CrossMark

Sensitivity of the Empirical Mode Decomposition to Interpolation Methodology and Data Non-stationarity

F. M. Z. Bahri[1] · J. J. Sharples[1]

Abstract

Empirical mode decomposition (EMD) is a commonly used method in environmental science to study environmental variability in specific time period. Empirical mode decomposition is a sifting process that aims to decompose non-stationary and non-linear data into their embedded modes based on the local extrema. The local extrema are connected by interpolation. The results of EMD strongly impact the environmental assessment and decision making. In this paper, the sensitivity of EMD to different interpolation methods, linear, cubic, and smoothing-spline, is examined. A range of non-stationary data, including linear, quadratic, Gaussian, and logarithmic trends as well as noise, is used to investigate the method's sensitivity to different types of non-stationarity. The EMD method is found to be sensitive to the type of non-stationarity of the input data, and to the interpolation method in recovering low-frequency signals. Smoothing-spline interpolation gave overall the best. The accuracy of the method is also limited by the type of non-stationarity: if the data have an abrupt change in amplitude or a large change in the variance, the EMD method cannot sift correctly.

Keywords Empirical mode decomposition · Non-linear and non-stationary data · Time series analysis · Sensitivity analysis · Interpolation method

1 Introduction

Data non-stationarity and non-linearity are two main sources of difficulty in data analysis. Time series data are non-stationary if the mean, variance, or both vary with time [18, 25]. Empirical mode decomposition (EMD) was pioneered by Huang in [13] as a fundamental component of the Hilbert-Huang transform (HHT), which was designed to facilitate spectral analysis of non-stationary and non-linear data. EMD is a sifting process that decomposes the data into component functions, called intrinsic mode functions (IMFs), which are analogous to the harmonic modes of Fourier analysis, but which can have variable amplitude and frequency throughout the time domain. A critical step in determining each of the IMFs involves constructing upper and lower envelopes of the local maxima and minima of the original series, or a residual series, depending on the stage of the sifting process. These envelopes are usually obtained using some form of interpolation, most commonly a cubic interpolating spline. Transformation of the IMFs using the Hilbert transform (HT) then delivers spectral information relating to the data [13].

The combination of EMD and the HT offers a practical alternative to more traditional time series decomposition and spectral analysis methods such as the Fourier transform and wavelet analysis, which are not generally well-suited to non-stationary and non-linear time series [5, 6, 19, 26]. Indeed, since its inception, EMD has been applied in a number of areas where non-stationary and non-linear data naturally arise. Examples include biomedicine [14, 17]; neuroscience [24]; chemical engineering [23]; finance [15]; atmospheric science [9, 27]; seismology [34]; and ocean dynamics [8, 28]. Given the broad application of EMD, it is important to have a sound understanding of its limitations and sensitivities so that the output it produces can be interpreted appropriately. A number of such limitations and sensitivities have been discussed in the literature; these include mode mixing [4, 12, 30], end effects [3, 32, 35], and sensitivity to interpolation methodology [20, 22].

✉ F. M. Z. Bahri
   f.bahri@adfa.edu.au

   J. J. Sharples
   J.sharples@adfa.edu.au

[1] School of Science, UNSW Canberra, Canberra, Australia

Mode mixing occurs when a single IMF consists of signals of widely disparate time scales, or when similar time scale signals reside in different IMF components [4, 12]. Mode mixing is known to occur when applying EMD to time series that contain noise or some other forms of signal intermittency [12]. To overcome the mode mixing problem, Wu [32] introduced the ensemble EMD (EEMD) procedure, which is a version of EMD based on multiple application of the EMD method to noise-adjusted data. Based on an analysis of tropospheric temperature data, Huang [12] demonstrated that EEMD resulted in less mode mixing than the straight EMD method. Torres [30] also addressed the mode mixing issue by introducing the complete ensemble EMD (CEEMD), in which noise is added at each stage of the sifting process. Based on an analysis of synthetic data and real electrocardiogram data, the CEEMD method was found to produce a smaller number of IMFs than the EEMD method. The high-frequency IMFs in both methods (CEEMD and EEMD) were similar, but the CEEMD method produced lower-frequency IMFs that were more symmetric than those obtained from the EEMD method [30].

As mentioned above, obtaining the IMFs typically involves the application of an interpolation method to the local extrema of a series. However, to support interpolation at the beginning and end of a time series, it is necessary to extend the series by adding points before the beginning and after the end of the time series. The way these additional points are defined will of course affect the final IMFs that are produced, and in the worst cases can produce highly undesirable end effects; for example, due to over- or under-shooting at the end points [3]. Moreover, these end effects can propagate inwards and corrupt the whole sifting process, resulting in inaccurate IMFs. A number of methods have been proposed to resolve the issue of end effects. These include the mirror-extending method [35], in which the data are reflected about the end points, and the slope-based method (SBM), in which new minima or maxima are appended to the beginning and end of the series based on the local slope of the series in their vicinity [3]. The SBM was subsequently modified by Wu [32] who noted that it can sometimes result in large oscillations near the ends. End effects have also been addressed by employing alternative interpolation techniques such as rational splines with tension [20, 22].

In the original presentation of the EMD methods, cubic-spline interpolation was employed to construct the upper and lower envelopes [13]. However, no a priori reason was given to justify the use of cubic-spline interpolation over the multitude of other interpolation techniques that are available. In fact, known issues with cubic-spline interpolation, such as overshooting and undershooting, can limit the accuracy of the overall EMD process [22]. A number of subsequent studies have investigated the use of other interpolation methods in the EMD process, and have shown that the choice of interpolation method used to obtain the envelopes of the maxima and minima can be quite critical [2, 7, 20].

Chen [2] introduced an alternative EMD method based on fitting B-splines to a weighted moving average of the extrema in the original series or subsequent residuals. This method removed the need for separate consideration of the upper and lower envelopes. They [2] used cubic and quadratic B-splines to compare their method with the classical EMD method (based on separate cubic interpolation of the minima and maxima), and found that the use of cubic B-splines gave a finer decomposition than the classical EMD method and performed better in terms of energy conservation. However, there is still no a priori reason justifying the use of cubic B-splines in this alternative formulation of EMD. Indeed, Chen note in [2] that the selection of an optimal B-spline order presents a new challenge.

Meignen [16] also proposed a method that removes the need for separate consideration of upper and lower envelopes. In their method, they define a mean envelope of the extreme points as the solution of a constrained quadratic optimization problem. This approach also circumvents one of the contentious issues with the classical EMD method, namely that of defining appropriate stoppage criteria, which is required to terminate the sifting process. Meignen [16] used synthetic time series to demonstrate that their mean-envelope method produced results very similar to classical EMD, although they only considered stationary time series in their comparisons.

A number of authors have also considered using alternative interpolation methods to define the upper and lower envelopes in the EMD process. Pegram [22] and Peel [20] advocated the use of rational splines with tension, as presented by Spath [29], in place of the cubic splines in the classical EMD method. By varying a pole parameter, which defines the "tautness" of the spline curve, rational splines can represent a spectrum of interpolation methods that encompasses quadratic splines, cubic splines, and piecewise-linear interpolation. The ability to vary the pole parameter permits consideration of the interplay between spline tension and the resulting IMFs. The performance of rational-spline EEMD was tested using synthetic time series by Peel [21], who found that it could reproduce the known structures well. Rational-spline EEMD was also found to perform well when applied to environmental time series [20–22], although there remains some ambiguity surrounding the optimal choice of the tension parameter.

Bahri [1] examined the performance of EMD and EEMD based on linear interpolation (i.e., the simplest form of interpolation) as applied to synthetic times series and to a

real geophysical time series. For the synthetic time series, Bahri [1] found that, on average, EMD and EEMD based on linear interpolation were able to reproduce the known signal components very well, with IMFs very similar to those obtained from the classical methods. Overall, the linear-interpolation-based methods were able to better capture the trend in the data, while the classical methods provided better estimates of both high- and low-frequency oscillatory components. Linear interpolation, however, did provide slightly more accurate results near the ends of the data series. For the geophysical time series (sea level data), the linear-interpolation-based EEMD was found to produce a different estimate of the long-term trend in the data than was obtained from classical EEMD, which raised a number of questions about the interpretation of EEMD output arising from actual environmental data.

In addition to the methodological limitations and sensitivities of EMD discussed above, there are also questions about how the nature of the non-linearity and non-stationarity of the time series might effect the efficacy of EMD. Environmental time series, like those that EMD has been applied to, can possess a variety of non-stationary and non-linear characteristics, and so it is important to understand the sensitivity of EMD to these aspects of the data. For example, Duffy [7] found that EMD performs well for periodic signals but can perform poorly for non-linear aperiodic time series, such as most meteorological data sets. The sensitivity of EMD to data non-stationarity has received only little attention in the literature. Peel [21] investigated the application of rational-spline EEMD to synthetic time series with a number of trend patterns, while Huang applied EMD to speech signals with time-dependent variance [10]. Knowing sensitivity of the EMD method to interpolation used and the nature of the input data is essential to obtain a better assessment of the environmental data.

In this study, we use synthetic time series to systematically investigate the combined sensitivity of EMD to data non-stationarity and interpolation method. In particular, we consider EMD methods based on linear interpolating splines, cubic interpolating splines, and smoothing splines, and assess their ability to reproduce the known signal components as IMFs.

We begin by providing a brief overview of the EMD algorithm and its extension to ensemble EMD. In particular, we highlight the role interpolation plays in the procedure and introduce the various interpolation methods used in the present study. We then define the various synthetic non-stationary data sets used to systematically analyze the performance of EMD based on the different interpolation methods. The performance of the various EMD and EEMD methods is then assessed based on their ability to accurately reproduce the known component signals.

2 Methods and Data

This section outlines the EMD and EEMD methods introduced by Huang [13] and [11], together with extensions based on alternative interpolation methods and describes the construction and use of synthetic data sets to assess the application of the alternative interpolation methods in EMD.

2.1 EMD Method

Huang [13] shows how a time series $y(t)$ can be decomposed into a number of intrinsic mode functions (IMFs), which satisfy certain defining conditions. These conditions are (i) symmetric upper and lower envelopes (mean is zero) and (ii) the number of zero crossings and extrema are either equal or differ by 1.

The IMFs of $y(t)$ are determined through the EMD process in Algorithm 2.1. Figure 1 provides a schematic flowchart of the EMD procedure.
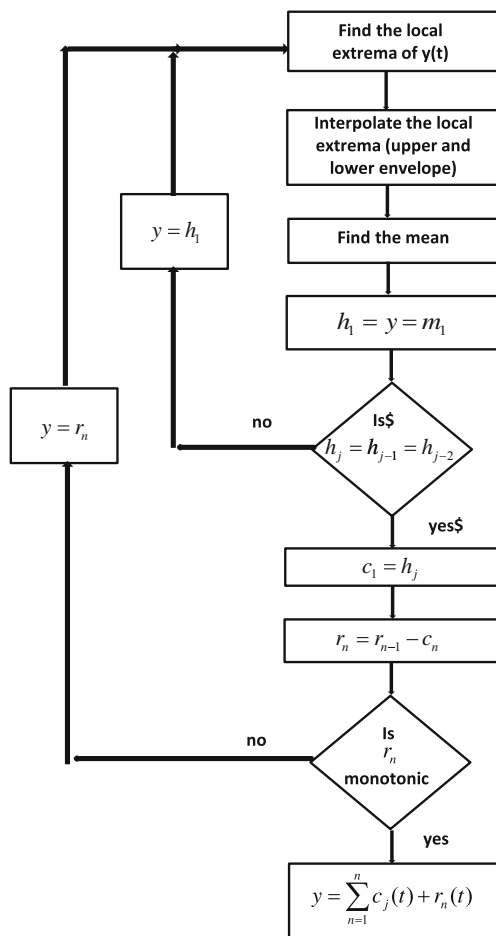

Algorithm 2.1 Empirical mode decomposition.

