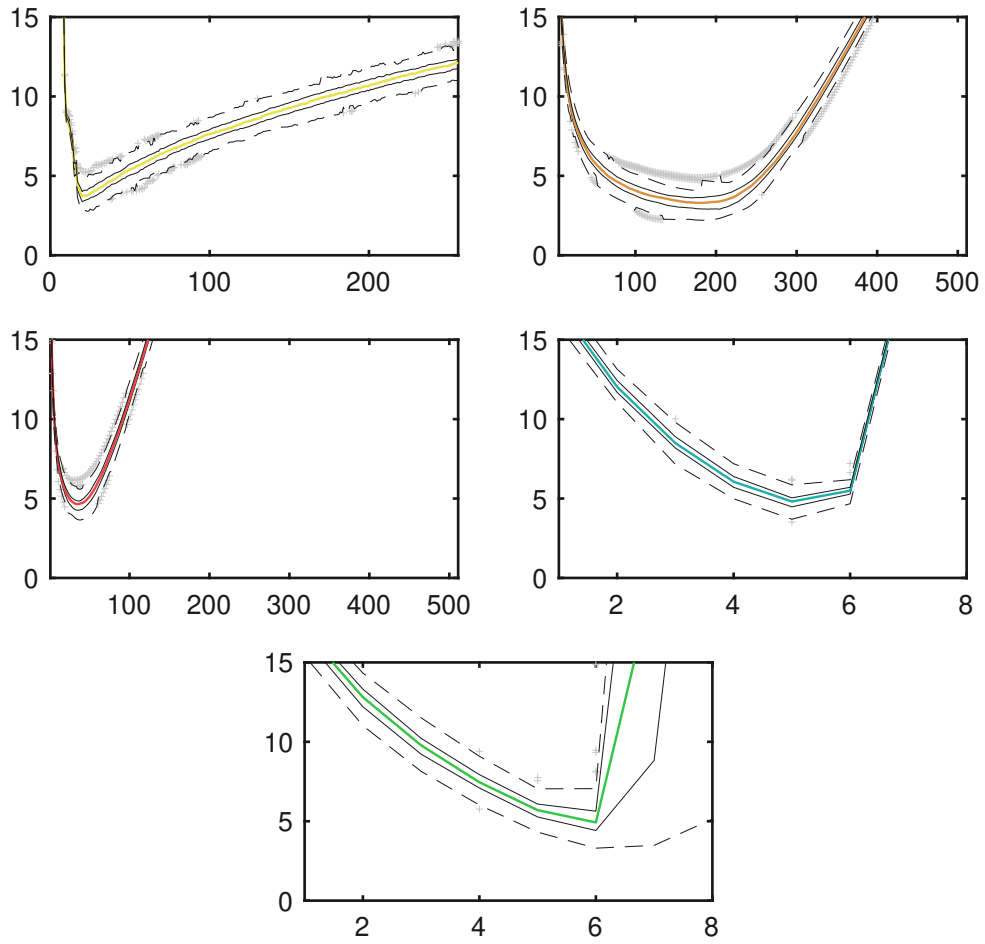

Figure 4.4. For each method, continuous boxplot displaying the error of the method with respect to the chosen parameter of interest. The colored line is the median, which is bounded by the lower and upper quartiles (solid black). The whiskers are the dashed black lines and outliers are displayed in gray. Left to right and top to bottom: OLS (yellow), LWP (orange), moving average (red), wavelet decomposition (cyan), EMD (green).

might lead to undesired oscillations and overfitting, while doing the opposite might lead to an estimated trend that does not properly capture the details of the actual trend.

In this case, EMD and wavelet decomposition offer an advantage compared to the previous methods, which is that the amount of reasonable values for the parameter is much lower. In practice, this means that whenever possible, one might be able to determine the optimal value for the parameter of interest by inspection. Additionally, for these two methods there are ways to estimate the best value for the maximum number of IMFs and levels, respectively. The ratio, energy and energy-ratio approaches proposed by Moghtaderi et al. [2] can be used for empirical mode decomposition, whereas statistical parameters can be used in the wavelet decomposition method to assess which approximation coefficients correspond to the trend [7].

4.4 Additional properties of the methods

In this section some properties which are usually desirable in the problem of trend extraction are mentioned and commented for each of the methods presented. A good property for a trend extraction method is *locality*, namely that the value of the trend at a given time is only affected by the values occurring at nearby times in the original time series. An example of situation in which this can be crucial is when considering a time series which is regularly updated with new values as time passes. For instance, in a time series where values are spaced in time by an hour, it might not make sense that a recently added value affects the value of the trend one month ago.