1. Identify all the local extrema in the signal $y(t)$.
2. Interpolate the maxima and minima separately to define upper and lower envelopes of $y(t)$.
3. Calculate the mean of the upper and lower envelopes $m_1(t)$.
4. The first protomode $h_1(t) = y(t) - m_1(t)$.
5. **If** $h_j = h_{j-1} = h_{j-2}$ **Then** Mode 1 is $c_1(t) = h_j(t)$ **Else** $y(t) = h_j(t)$ **Go to 1**.
6. Find the residual time series $r_1(t) = y(t) - c_1(t)$.
7. For the next IMF, set $y(t) = r_1(t)$ **Go to 1**.
8. Subsequent IMFs $c_3, \ldots, c_n$ and residuals $r_3, \ldots, r_n$ are obtained similarly.
    If $r_n(t)$ has at most one extremum **Stop**. EMD is now complete.

Upon completion of the EMD algorithm, the original time series has been decomposed as follows:

$$y(t) = \sum_{j=1}^{n} c_j(t) + r_n(t), \qquad (1)$$

where $c_j(t)$, $j = 1, \ldots, n$ are the IMFs and $r_n(t)$ is the residual (for a more detailed exposition of the EMD procedure, the reader is referred to Huang [12]).

As discussed above, noisy input data can result in mode mixing across the IMFs. Hence, when dealing with noisy time series, the EMD procedure is implemented as part of an ensemble process, which is referred to as ensemble EMD (EEMD). In the EEMD procedure, noise ($\epsilon \sim N(0, \delta)$) is added to the EMD; the ensemble average of the IMFs found from numerous EMD applications defines the final

Fig. 1 Empirical mode decomposition flowchart, n is the total number of modes

For a time series $y(j)$ defined at knot points $t(j)$, the smoothing spline f is defined as the minimizer (over an appropriate space of functions [33]) of the following functional:

$$p \sum_{j=1}^{n} |y(j) - f(t(j))|^2 + (1-p) \int |f''(t)|^2 \, dt. \qquad (2)$$

Here, f'' denotes the second derivative of the function f and $p \in (0, 1)$ is the smoothing parameter, which determines a balance between fidelity of the spline to the data and its smoothness. As $p \to 0$, f approaches the least squares straight-line fit to the data; as $p \to 1$, f approaches the variational or "natural" cubic-spline interpolant. Smoothing splines were determined using the csaps routine in MATLAB.

Of course, the choice of smoothing parameter p will affect the final IMFs produced by the method. In this initial work, however, the sensitivity of EMD to variation in the smoothing parameter will not be investigated in detail. While we do provide an example of the effect of varying the smoothing parameter on EMD performance, for the majority of the analyses, we choose a value of the smoothing parameter, $p = 0.0015$, that results in recovery of lower-frequency signal components to a reasonable degree of accuracy.

Despite the availability of a number of EMD codes (e.g., Huang's code,[1] Flandrin's EEMD codes[2] and the MATLAB HHT toolbox[3]), we developed our own implementation of the EEMD algorithm in MATLAB following the method espoused by Huang [12]. In particular, we have not attempted to ameliorate any of the known issues with end effects (the reader is referred to [3] and [20] for further discussion of these issues).

2.2 Synthetic Time Series

To examine the sensitivity of the EMD procedure to the different interpolation methods used in its implementation, we consider synthetic time series data. The advantage of using synthetic data is that the actual signal components are known, which permits direct comparison with the IMFs obtained from EMD and EEMD.

We are specifically interested in examining the sensitivity of the EMD/EEMD method on time series that exhibit non-linear and/or non-stationary behavior. To this end, we begin by considering a time series $x_0(t)$ comprised of

ensemble IMFs [12]. Wu and Huang [31] added noise with δ equal to 0.1, 0.2, and 0.4 times the standard deviation of the data, and demonstrated that in each case, EEMD performed better than EMD. However, as a general rule, they advocated adding noise with δ equal to 0.2 of the standard deviation of the data. For this reason, in this paper, we implement the EEMD method by adding noise ($\epsilon \sim N(0, \delta)$) with δ equal to 0.2 times the standard deviation of the data.

In this study, we focus on Step 2 of the EMD algorithm (Algorithm 2.1), which determines the upper and lower envelopes using some form of interpolation. In addition to the Hermite cubic-spline interpolant (implemented using *spline* function in MATLAB) originally advocated by Huang [13], we use piecewise-linear interpolation between adjacent nodes and smoothing splines. For brevity, we refer to the EMD procedure based on linear interpolation as "linear EMD," the EMD procedure based on cubic interpolation as "cubic EMD," and the EMD procedure based on smoothing splines as "smoothing EMD." Analogous terminology will also be adopted for the various EEMD methods.

---

[1] http://rcada.ncu.edu.tw/intro.html

[2] http://perso.ens-lyon.fr/patrick.flandrin/emd.html

[3] http://au.mathworks.com/matlabcentral/fileexchange/19681-hilbert-huang-transform

two (deterministic) sinusoidal signals with low and high frequencies:

$$x_0(t) = L(t) + H(t); \quad t = 1, 2, \ldots, 1000, \tag{3}$$

where $L(t)$ and $H(t)$ are defined as

$$L(t) = 0.4 \sin\left(\frac{2\pi t}{200}\right); \quad H(t) = 0.2 \sin\left(\frac{2\pi t}{25}\right);$$
$$t = 1, 2, \ldots, 1000. \tag{4}$$

A perfect transform will recover $L(t)$ and $H(t)$ exactly as IMFs.

To examine the ability of EMD/EEMD to recover linear and non-linear trend components from an input signal, we also consider (deterministic) time series defined by adding a variety of trend functions $T_i(t)$ to $x_0(t)$. Specifically, we consider the following time series:

$$x_i(t) = x_0(t) + T_i(t), \quad i = 1, \ldots, 4; \quad t = 1, 2, \ldots, 1000, \tag{5}$$

where

$$T_1(t) = 1 + \frac{t}{2000}, \tag{6}$$

$$2mm]T_2(t) = 1 + \frac{(t + 300)^2}{200000}, \tag{7}$$

$$T_3(t) = 7 - \ln(1002 - t), \tag{8}$$

$$T_4(t) = 2 \exp(-0.09(20t - 1)^2). \tag{9}$$

The trend functions have been chosen to emulate patterns that might be found in environmental time series: T_1 is a constant-linear trend; T_2 is a non-monotonic (quadratic) trend, which initially decreases then increases; T_3 is a logarithmic trend, which exhibits a sharp rise near the end of the time series—similar to the "hockey-stick" trends encountered in global temperature data associated with global warming; and T_4 is a Gaussian trend, with a sharp increase followed by a sharp decrease.

A perfect transform applied to the time series $x_i(t)$ would recover $L(t)$ and $H(t)$ exactly again as IMFs, and the trend $T_i(t)$ as the residual $r_n(t)$ (cf. Eq. 1).

To test the sensitivity of EMD/EEMD in the presence of noise, we also consider time series obtained by adding stationary and non-stationary noise to those defined by Eqs. 3 and 5. Specifically, we define:

$$y_0(t) = x_0(t) + \epsilon_0^0(t); \quad t = 1, 2, \ldots, 1000, \tag{10}$$

where $\epsilon_0^0 \sim N(0, \delta_0^0)$ represents stationary (constant variance) noise, with δ_0^0 taken as 20% of the standard deviation of x_0 (calculated as the mean squared difference from the average of the signal over time).

Similarly, we define the following time series:

$$y_i(t) = x_i(t) + \epsilon_i^0, \quad i = 1, \ldots, 4; \quad t = 1, 2, \ldots, 1000, \tag{11}$$

To further examine the sensitivity of EMD/EEMD in the presence of data non-stationarity, we also consider the following time series:

$$z_{i0}(t) = x_0(t) + \epsilon_i(t), \quad i = 1, 2; \quad t = 1, 2, \ldots, 1000, \tag{12}$$

and

$$z_{ij}(t) = x_j(t) + \epsilon_i(t), \quad j = 1, 2, \ldots 4; \quad i = 1, 2;$$
$$t = 1, 2, \ldots, 1000. \tag{13}$$

Here, the time-dependent noise terms are taken as $\epsilon_i(t) \sim N(0, \delta_i(t))$, with:

$$\delta_1(t) = 0.0005t, \quad t = 1, 2, \ldots, 1000; \tag{14}$$

and

$$\delta_2(t) = 0.0005 + 0.5 \exp\left(-(0.006t - 3)^2\right),$$
$$t = 1, 2, \ldots, 1000. \tag{15}$$

Hence, the time series $z_{i0}(t)$ represents data with no trend and two different types of non-stationary noise: monotonically increasing standard deviation and standard deviation that varies in a Gaussian manner (increases and then decreases) over time; while $z_{ij}(t)$ represents data with various trend components and the two different types of non-stationary noise. The time series defined by Eq. 12 are shown in the bottom two panels of Fig. 2.

The various time series considered in this study are summarized in Table 1.

3 Sensitivity of EMD and EEMD to Interpolation Methodology

The performance of the EMD and EEMD procedures with the three different interpolation methods is assessed using the twenty synthetic time series introduced in the last section. We begin by considering how well EMD and EEMD are able to recover the sinusoidal signals $L(t)$ and $H(t)$ from the input signals x_0 and y_0. We then investigate their performance in recovering the sinusoidal signals and the various trends (6 to 9) from the input time series x_i and y_i, $i = 1, \ldots, 4$.

We first show the EMD/EEMD original results for two selected data sets and then discuss how we analyze the results. It should be mentioned that in the EEMD results, it is normal to obtain more modes than we expect but the amplitude of some of these modes is too small for the modes to carry valuable information. Therefore, eliminating

Fig. 2 Synthetic data with stationary noise and **(a)** constant trend y_1; **(b)** quadratic trend y_2; **(c)** logarithmic trend y_3; **(d)** Gaussian trend y_4; **(e)** non-stationary noise z_{10}; **(f)** non-stationary noise z_{20}
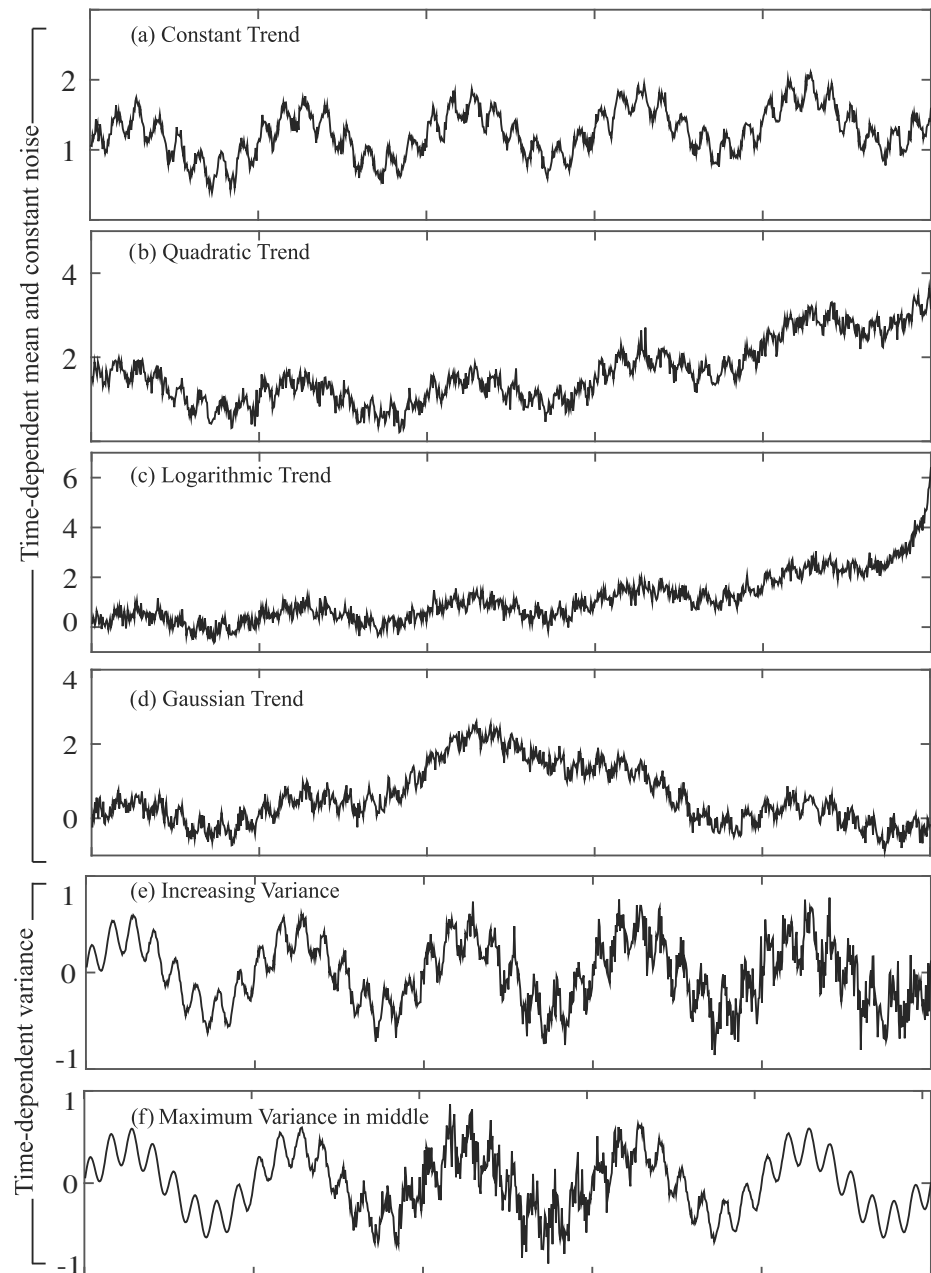
these modes will not change the final analysis. In this study, the modes with small amplitudes (smaller than 1×10^{-2}) are eliminated following the selection criteria introduced in