Moving average and LWP are clearly local trend extraction methods, whose locality can be easily adjusted by varying the value of k. On the other hand, the locality of OLS depends on the choice of functions. Polynomial regressions, which are the most commonly chosen option, are nonlocal. The wavelet decomposition method is also local provided that the right wavelet family is chosen. EMD might seem local at first sight, because the trend is obtained using natural cubic splines, which are local. Nonetheless, the common sifting stopping criteria considered in Section 3.5 are both global, making the method also global. A way to tackle this issue is to only sift the regions of the IMF candidate which need sifting, as proposed by Rilling et al. [24]. This property is illustrated in Figure 4.5, where a comparison between the trend extracted from the whole time series and only from a subsequence of it is shown.

The *simplicity* of a method might also be desired, perhaps as a first approach to a problem which might require trend extraction. In this case moving average and OLS most likely become the best options, followed closely by LWP. Wavelet decomposition and EMD might be seem as more complex, although the fact that there already exist ready-to-use implementations make these methods also reasonably accessible.

When attempting to understand the effect of a trend extraction method on a given time series, a method supported by a *rigorous mathematical framework* becomes of special interest. Unlike the other methods presented here, EMD suffers greatly from a lack of such theory, as it is essentially defined by an algorithm which relies on heuristics [5, 30]. Despite that, EMD has an advantage over the other methods, which is the fact that it is *adaptative*. The time-scale of the trend is clearly determined by k in moving average, and also by the weight function in LWP. For OLS, the scale is fixed by the chosen basis functions, while wavelet decomposition fixes the scale when the number of levels in the decomposition is chosen. These parameters, which fix the

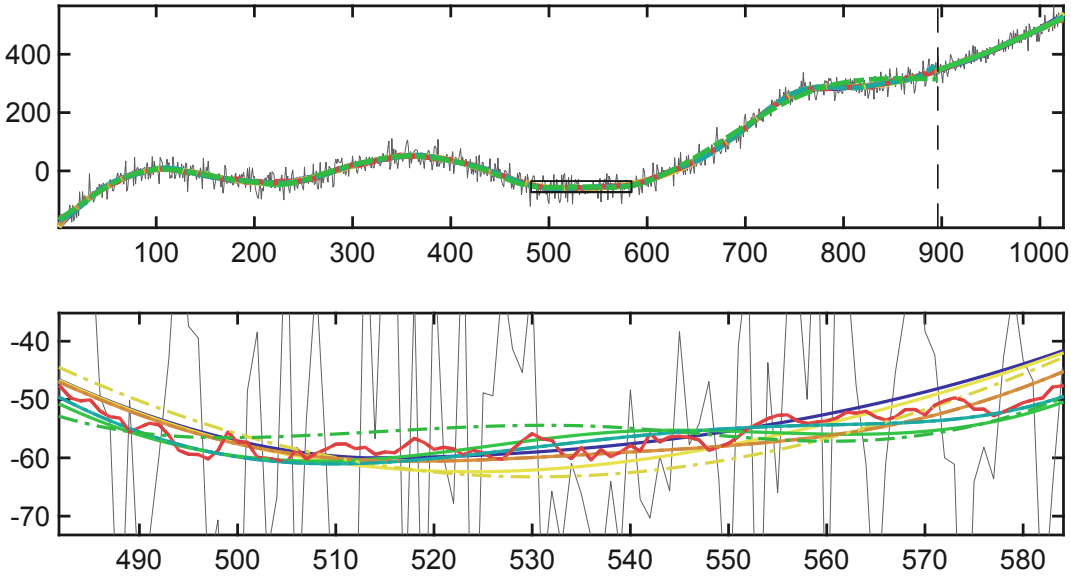

Figure 4.5. Same content as in Figure 4.2, but adding also the optimal estimated trends obtained from considering the subsequence starting at $t = 1$ and finishing at $t = 896$ (colored dash-dotted lines).

time-scale of the trend, are oftentimes predefined, although they can have a major effect on the estimated trend, as seen in Subsection 4.3.2. This is not the case in EMD, as each of the components is not necessarily associated with a single time-scale. In fact, there can be several time-scales in a component.

In some cases, it might be useful to choose a method for which the resulting trend can be easily summarized by a few values. Using OLS is then reasonable, as the trend can be described merely using the coefficients of the predefined basis functions. The same happens with the wavelet decomposition method, where the approximation coefficients characterize the trend once the wavelet function and the way to treat the boundaries has been predefined. However, for the remaining methods there is not an immediate way to summarize the trend.