Table 1 Summary of the time series used as input to EMD/EEMD

Data	Description	Equation
x_0	Sum of sinusoids L & H	Eq. 3
x_i	x_0 + various trends (T_i)	Eqs. 5; 6–9
y_0	x_0 + stationary noise (ϵ_0)	Eq. 10
y_i	x_0 + various trends (T_i) + stationary noise (ϵ_i)	Eq. 11
z_{i0}	x_0 + non-stationary noise (ϵ_i)	Eqs. 12; 14
z_{ij}	x_0 + various trends (T_j)+ non-stationary noise (ϵ_i)	Eq. 13

[19]. Figure 3 shows the results of LEEMD (linear-EEMD), CEEMD (cubic-EEMD) and SEEMD (smoothing-EEMD) applied to y_1.

In the presence of noise, the results are not as straightforward. Figure 3 shows the EMD outputs for the three interpolation methods for the y_0 data (10). The first two modes in LEEMD and CEEMD results correspond to the noise, and the third mode is the high-frequency mode. For the SEEMD results, the first mode is the noise, and the second mode is the high-frequency mode. The amplitude of this mode in SEEMD is underestimated due to the smoothing parameter being set to extract the low-frequency mode more accurately (see Section 3.6).
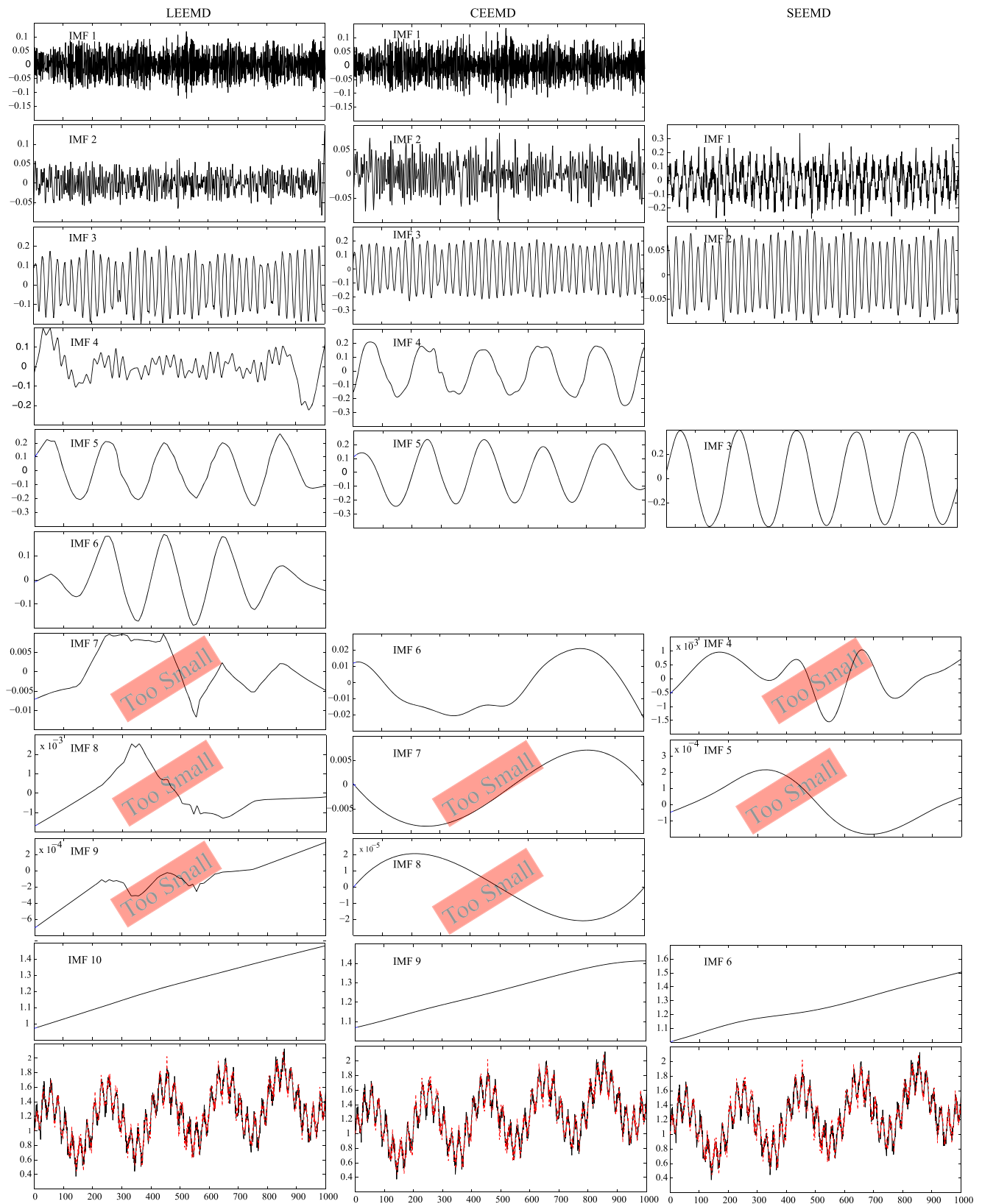
Fig. 3 EEMD results for the noisy y_1 data (10) using LEEMD (left column), CEEMD (middle column), and SEEMD (right column). The top two rows are noise, the third row is the high-frequency modes, and rows 4 to 6 are low-frequency modes extracted by three methods. The bottom row shows the sum of the two extracted modes (red dash line) compared with the original signal y_0 (black solid line)

LEEMD and CEEMD struggle to extract the low-frequency mode. In LEEMD, the low-frequency mode appears in three IMFs (the first of which still has high-frequency mode mixing) in CEEMD in two IMFs. This is a common challenge of the method; the user can either choose one of the two or three modes or add all the modes with similar time periods together [19]. It is not clear which is the best method.

Adding the modes and then adding the trend do not cause loss of any information but, in some cases, we are interested in studying a mode with a particular time period; then by adding the modes, we lose some of the information from that mode. Figure 4 shows the added modes along with the original low-frequency mode resulting from LEEMD, CEEMD, and SEEMD. This figure shows that although adding the modes with similar frequency components resulting from LEEMD and CEEMD produces results quite close to the actual mode, they are still less accurate than that produced in a single mode using SEEMD. In LEEMD, as can be seen in Fig. 4, some of the information from the higher-frequency signal are producing error, and in the CEEMD results, there is underestimation near the end points. SEEM, however, produces the most accurate results of the three interpolation methods.

Another problem is in deciding which modes to add up. This is quite challenging, especially when dealing with data with several high- or low-frequency components. Then, the IMFs are often a combination of two or more modes. Obviously, in both of these situations, some information will be lost from the analysis. The ideal situation is that the method should result in pure IMFs in the first place. Treating the problem objectively, we did not add modes together but chose the modes that had the closest frequency and amplitude to the actual modes. For example, for the low-frequency modes in the y_1 results, in LEEMD the fifth IMF, in CEEMD the fourth IMF, and in SEEMD the third
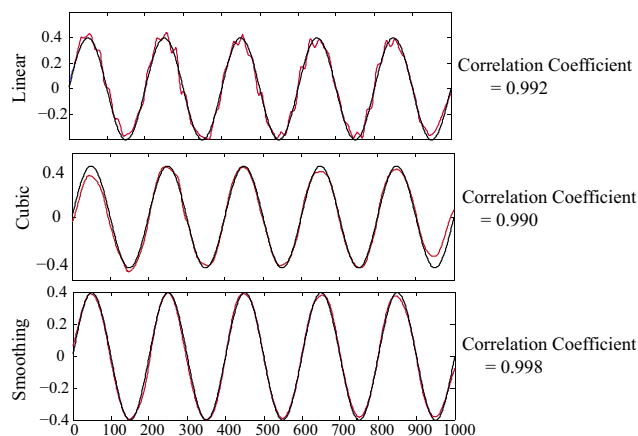
IMF have the closest frequency and amplitude to the actual low-frequency mode.

The mean absolute error (MAE) is defined to compare the modes (IMFs) from EMD/EEMD with the actual modes and trends:

$$\text{MAE}_I = \frac{\sum_{k=1}^{N} |M(t_k) - \text{IMF}_I(t_k)|}{N}. \tag{16}$$

Here, $M(t_k)$ can be either $L(t_k)$ or $H(t_k)$ or one of the trends $T_i(t_k)$ (6 to 9), and $\text{IMF}_I(t_k)$ is the corresponding IMF. The subscript I denotes the interpolation method: L linear; C cubic; S smoothing; and N is the number of the data points.

3.1 EMD/EEMD Applied to x_0 and y_0

We consider first the performance of the EMD procedure applied to the noiseless input time series $x_0(t)$ and that of the EEMD procedure applied to the corresponding time series with noise, $y_0(t)$. The ability of EMD and EEMD to retrieve the component signals in these cases will establish a benchmark for subsequent analyses of the more complex synthetic time series.
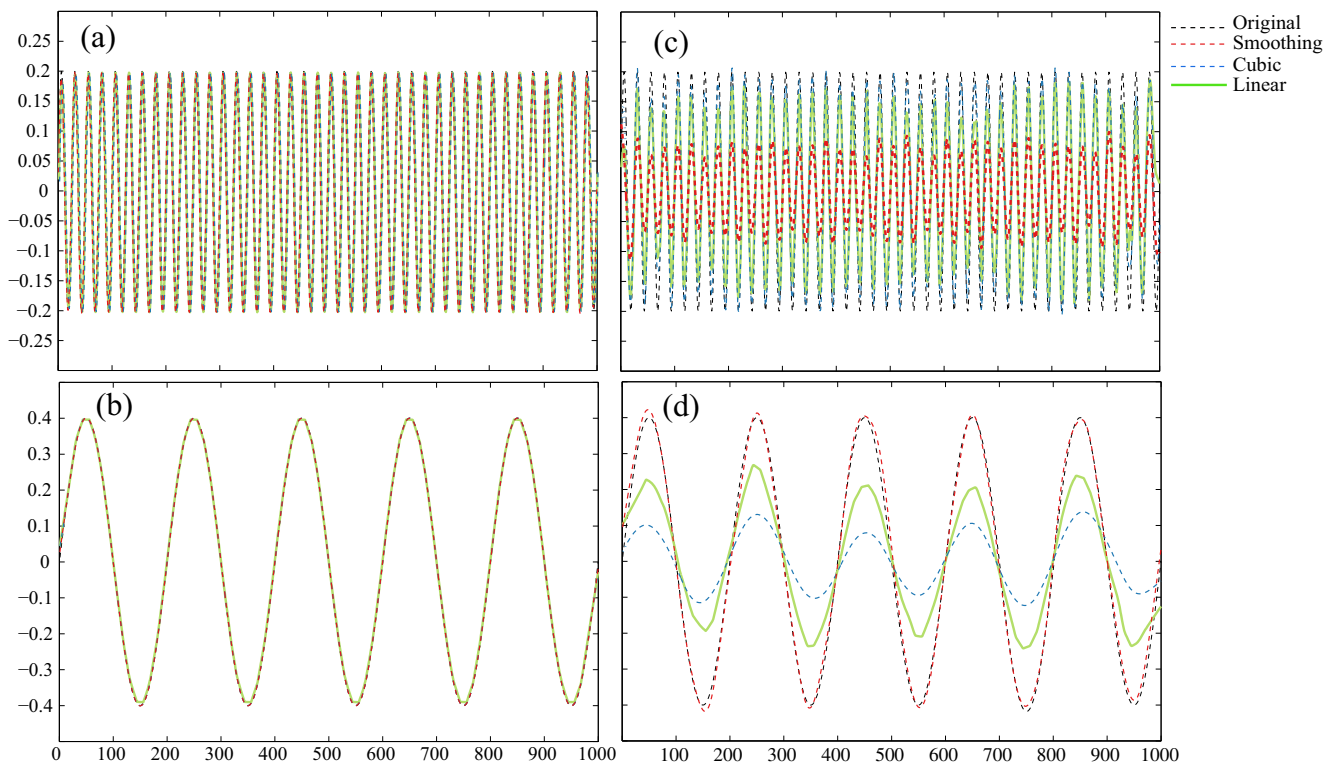
Figure 5 shows that EMD applied to x_0 is able to recover the high- and low-frequency sinusoidal components to a very high degree of accuracy. EMD with cubic- and smoothing-spline interpolations performed the best, with the MAEs of 6.68×10^{-4} and 5.22×10^{-4}, respectively, for the low-frequency component, and 6.68×10^{-4} and 4.56×10^{-4}, respectively, for the high-frequency component. Linear EMD exhibited poorer performance, with an MAE of 3.25×10^{-3} for both the low- and high-frequency components.

EEMD applied to the noisy time series y_0 is able to accurately recover the frequencies of the sinusoidal components, but performs less well in capturing their amplitudes, as can be seen in Fig. 5c, d. For the high-frequency component, cubic EEMD performs the best with an MAE of 2.22×10^{-2}, while smoothing EEMD performs the worst with an MAE of 7.86×10^{-2}. Linear EEMD is able to capture the high-frequency component to a similar degree of accuracy as the cubic EEMD, with an MAE of 3.69×10^{-2}, but produces very inaccurate results for the low-frequency component, with an MAE of 1.15×10^{-1}. For the low-frequency component, smoothing EEMD performs the best, with an MAE of 1.61×10^{-2}, while cubic EEMD yielded an MAE of 1.86×10^{-1}.

The large error in smoothing EEMD for the high-frequency component can be reduced by changing the smoothing parameter. This will be discussed in more detail in Section 3.6.

We note the considerable damping of the amplitude in the low-frequency signals retrieved from the cubic and linear

Fig. 4 EEMD low-frequency modes for y_1 data (10) using LEEMD (top row), CEEMD (middle row), and SEEMD (bottom row). The red lines show the sum of the extracted low-frequency modes; the black lines are the original low-frequency modes

Fig. 5 Panels **a** and **b** EMD results for the noiseless time series x_0 (3); panels **c** and **d** EEMD results for the corresponding data with noise y_0 (10). Panels **a** and **c** high-frequency IMFs; panels **b** and **d** low-frequency IMFs. Original is either the high-frequency component $H(t)$ (top panels) or the low-frequency component $L(t)$ (bottom panels) of the input signal

Table 2 Mean absolute error (MAE) of the EMD (x_1–x_4) or EEMD IMFs $\times 10^{-3}$ for all the data inputs

Input	High frequency			Low frequency			Trend		
data	L(E)EMD	C(E)EMD	S(E)EMD	L(E)EMD	C(E)EMD	S(E)EMD	L(E)EMD	C(E)EMD	S(E)EMD
x_0	3.30	0.66	0.45	3.2	0.67	0.52	NA	NA	NA
x_1	3.30	0.66	0.45	3.2	0.67	0.52	0.092	0.095	0.17
x_2	3.20	0.63	0.42	16.6	13.6	14.3	16.12	13.67	14.36
x_3	6.80	3.30	5.6	81.3	56.6	72.0	350.23	284.43	255.19
x_4	3.20	0.65	0.47	135	31.3	25.9	319.8	56.7	73.7
y_0	36.9	22.2	78.6	115.7	186.5	16.1	NA	NA	NA
y_1	36.3.8	22	78.9	123.6	131.3	15.2	19.38	25.4	7.22
y_2	82	42.9	82.5	100.3	48.4	33.7	57.06	26.25	22.09
y_3	76.7	56.0	81.40	114.3	69.5	166.1	349.7	327.9	267.4
y_4	81.2	53.5	80.1	123.2	55.3	41.3	104.86	53.16	62
z_{10}	50.2	41.4	77.3	167.8	941.0	16.5	NA	NA	NA
z_{11}	54.1	46.4	75.9	158.4	96.5	19	32.65	19.97	12.51
z_{12}	53.1	47.2	79.5	143.8	108.0	72.8	84.78	16.48	26.17
z_{13}	58.0	47.8	79.6	178.1	118.1	55.5	389.9	419.8	283.76
z_{14}	60.6	50.7	79.4	174.2	116.4	35.4	90.99	224.74	79.27
z_{20}	46.6	48.7	73.2	160.3	147.3	10.4	NA	NA	NA
z_{21}	46.9	48.5	74.8	146.9	150.4	12.0	45.9	12.08	16.2
z_{22}	51.4	46.2	76.2	162.3	135.6	21.9	104.102	36.109	23.78
z_{23}	48.1	49.0	77.9	171.2	193.8	75.7	319.77	388.4	189.25
z_{24}	50.0	49.2	76.4	184.7	177.6	25.5	124.8	99.6	74.8

L(E)EMD, EMD/EEMD with linear interpolation; *C(E)EMD*, EMD/EEMD with cubic-spline interpolation; *S(E)EMD*, EMD/EEMD with smoothing-spline interpolation

EEMD is due to the fact that in these cases, the signal is divided over more than one IMF (Fig. 3).

3.2 EEMD Applied to x_i and y_i

EMD applied to the noiseless time series with constant-linear trend x_1 performed very well regardless of the interpolation method used, producing identical MAE statistics to those obtained from x_0 for the sinusoidal components. Similarly, EMD based on each of the interpolation methods was able to accurately recover the constant linear trend, with linear EMD, cubic EMD, and smoothing EMD yielding MAEs of 9.25×10^{-5}, 9.50×10^{-5}, and 1.79×10^{-4}, respectively.

The corresponding MAE statistics for the time series x_2, x_3, and x_4 can be seen in Table 2. In the absence of noise, cubic EMD and smoothing EMD provided the better estimates of the sinusoidal components and the various trend components. Linear EMD performed considerably worse in estimating the logarithmic trend component of x_3, while EMD with all the interpolation methods had difficulty in recovering the Gaussian trend component of x_4.

Figure 6 shows the results of the various EMD procedures applied to x_3 and x_4. It is evident that the nature of the logarithmic trend has resulted in poor EMD performance; this is particularly noticeable in the inability of the linear EMD to estimate the low-frequency sinusoidal component (Fig. 6b), though similar issues can be seen
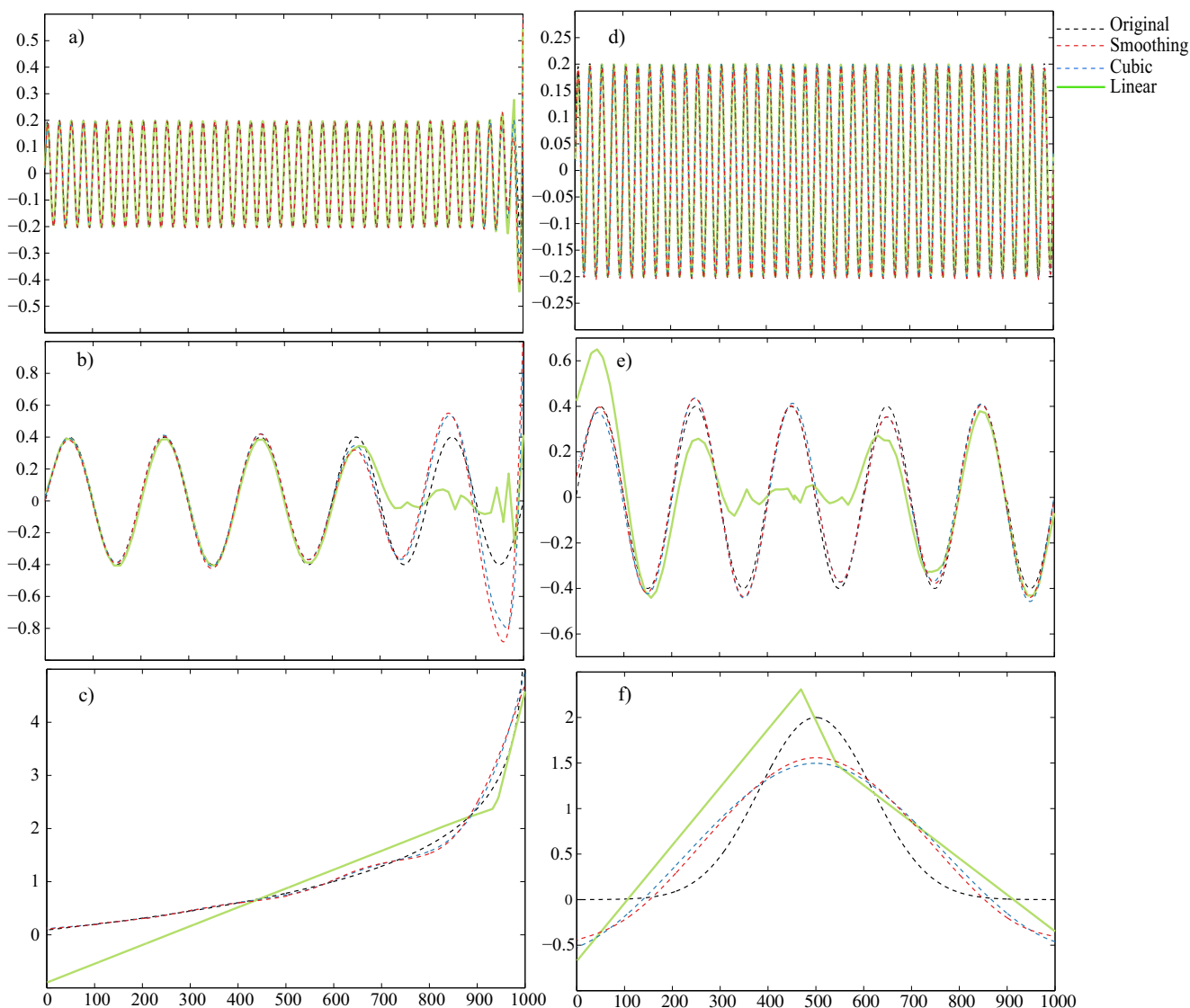
Fig. 6 Panels **a**, **b**, and **c** EMD results for x_3 data (5); panels **d**, **e**, and **f** EMD results for the x_4 data (5). Top panels: high-frequency IMFs; middle panels: low-frequency IMFs; and bottom panels: the residuals and original trends

with the cubic and smoothing EMD results (Fig. 6b). It has also problems with the estimation of the high-frequency component near the end of the time series for the all interpolation methods (Fig. 6a).