A common assumption for the trend is that it is a *smooth enough* function. When this assumption is made, it makes sense to choose a method which yields a smooth estimated trend. This is guaranteed in EMD, as the trend is a finite linear combination of natural cubic splines, which are two times differentiable with continuity. When using OLS, this is also true provided that the basis function are also smooth. As can be seen in Figure 4.2, the same cannot be said for moving average, which offers no guarantee of smoothness. Although this is also the case for LWP, in practice it seems to produce smooth results. The wavelet decomposition method also seems to yield smooth trends, as shown in Figure 4.2.

5 Conclusions

Five trend extraction methods have been presented and compared using an arbitrary simulated time series. The main conclusion of this work is that there is no overall best trend extraction method among the ones presented. This is partly due to the fact that the trend extraction problem is itself ill-defined in the sense that it lacks a rigorous formal definition. The definition of trend is oftentimes strongly linked to the model chosen to explain a particular time series. When in need to extract the trend from a time series, one should consider the desired properties of the method in order to choose the most appropriate one for that particular case.

The results obtained are in line with the literature that has been reviewed, although there seems to be a positive bias towards adaptative methods such as those belonging to the empirical mode decomposition family. It has also been noticed that it is relatively common in the reviewed literature to ignore and not sufficiently specify the details of the empirical mode decomposition algorithm used, sometimes relying on third-party implementations. This practice hinders the reproducibility of the results and should be avoided whenever possible. Additionally, it also appears to be common to base the method comparisons on a very small set of examples, sometimes rather simple. While doing so can be useful in order to have an idea of the behavior of each method considered, one can be easily tempted to immediately extrapolate such results to any time series, which is something that should be done carefully.

The present work uses one example to showcase the results of applying each method to a time series, which can be seen as insufficient. Another reasonable criticism is the lack of a more in-depth explanation of some parts which might be considered nontrivial, such as the wavelet decomposition. This is justified by format limitations.

The possibilities to extend the present work are numerous. One option could be to analyze other trend extraction methods, presenting also methods which have been invented in recent years. Following this line of work, it could also be interesting to consider other variations of the methods presented, such as different ways to treat the boundaries, which might lead to better results. Another possibility would be to propose a more extensive comparative study by considering several examples, in hopes of being able to produce results which can be extrapolated to more general cases. Similarly, one might also choose to limit the comparison to a very specific situation where the need to extract a trend appears.

Bibliography

[1] T. Alexandrov et al. “A Review of Some Modern Approaches to the Problem of Trend Extraction”. In: *Econometric Reviews* 31.6 (2012), pp. 593–624. DOI: 10.1080/07474938.2011.608032.

[2] A. Moghtaderi et al. “Trend filtering via empirical mode decompositions”. In: *Computational Statistics & Data Analysis* 58 (2013), pp. 114–126. DOI: 10.1016/j.csda.2011.05.015.

[3] Z. Wu and N. E. Huang. “Ensemble empirical mode decomposition: a noise-assisted data analysis method”. In: *Advances in adaptive data analysis* 1.01 (2009), pp. 1–41. DOI: 10.1142/S1793536909000047.

[4] X. Zhang and Z. Chen. “A new method to remove the tree growth trend based on ensemble empirical mode decomposition”. In: *Trees* 31.2 (2017), pp. 405–413. DOI: 10.1007/s00468-015-1295-z.

[5] I. Daubechies et al. “Synchrosqueezed wavelet transforms: An empirical mode decomposition-like tool”. In: *Applied and Computational Harmonic Analysis* 30.2 (2011), pp. 243–261. DOI: 10.1016/j.acha.2010.08.002.

[6] K. Dragomiretskiy and D. Zosso. “Variational Mode Decomposition”. In: *IEEE Transactions on Signal Processing* 62.3 (2014), pp. 531–544. DOI: 10.1109/TSP.2013.2288675.

[7] L. Lentka and J. Smulko. “Methods of trend removal in electrochemical noise data – Overview”. In: *Measurement* 131 (2019), pp. 569–581. DOI: 10.1016/j.measurement.2018.08.023.

[8] P. J. Brockwell and R. A. Davis. “Introduction”. In: *Introduction to Time Series and Forecasting*. Springer International Publishing, 2016, pp. 1–37. DOI: 10.1007/978-3-319-29854-2_1.

[9] N. E. Huang et al. “The Empirical Mode Decomposition and the Hilbert Spectrum for Nonlinear and Non-Stationary Time Series Analysis”. In: *Proceedings: Mathematical, Physical and Engineering Sciences* 454.1971 (1998), pp. 903–995. DOI: 10.1098/rspa.1998.0193.

[10] D. C. Montgomery et al. *Introduction to time series analysis and forecasting*. John Wiley & Sons, 2015.

[11] Z. Wu et al. “On the trend, detrending, and variability of nonlinear and nonstationary time series”. In: *Proceedings of the National Academy of Sciences* 104.38 (2007), pp. 14889–14894. DOI: 10.1073/pnas.0701020104.