With the Gaussian trend, the high-frequency component was recovered well by all three interpolation methods but linear EMD again performed poorly for the low-frequency component. Comparing Fig. 6b and e shows that the region of particularly poor performance of the linear EMD coincides with where the trend component exhibits the greatest variability.

The various EMD methods, particularly linear EMD, performed relatively poorly in estimating the Gaussian trend component of x_4 (Fig. 6f). The cubic and smoothing EMD methods were able to capture the general pattern and symmetry of the trend component, but underestimated its variability.

The MAE statistics resulting from application of EEMD to the time series constant noise, y_1, y_2, y_3, and y_4, are given in Table 2. For the constant-linear trend time series y_1, smoothing EEMD resulted in a more accurate recovery of the three signal components, in particular the

Fig. 7 EEMD results for the y_3 data (5)
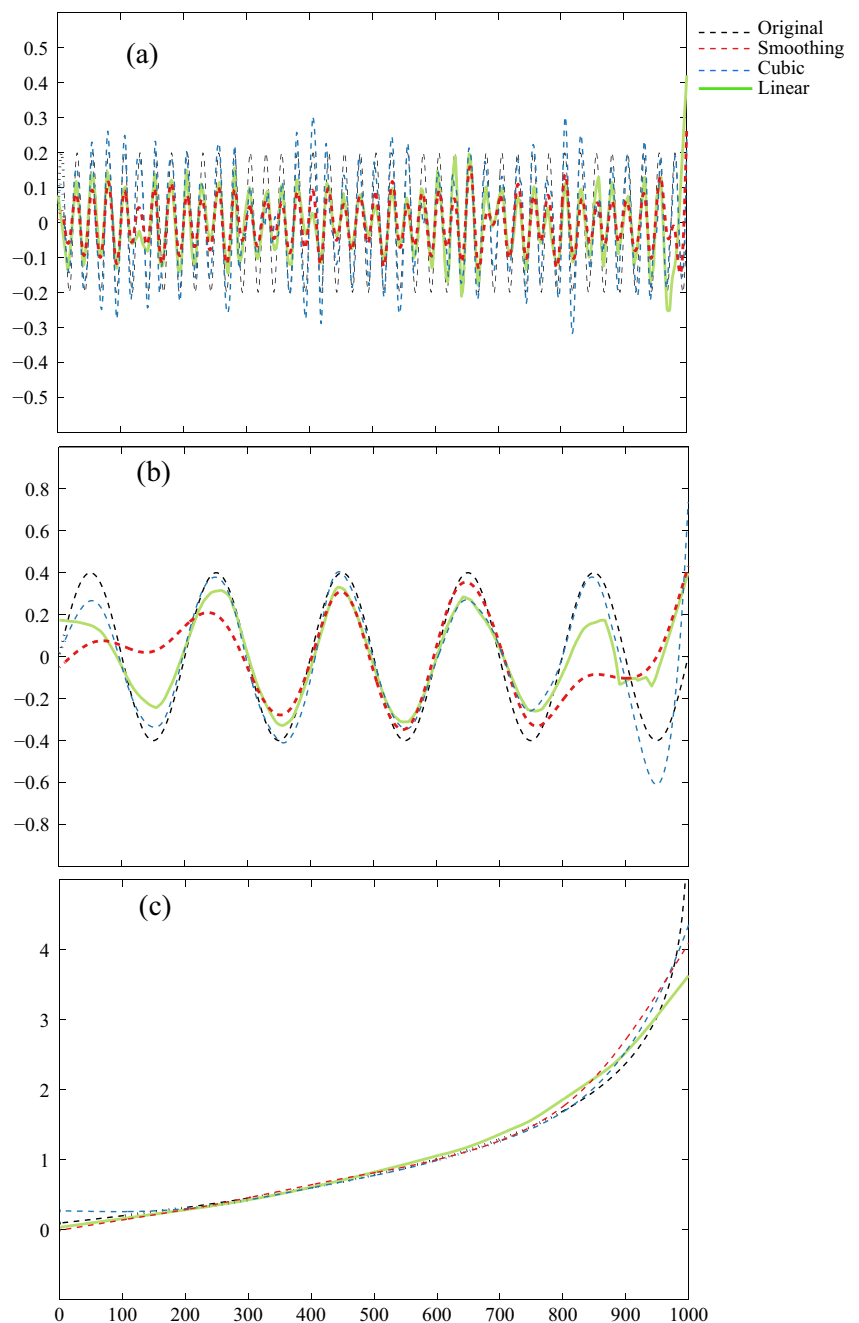
constant-linear trend and the low-frequency sinusoidal components, much better than that of the linear and cubic EEMD. Indeed, the MAE for the low-frequency component was an order of magnitude smaller with smoothing EEMD. Smoothing EEMD also dealt better with noise but it is important to set the right smoothing parameter (see Section 3.6).

For the time series y_2 with quadratic trend, smoothing EEMD again provided the lowest MAE value for the low-frequency component, although those from cubic EEMD are only slightly higher. Similarly, for the time series y_4 with Gaussian trend, smoothing EEMD yielded the lowest MAE for the low-frequency component. The exception was the logarithmic trend in y_3, for which cubic EEMD yielded a lower MAE than smoothing EEMD (see Table 2). As can be seen in Fig. 7, linear EEMD had considerable difficulty in accurately reproducing the amplitudes of both sinusoidal components.

For the noisy data with trends, y_1 to y_4, the cubic EEMD and linear EEMD results both had higher MAEs than for the corresponding noiseless data, x_1 to x_4. These higher values are because of considerably more mode mixing than the noiseless data. This shows that cubic EEMD and linear EEMD are sensitive to noise. The smoothing EEMD is less prone to noise, therefore to mode mixing, especially for low-frequency components.

3.3 EEMD Applied to z_{10} and z_{20}

The results of applying the various EEMD methods to the time series z_{10} and z_{20} are presented in Fig. 8. The ability of linear EEMD to recover the high-frequency sinusoidal component was considerably affected by the noise, a result which was consistent with those relating to the application of linear EEMD to the time series y_i. Again, it is the amplitudes of the estimated components that were most sensitive to the presence of noise, whereas their frequencies appear to be far more robust. The regions of poorest performance of linear EEMD coincided with the regions where the amplitude of the noise was greatest, that is, near the end of the time series for z_{10} and in the middle of the time series for z_{20}. For z_{10}, linear EEMD yielded an MAE for the high-frequency component of 5.02×10^{-2}, while cubic and smoothing EEMDs produced MAEs of 4.14×10^{-2} and 7.73×10^{-2}, respectively. For the case of z_{20}, in which the amplitude of the noise is greatest in the middle of the the time series, linear EEMD produced an MAE of 4.66×10^{-2}, while cubic and smoothing EEMDs produced MAEs of 4.87×10^{-2} and 7.32×10^{-2}, respectively.

As mentioned earlier, due to the interest of this study in long-term variability, the smoothing parameter is biased towards the accuracy of the low-frequency IMFs. Unfortunately, that means the high-frequency IMFs
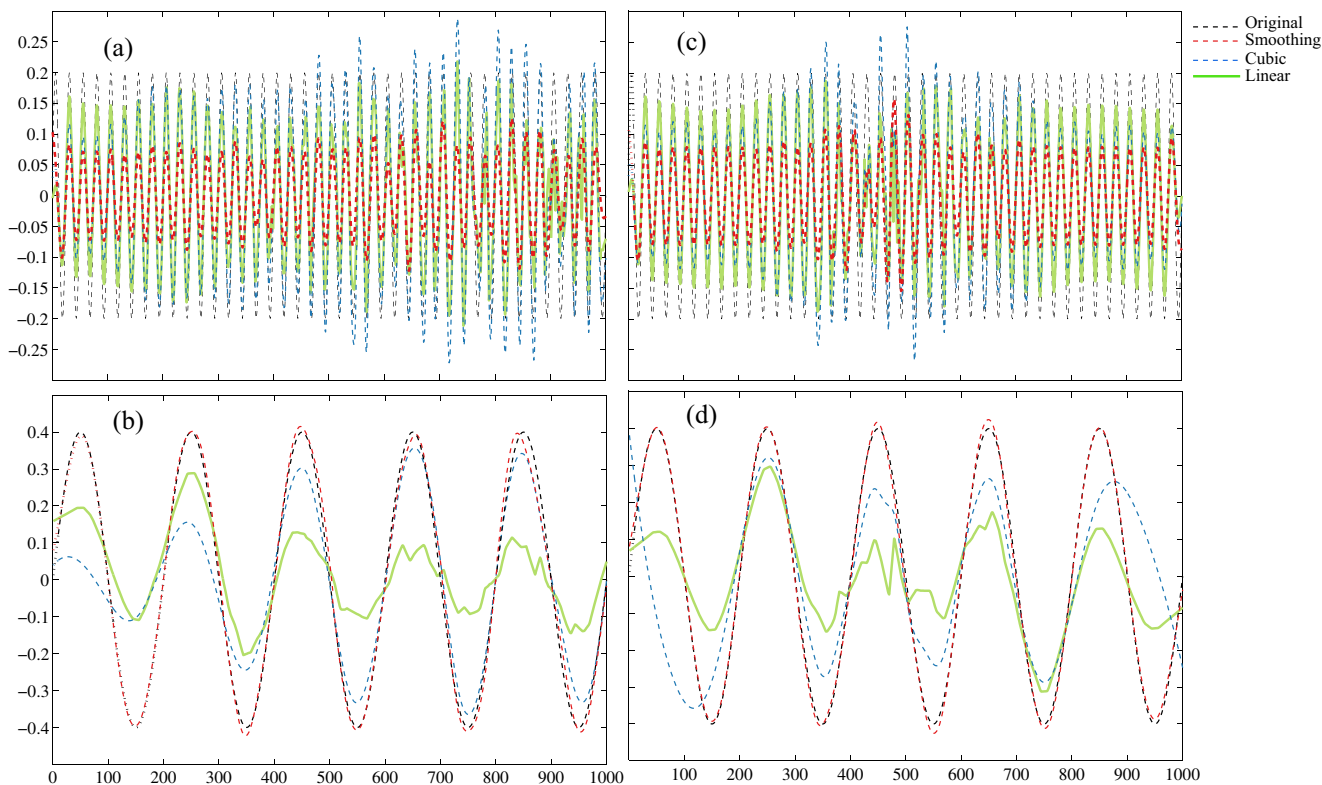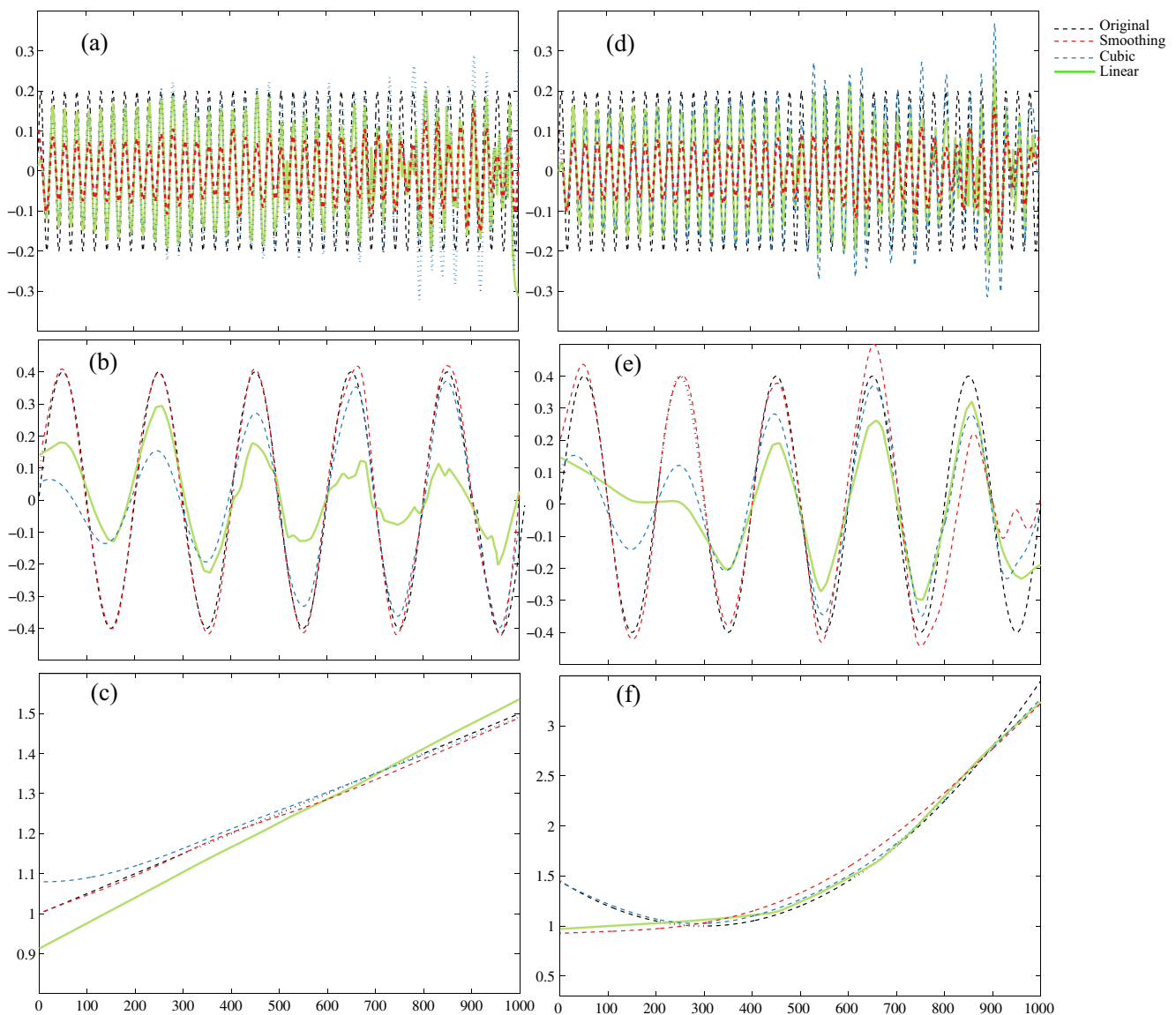

Fig. 8 EEMD results for the data with non-stationary noise. Panels **a** and **b** z_{10}; panel s**c** and **d** z_{20} (12)

obtained from smoothing EEMD are not as accurate as they could be, and the values of the corresponding MAEs are relatively high. These high MAEs are due to errors in estimating amplitude rather than frequency. So if the smoothing parameter in smoothing EEMD is not suitable for a specific frequency range, the method can still find the correct frequency but will underestimate the amplitude. This issue can easily be fixed in individual cases by changing the smoothing parameter to a more suitable value (for more details about choice of smoothing parameter, see Section 3.6).

For the low-frequency sinusoidal component, Fig. 8 shows that linear EEMD is badly affected by the presence of noise, and worst affected where the amplitude of the noise is greatest. Cubic EEMD also seems to perform poorly

in these cases, though the regions of poorest performance occur at the ends of the time series rather than where the noise is greatest in amplitude. This suggests that end effects may have some bearing on the results of cubic EEMD; further analysis is required to confirm this possibility. For z_{10}, linear EEMD yielded an MAE for the low-frequency component of 1.67×10^{-1}, while cubic and smoothing EEMDs produced MAEs of 9.41×10^{-1} and 1.65×10^{-2}, respectively. For z_{20}, linear EEMD produced an MAE of 1.60×10^{-1}, while cubic and smoothing EEMD produced MAEs of 1.47×10^{-1} and 1.04×10^{-2}, respectively. These results indicate that smoothing EEMD also provides the most accurate estimates of the low-frequency sinusoidal component for the two cases of time-dependent variance considered; there is no amplitude underestimation in the

Fig. 9 EEMD results for data with the non-stationary noise and various trends, z_{11} (constant-linear trend) and z_{12} (quadratic trend; Eq. 12)

results and therefore no artificially high MAEs. It is worth noting that the performance of smoothing EEMD does appear to also work well near the end of the time series in the z_{20} case (Fig. 8). This is evidence that shows the smoothing EEMD with a suitable smoothing parameter is a good alternative to the standard (cubic interpolation) EEMD, with greater accuracy and less mode mixing.

3.4 EEMD Applied to z_{ij}, $i = 1, 2, j = 1, \ldots, 4$

Here, we consider the results of applying the various EEMD methods to the time series z_{1j} and z_{2j} with $j = 1, \ldots, 4$. Table 2 details the resulting MAE statistics for the various cases, while Figs. 9 and 10 show some selected examples of EEMD performance.

Consistent with the results in the previous section, linear EEMD was highly sensitive to the presence of noise in the input time series. This is clearer when we consider the MAEs for the x_i ($3.3 - 6.8$) and the y_i ($36.9 - 81.2$) in Table 2. Again, it was the amplitude of the estimated signals that was most badly affected (e.g., see Fig. 9). Interestingly however, linear EEMD was better able to accurately reproduce the trend component of the input signals, in contrast to its ability to accurately reproduce their sinusoidal components. Figures 9 and 10 show that cubic EEMD was also prone to poor performance, particularly in its ability to reproduce the low-frequency sinusoidal component. Table 2 shows that apart from the data with logarithmic and Gaussian trends, linear EEMD results for the trend have smaller MAEs than for the low-frequency
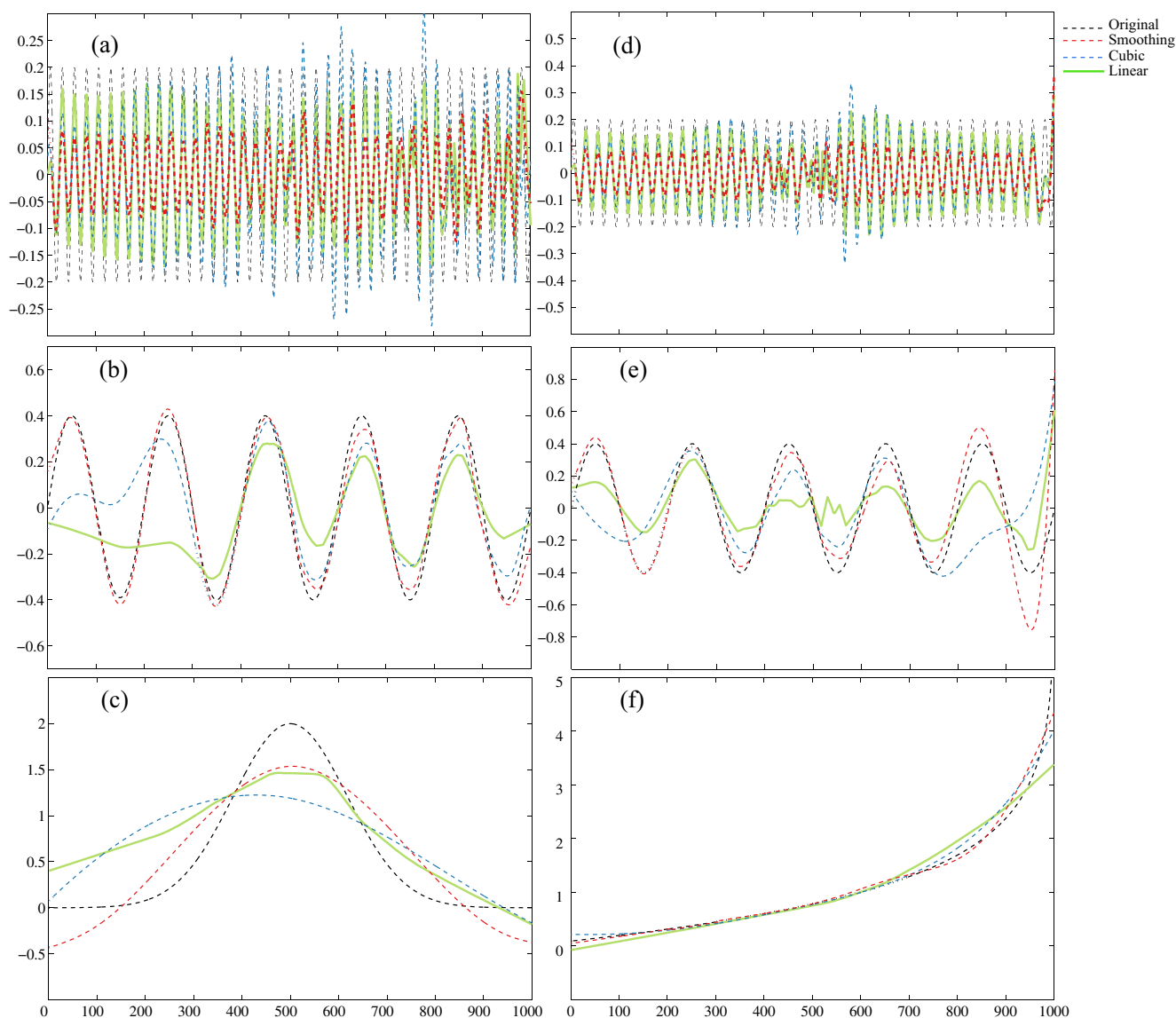
Fig. 10 EEMD results for the data with non-stationary noise and various trends, z_{23}, (logarithmic trend) and z_{14}, logarithmic tend; Eq. 12)

mode. Figure 9 shows that cubic EEMD is also prone to poor performance, particularly in its ability to reproduce the low-frequency sinusoidal component.

Comparison of MAEs in Table 2 shows that, overall, smoothing EEMD provided the most accurate estimation of the original signal components. Indeed, smoothing EEMD produced the smallest MAE values in all but two of the cases considered, namely z_{12} and z_{21}, for which cubic EEMD yielded the smallest MAE values. However, in these cases, the smoothing EEMD mean absolute errors were only slightly larger. For the sinusoidal components, smoothing EEMD consistently produced the smallest MAE values. For the low-frequency components, smoothing EEMD produced MAE values an order of magnitude smaller than those obtained from linear and cubic EEMD.

In terms of estimating the trend components of the input signal, it was the logarithmic and Gaussian trends that most confound the various EEMD procedures. Figure 10

shows that it was the abrupt changes in these trend components that represented by these trend components resulted in poor EEMD performance. For the logarithmic trend cases, this is particularly evident near the end of the time series, whereas for the Gaussian trend cases, the EEMD methods failed to accurately estimate the full extent of the trend, similar to the case of y_4. However, despite its relatively poor performance, smoothing EEMD better captured the variability and symmetry of the Gaussian trend component for the two cases, z_{14} and z_{24}, as was the case with y_4.