[12] Z. Zhou et al. “Techniques to improve the accuracy of noise power spectrum measurements in digital x-ray imaging based on background trends removal”. In: *Medical Physics* 38.3 (2011), pp. 1600–1610. DOI: 10.1118/1.3556566.

[13] A. M. Homborg et al. “Time-frequency methods for trend removal in electrochemical noise data”. In: *Electrochimica Acta* 70 (2012), pp. 199–209. DOI: 10.1016/j.electacta.2012.03.062.

[14] N. Bigdely-Shamlo et al. “The PREP pipeline: standardized preprocessing for large-scale EEG analysis”. In: *Frontiers in Neuroinformatics* 9 (2015), p. 16. DOI: 10.3389/fninf.2015.00016.

[15] Z. He et al. “An incipient fault detection approach via detrending and denoising”. In: *Control Engineering Practice* 74 (2018), pp. 1–12. DOI: 10.1016/j.conengprac.2018.02.005.

[16] A. de Cheveigné and D. Arzounian. “Robust detrending, rereferencing, outlier detection, and inpainting for multichannel data”. In: *NeuroImage* 172 (2018), pp. 903–912. DOI: 10.1016/j.neuroimage.2018.01.035.

[17] W. S. Cleveland. “Robust Locally Weighted Regression and Smoothing Scatterplots”. In: *Journal of the American Statistical Association* 74.368 (1979), pp. 829–836. DOI: 10.1080/01621459.1979.10481038.

[18] J. Lu et al. “Detrending crop yield data for spatial visualization of drought impacts in the United States, 1895-2014”. In: *Agricultural and Forest Meteorology* 237-238 (2017), pp. 196–208. DOI: 10.1016/j.agrformet.2017.02.001.

[19] S. C. Mushini et al. “Improved amplitude- and phase-scintillation indices derived from wavelet detrended high-latitude GPS data”. In: *GPS Solutions* 16.3 (2012), pp. 363–373. DOI: 10.1007/s10291-011-0238-4.

[20] F. Zhang et al. “Bioelectric signal detrending using smoothness prior approach”. In: *Medical Engineering & Physics* 36.8 (2014), pp. 1007–1013. DOI: 10.1016/j.medengphy.2014.05.009.

[21] R. Kopel et al. “No time for drifting: Comparing performance and applicability of signal detrending algorithms for real-time fMRI”. In: *NeuroImage* 191 (2019), pp. 421–429. DOI: 10.1016/j.neuroimage.2019.02.058.

[22] M. W. Frazier. “Wavelets on $\mathbb{Z}_N$”. In: *An Introduction to Wavelets Through Linear Algebra*. New York, NY: Springer New York, 1999, pp. 165–263. DOI: 10.1007/0-387-22653-2_4.

[23] A. González-López and P.-A. Campos-Morcillo. “Efficient detrending of uniform images for accurate determination of the noise power spectrum at low frequencies”. In: *Physics in Medicine & Biology* 64.10 (2019), p. 105001. DOI: 10.1088/1361-6560/ab1a68.

[24] G. Rilling et al. “On empirical mode decomposition and its algorithms”. In: *IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing (NSIP)*. Vol. 3. 2003, pp. 8–11.

[25] N. E. Huang et al. “On instantaneous frequency”. In: *Advances in adaptive data analysis* 1.02 (2009), pp. 177–229. DOI: 10.1142/S1793536909000096.

[26] Q. Chen et al. “A B-spline approach for empirical mode decompositions”. In: *Advances in Computational Mathematics* 24.1 (2006), pp. 171–195. DOI: 10.1007/s10444-004-7614-3.

[27] K. Zeng and M.-X. He. “A simple boundary process technique for empirical mode decomposition”. In: *IGARSS 2004. 2004 IEEE International Geoscience and Remote Sensing Symposium*. Vol. 6. 2004, pp. 4258–4261. DOI: 10.1109/IGARSS.2004.1370076.

[28] N. E. Huang et al. “A new view of nonlinear waves: the Hilbert spectrum”. In: *Annual Review of Fluid Mechanics* 31.1 (1999), pp. 417–457. DOI: 10.1146/annurev.fluid.31.1.417.

[29] N. E. Huang et al. “A confidence limit for the empirical mode decomposition and Hilbert spectral analysis”. In: *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences* 459.2037 (2003), pp. 2317–2345. DOI: 10.1098/rspa.2003.1123.

[30] J. Gilles. “Empirical Wavelet Transform”. In: *IEEE Transactions on Signal Processing* 61.16 (2013), pp. 3999–4010. DOI: 10.1109/TSP.2013.2265222.