3.5 Robustness of the EEMD Ensemble Members

In this section, we consider the consistency of the three interpolation methods in EMD/EEMD in estimating the trend component of the various input time series. To do this, we construct confidence intervals based on the full spectrum
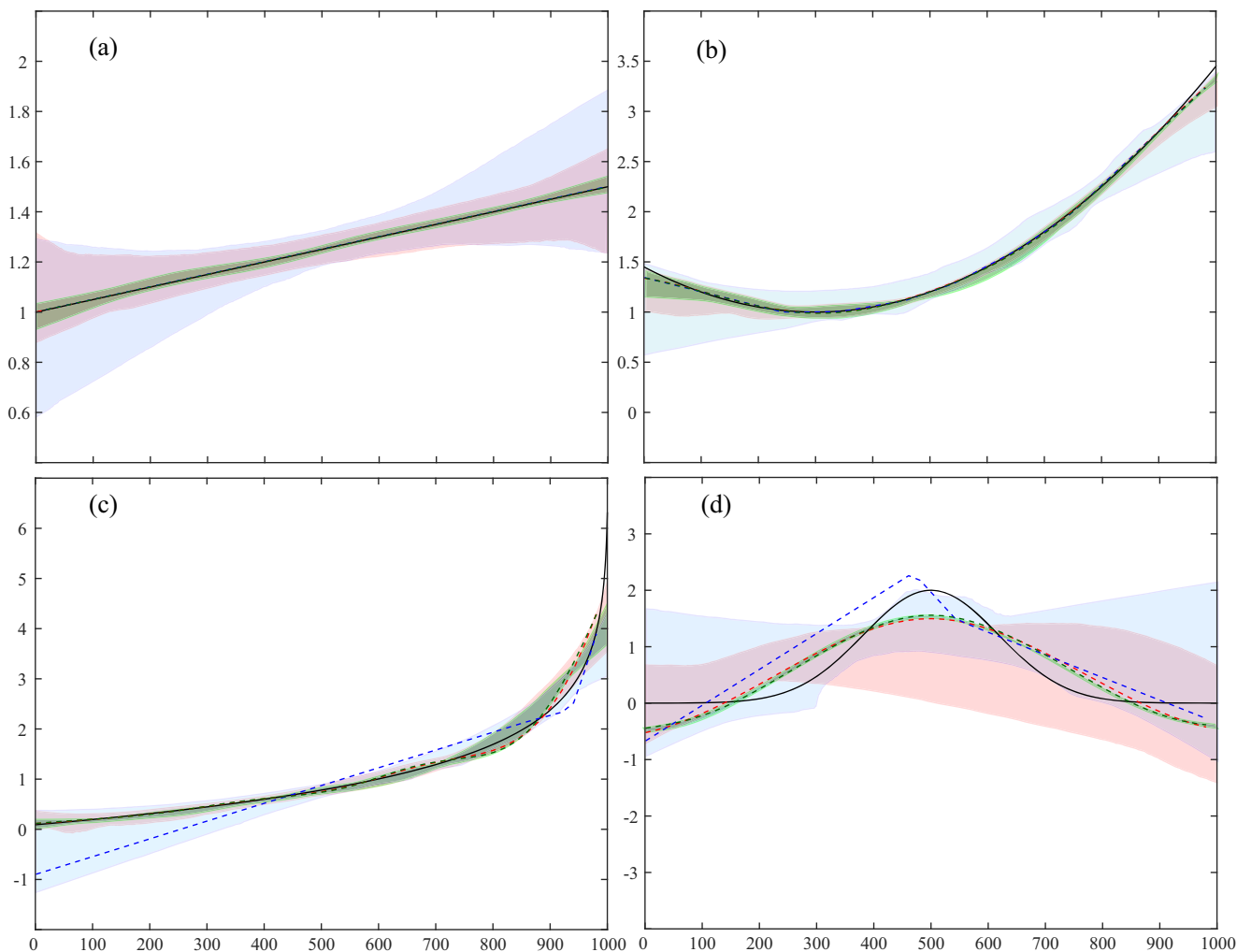
Fig. 11 95% confidence intervals for the trend results for the x_1 (**a**), x_2 (**b**), x_3 (**c**), and x_4 (**d**) data. Blue shading, linear EMD; red shading, cubic EMD; and green shading, smoothing EMD

of the EEMD ensemble members. Figures 11 and 12 show EEMD trend components together with 95% confidence intervals (shaded), which are determined by considering the point-wise distribution of the 3000 estimated trend components (i.e., the ensemble members) across the time series domain. Thus, at each point in time, the lower confidence limit defines the value below which 2.5% of the ensemble member values fell, while the upper confidence limit defines the value above which 2.5% of the ensemble member values fell.

Figure 11 shows the trend components estimated by applying the various EEMD methods to the time series x_1, x_2, x_3, and x_4. The confidence intervals in Fig. 11a indicate that there is a considerable degree of variability in the constant-linear trends determined in each of linear EEMD ensemble runs. In particular, the confidence intervals indicate that linear EMD produced constant-linear trend estimates with gradients ranging from 0.58 to 1.3

(width of trend estimation = 0.72). Figure 11a shows that cubic EEMD also produced a fairly wide spread of linear trend estimates, from 0.88 to 1.32 (width of trend estimation= 0.44). In contrast, the confidence intervals associated with smoothing EEMD were very small, from 0.93 to 1.03 (width of trend estimation = 0.1), indicating greater consistency in the trend components produced over the 3000 smoothing EMD ensemble runs.

Similar results are evident for the quadratic, logarithmic, and Gaussian trend components in Fig. 11b, c, and d, respectively. Linear EMD produced the broadest range of trend estimates, while smoothing EMD exhibited considerable robustness, producing consistent estimates of the various trends across the 3000 ensemble runs. In the case of the Gaussian trend in Fig. 11d, the confidence intervals associated with linear and cubic EEMD indicate that they were far less consistent than smoothing EMD in their ability to estimate the trend. However, it should be noted that the
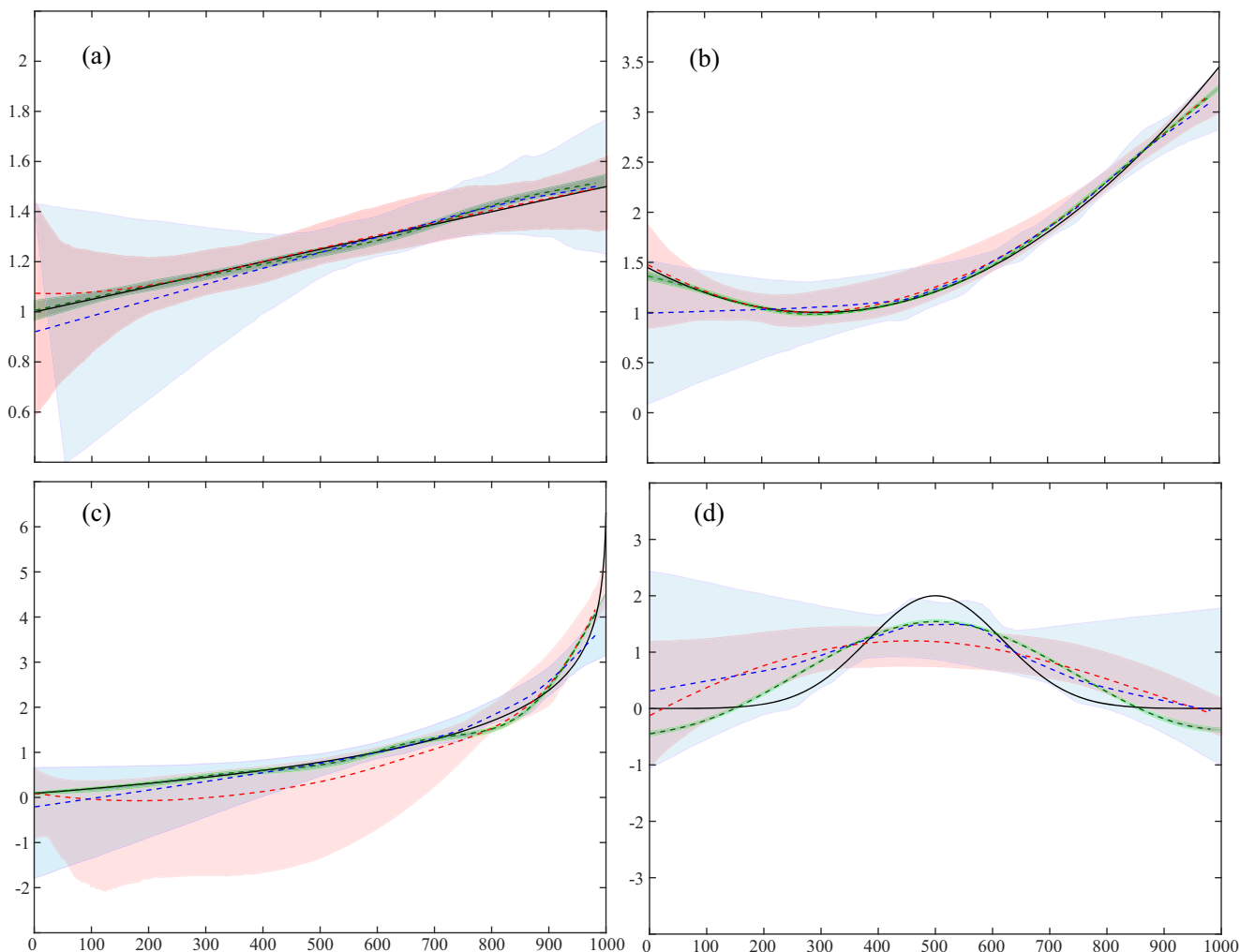

Fig. 12 As for Fig. 11 but for the trend results for z_{1j}, $j = 1, 2, 3, 4$. Blue shading, linear EEMD; red shading, cubic EEMD; and green shading, smoothing EEMD
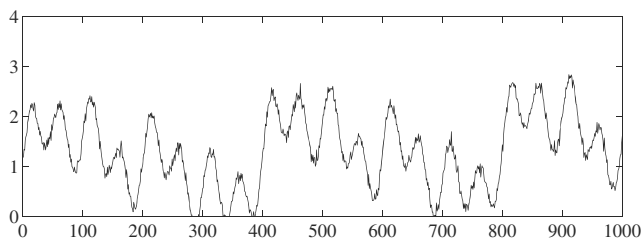
Fig. 13 Input data $Y(t)$ (17)

linear EEMD confidence intervals are the only ones that encompassed the actual trend component.

Figure 12 shows that similar patterns in the confidence intervals arose for the time series data with time-dependent mean and time-dependent variance (z_{1j}, $j \geq 1$). The broadest confidence intervals arose in connection with linear EMD, while smoothing EMD again exhibited a considerable degree of robustness across the ensemble members.

Similar results were found for the sinusoidal signal components, but will not be discussed here.

3.6 Varying the Smoothing Parameter in Smoothing EEMD

In the previous sections, smoothing EEMD was implemented assuming a smoothing parameter $p = 0.0015$. This value of the smoothing parameter was shown to result in accurate reproduction of the trend and low-frequency

sinusoidal component, but poor reproduction of the high-frequency sinusoidal component. In this section, we examine if increasing the smoothing parameter results in a more accurate estimate of the high-frequency component, and what effect it has on the accuracy of estimates of the trend and low-frequency component. As $p \rightarrow 1$, smoothing spline interpolation approaches cubic interpolation, so the expectation is that increasing p will produce results more akin to those obtained using cubic EEMD.

In order to do this, we construct synthetic data with four sinusoidal components in which we have inserted two more sinusoidal components between the high-frequency and low-frequency component used previously:

$$Y(t) = S(t) + T_1(t) + \epsilon_0, \tag{17}$$

where $\epsilon_0 \approx N(0, \delta)$, $\delta = 0.2 \, \text{std}(S(t))$, $T_1(t) = 1 + \frac{t}{2000}$, and:

$$S(t) = 0.6 \sin\left(\frac{2\pi t}{400}\right) + 0.4 \sin\left(\frac{2\pi t}{200}\right) + 0.4 \sin\left(\frac{2\pi t}{100}\right)$$
$$+ 0.6 \sin\left(\frac{2\pi t}{50}\right). \tag{18}$$

where $t = 1, 2, \ldots, 1000$. These data are shown in Fig. 13.

SEEMD is used on these data with the smoothing parameter varying in the range $0 < p < 10^{-6}$. The IMFs for each p value are compared with the corresponding sinusoidal components in Eq. 18; the MAEs for each component are shown in Fig. 14.

For $p > 10^{-1}$ and $p < 10^{-3}$, the MAEs for all IMFs fluctuate considerably and are relatively high: some value

Fig. 14 MAEs for the IMFs corresponding to the four sinusoidal components in Eq. 18, as a function of the smoothing parameter p
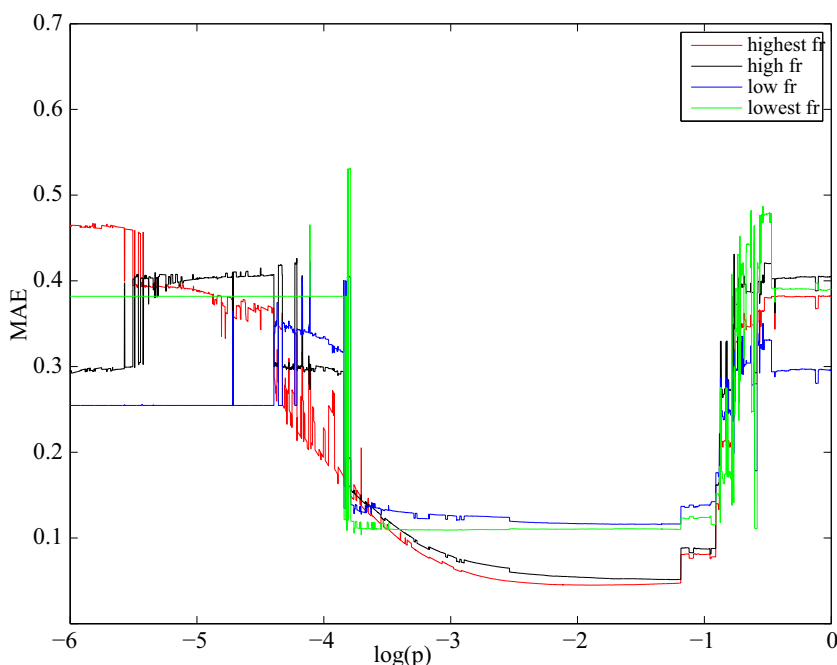
Table 3 Summary of the results

Interpolation	Advantages	Disadvantages
LEEMD	Works equally well with other interpolations in absence of noise	Prone to noise
	Estimate the logarithmic and Gaussian trend well	Prone to end effects
		Extremely sensitive to type of the data
CEEMD	Works generally well and better than linear interpolation	Prone to noise
		Prone to end effects
		Sensitive to type of the data
SEEMD	Accurate results in all cases	
	Not sensitive to noise	Needs adjustment of smoothing parameter
	Not sensitive to data type	
	Not sensitive to end effects	

of p give good results for the higher-frequency IMFs but not for the lower-frequency IMFs, and vice versa (Fig. 14). However, for $10^{-3} < p < 10^{-1}$, the MAEs for all four IMFs are low, making this range suitable values for the smoothing parameter. However, this is for the synthetic data presented here; the range may be different for different data (Table 3). Finding suitable values for the smoothing parameter needs further study.

4 Discussion and Conclusions

Lack of a complete formal mathematical framework underpinning EMD means that a theoretical analysis of the sensitivity of the EMD method is not possible. Studies like this one that examine the sensitivity of EMD using synthetic data are one of the only means of understanding the limitations of the method. In this study, the application of EMD to deterministic time series data and EEMD to noisy time series data has been investigated in the presence of different types of trend and two types of noise (stationary and non-stationary). The effects of varying the method of interpolation, linear, cubic, and smoothing-spline, in the EMD algorithm have also been investigated.

Linear interpolation performed just as well as other interpolation methods in EMD applied to data without noise, but performed poorly in EEMD applied to noisy data. In particular, it suffered from mode mixing and end effects, and consistently underestimated the amplitude of the sinusoidal signal components. Linear interpolation did however provide reasonable estimates of the trend components (except for the data with logarithmic or Gaussian trends).

Cubic EEMD exhibited overall better performance than linear EEMD, but was still prone to mode mixing and end effects in certain cases. Smoothing EEMD (SEEMD) consistently provided the most accurate estimates of the low-frequency sinusoidal component and the various trends.

Because the smoothing parameter is set to favor the low-frequency components, the amplitude of the high-frequency component in most of the cases was underestimated. However, we have shown that it is possible to find a smoothing parameter that gives accurate results for all frequencies. There is a need for further study on how to find suitable smoothing parameter for data in general.

Another advantage of SEEMD is that it is more robust and less sensitive to noise, so that there is less mode mixing. In addition, SEEMD is less sensitive to non-stationarity of the input data. In all cases, there is much less variation in IMF ensemble members using SEEMD.

It is important to note that there are a number of additional interpolation methods that could be used in place of smoothing splines, and that may produce similar stabilizing effects on the EMD procedure—one such method is Gaussian-process interpolation. Investigating the performance of such methods would be an interesting extension of the current study.

We have also shown that EEMD performance depends on the nature of the input data. In particular, it is sensitive to abrupt changes in the underlying trend in the data; for example, the "hockey-stick" logarithmic trend, and the rise and fall of the Gaussian trend. When the data are more complex or there are sudden changes in the trend, EEMD is sensitive to the choice of interpolation method. The reason for this sensitivity is difficult to ascertain in the absence of a sound theoretical basis for EMD, but further analyses using carefully targeted synthetic time series could provide additional insight into this issue.

Estimation of the frequency of the sinusoidal components by the IMFs was less sensitive to the data than estimation of the amplitude—this was because component signals were divided over multiple IMFs for some of the interpolation methods. All the interpolation methods worked reasonably well in recovering the trend with no one method always better than the others. Finding the correct shape of the residual is important in analyzing long-term changes, and

critically important to the accurate study of environmental data [8].

For completeness and future reference, the results of our study for each type of the time series considered in the study are summarized in Table 2.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Bahri, F.M., & Sharples, J.J. (2015). Sensitivity of the Hilbert-Huang transform to interpolation methodology: examples using synthetic and ocean data. In *MODSIM2015, 21st international congress on modelling and simulation. Modelling and simulation society of Australia and New Zealand* (pp. 1324–1330).

2. Chen, Q., Huang, N., Riemenschneider, S., Xu, Y.A. (2006). B-spline approach for empirical mode decompositions. *Advances in Computational Mathematics*, *24*, 171–195.

3. Dätig, M., & Schlurmann, T. (2004). Performance and limitations of the Hilbert–Huang transformation (HHT) with an application to irregular water waves. *Ocean Engineering*, *31*, 1783–1834.

4. Deering, R., & Kaiser, J.F. (2005). The use of a masking signal to improve empirical mode decomposition. In *IEEE international conference, acoustics, speech, and signal processing, 2005. Proceedings. (ICASSP'05)* (Vol. 4, p. iv–485).

5. Donnelly, D. (2006). The fast Fourier and Hilbert-Huang transforms: a comparison. *Computational Engineering in Systems Applications*, *1*, 84–88.

6. Du, Q., & Yang, S. (2007). Application of the EMD method in the vibration analysis of ball bearings. *Mechanical Systems and Signal Processing*, *21*, 2634–2644.

7. Duffy, D.G. (2005). The application of Hilbert-Huang transforms to meteorological datasets. *Hilbert-Huang Transform and Its Applications* (pp. 129–147).

8. Ezer, T., Atkinson, L.P., Corlett, W.B., Blanco, J.L. (2013). Gulf Stream's induced sea level rise and variability along the US mid-Atlantic coast. *Journal of Geophysical Research: Oceans*, *118*, 685–697.

9. Hong, J., Kim, J., Ishikawa, H., Ma, Y. (2010). Surface layer similarity in the nocturnal boundary layer: the application of Hilbert-Huang transform. *Biogeosciences*, *7*, 1271–1278.

10. Huang, H., & Pan, J. (2006). Speech pitch determination based on Hilbert-Huang transform. *Signal Processing*, *86*, 792–803.

11. Huang, N.E., & Shen, S.S. (2005). *Hilbert-Huang transform and its applications*. Singapore: World Scientific.

12. Huang, N.E., & Wu, Z. (2008). A review on Hilbert-Huang transform: method and its applications to geophysical studies. *Reviews of Geophysics*, *46*.

13. Huang, N.E., Shen, Z., Long, S.R., Wu, M.C., Shih, H.H., Zheng, Q., Yen, N.-C., Tung, C.C., Liu, H.H. (1998). The empirical mode decomposition and the 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.

14. Huang, W., Shen, Z., Huang, N.E., Fung, Y.C. (1998). Engineering analysis of biological variables: an example of blood pressure over 1 day. *Proceedings of the National Academy of Sciences*, *95*, 4816–4821.

15. Huang, N.E., Wu, M.-L., Qu, W., Long, S.R., Shen, S.S.P. (2003). Applications of Hilbert–Huang transform to non-stationary financial time series analysis. *Applied Stochastic Models in Business and Industry*, *19*, 245–268.

16. Meignen, S., & Perrier, V. (2007). A new formulation for empirical mode decomposition based on constrained optimization. *Signal Processing Letters IEEE*, *14*, 932–935.

17. Pachori, R.B. (2008). Discrimination between ictal and seizure-free EEG signals using empirical mode decomposition. *Research Letters in Signal Processing*, *14*.

18. Parzen, E. (1999). Stochastic processes. *SIAM*, *24*.

19. Peng, Z.K., Peter, W.T., Chu, F.L. (2005). A comparison study of improved Hilbert–Huang transform and wavelet transform: application to fault diagnosis for rolling bearing. *Mechanical Systems and Signal Processing*, *19*, 974–988.

20. Peel, M.C., McMahon, T.A., Pegram, G.G.S. (2009). Assessing the performance of rational spline-based empirical mode decomposition using a global annual precipitation dataset. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, *465*, 1919–1937.

21. Peel, M.C., McMahon, T.A., Srikanthan, R., Tan, K.S. (2011). Ensemble empirical mode decomposition: testing and objective automation. In *Proceedings of the 34th world congress of the international association for hydro-environment research and engineering: 33rd hydrology and water resources symposium and 10th conference on hydraulics in water engineering* (p. 702).

22. Pegram, G.G.S., Peel, M.C., McMahon, T.A. (2008). Empirical mode decomposition using rational splines: an application to rainfall time series. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, *464*, 1483–1501.

23. Phillips, S.C., Swain, M.T., Wiley, A.P., Essex, J.W., Edge, C.M. (2003). Reversible digitally filtered molecular dynamics. *The Journal of Physical Chemistry B*, *107*, 2098–2110.

24. Pigorini, A., Casali, A.G., Casarotto, S., Ferrarelli, F., Baselli, G., Mariotti, M., Massimini, M., Rosanova, M. (2011). Time–frequency spectral analysis of TMS-evoked EEG oscillations by means of Hilbert–Huang transform. *Journal of Neuroscience Methods*, *198*, 236–245.

25. Priestley, M.B. (1988). *Non-linear and non-stationary time series analysis*. London: Academic Press.

26. Rai, V.K., & Mohanty, A.R. (2007). Bearing fault diagnosis using FFT of intrinsic mode functions in Hilbert–Huang transform. *Mechanical Systems and Signal Processing*, *21*, 2607–2615.

27. Salisbury, J.I., & Wimbush, M. (2002). Using modern time series analysis techniques to predict ENSO events from the SOI time series. *Nonlinear Processes in Geophysics*, *9*, 341–345.

28. Schlurmann, T. (2002). Spectral analysis of nonlinear water waves based on the Hilbert-Huang transformation. *Transaction-American society of Mechanical Engineers Journal of Offshore Mechanics and Arctic Engineering*, *124*, 22–27.

29. Späth, H. (1995). *One dimensional spline interpolation algorithms*. Wellesley: AK Peters/CRC Press.

30. Torres, M.E., Colominas, M., Schlotthauer, G., Flandrin, P., et al. (2011). A complete ensemble empirical mode decomposition with adaptive noise. In *2011 IEEE international conference on acoustics, speech and signal processing (ICASSP)* (Vol. 2011, pp. 4144–4147).

31. Wu, Z., & Huang, N.E. (2009). Ensemble empirical mode decomposition: a noise-assisted data analysis method. *Advances in Adaptive Data Analysis, World Scientific*, *1*(1), 1–41.

32. Wu, F., & Qu, L. (2008). An improved method for restraining the end effect in empirical mode decomposition and its applications to the fault diagnosis of large rotating machinery. *Journal of Sound and Vibration*, *314*, 586–602.

33. Wahba, G., & Wang, Y. (1995). Behavior near zero of the distribution of GCV smoothing parameter estimates. *Statistics & Probability Letters*, *25*, 105–111.

34. Yang, C., Zhang, J., Fan, G., Huang, Z., Zhang, C. (2012). Time-frequency analysis of seismic response of a high steep hill with two side slopes when subjected to ground shaking by using HHT. In *Sustainable transportation systems: plan, design, build, manage, and maintain*.

35. Zhao, J.-P., & Huang, D.-J. (2001). Mirror extending and circular spline function for empirical mode decomposition method. *Journal of Zhejiang University Science*, *2*, 247–252.