The sensitivity of Empirical Mode Decomposition and its Applicationto Environmental Data

Thesis · May 2018

CITATIONS
0

READS
62

1 author:

Mona Bahri
UNSW Sydney
15 PUBLICATIONS **2** CITATIONS

SEE PROFILE

SENSITIVITY OF THE EMPIRICAL MODE DECOMPOSITION AND ITS APPLICATION TO ENVIRONMENTAL DATA

A thesis submitted for the degree of
Doctor of Philosophy

By
Fatemeh (Mona) Ziaeyan Bahri

School of Physical, Environmental and Mathematical Sciences,
The University of New South Wales,
Canberra.

UNSW SYDNEY | Australia's Global University

Thesis/Dissertation Sheet

Surname/Family Name	:	**Ziaeyan Bahri**
Given Name/s	:	**Fatemeh (Mona)**
Abbreviation for degree as give in the University calendar	:	**PhD**
Faculty	:	**UNSW in Canberra**
School	:	**School of Physical, Environmental and Mathematical Sciences**
Thesis Title	:	**The sensitivity of the Empirical Mode Decomposition and its application on Environmental data**

Abstract 350 words maximum: (PLEASE TYPE)

The Empirical Mode Decomposition (EMD) is a powerful data analysis method that can be used to extract embedded components within time series and other data. The method is advantageous in that it is able to analyse nonlinear and non-stationary data. Surprisingly, the literature exploring the ability of EMD to analyse data with different types of non-stationarity is relatively sparse. Also, the sensitivities associated with the critical initial steps of the EMD procedure are not well understood. One of the critical steps in determining each of the IMFs involves constructing upper and lower envelopes of the local maxima and minima of the time series. In the original presentation of the EMD methodology, cubic-spline interpolation was used to construct these envelopes. However, there is no *a priori* reason to support the use of cubic splines, and it is natural to wonder how employing alternative interpolation methods might affect the ultimate outcome of the EMD method.

This dissertation is dedicated to providing a more comprehensive understanding of the sensitivity of EMD to different interpolation methodologies, and to different types of data non-stationarity. These sensitivities are investigated systematically using synthetic time series data that cover a range of interesting features. In addition, a number of environmental data sets are used to explore how EMD sensitivity can influence the inferences that might be drawn from modes extracted using various forms of the EMD method. The analyses conducted in this thesis suggest that the robustness and accuracy of EMD is improved when smoothing-spline interpolation is employed as its underlying interpolation method. In particular, Smoothing Ensemble EMD (SEEMD) is introduced, and is shown to out-perform other EMD methods in a number of important contexts. Specifically, SEEMD is shown to be more robust in the presence of noise and is able to extract more meaningful features from the environmental times series. It is also found, however, that all EMD-based methods perform poorly when confronted with time series exhibiting abrupt non-stationarity.

FOR OFFICE USE ONLY Date of completion of requirements for Award:

COPYRIGHT STATEMENT

'I hereby grant the University of New South Wales or its agents the right to archive and to make available my thesis or dissertation in whole or part in the University libraries in all forms of media, now or here after known, subject to the provisions of the Copyright Act 1968. I retain all proprietary rights, such as patent rights. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.
I also authorise University Microfilms to use the 350 word abstract of my thesis in Dissertation Abstract International (this is applicable to doctoral theses only).
I have either used no substantial portions of copyright material in my thesis or I have obtained permission to use copyright material; where permission has not been granted I have applied/will apply for a partial restriction of the digital copy of my thesis or dissertation.'

Signed .....

Date .....4.12.2018.....

AUTHENTICITY STATEMENT

'I certify that the Library deposit digital copy is a direct equivalent of the final officially approved version of my thesis. No emendation of content has occurred and if there are any minor variations in formatting, they are the result of the conversion to digital format.'

Signed .....

Date .....4.12.2018.....

ORIGINALITY STATEMENT

'I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, or substantial proportions of material which have been accepted for the award of any other degree or diploma at UNSW or any other educational institution, except where due acknowledgement is made in the thesis. Any contribution made to the research by others, with whom I have worked at UNSW or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception or in style, presentation and linguistic expression is acknowledged.'

Signed .....

Date .....4/12/2018.....

I hereby declare that this submission is my own work and to the best of my knowledge it contains no material previously published or written by another person, nor material which to a substantial extent has been accepted for the award of any other degree or diploma at UNSW or any other educational institution, except where due acknowledgement is made in the thesis. Any contribution made to the research by colleagues, with whom I have worked at UNSW or elsewhere, during my candidature, is fully acknowledged.

I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception or in style, presentation and linguistic expression is acknowledged.

Fatemeh (Mona) Ziaeyan Bahri

Abstract

The Empirical Mode Decomposition (EMD) is a powerful data analysis method that can be used to extract embedded components within time series and other data. In the context of EMD, these embedded components are called Intrinsic Mode Functions (IMF). Since its inception in the late 1990s, EMD has been applied in a number of areas, including biomedicine, neuroscience, epidemiology, chemical engineering, finance, atmospheric turbulence, seismology and oceanography. The method is advantageous in that it is able to analyse nonlinear and non-stationary data. Surprisingly, the literature exploring the ability of EMD to analyse data with different types of non-stationarity is relatively sparse. Also, the sensitivities associated with the critical initial steps of the EMD procedure are not well understood. One of the critical steps in determining each of the IMFs involves constructing upper and lower envelopes of the local maxima and minima of the time series. In the original presentation of the EMD methodology, cubic-spline interpolation was used to construct these envelopes. However, there is no *a priori* reason to support the use of cubic splines, and it is natural to wonder how employing alternative interpolation methods might affect the ultimate outcome of the EMD method.

This dissertation is dedicated to providing a more comprehensive understanding of the sensitivity of EMD to different interpolation methodologies, and to different types of data non-stationarity. These sensitivities are investigated systematically using synthetic time series data that cover a range of interesting features. In addition, a number of environmental data sets are used to explore how EMD sensitivity can influence the inferences that might be drawn from IMFs extracted using various forms of the EMD method. In this part of the study, temperature, sea level and forest fire danger rating time series are chosen. There is significant interest in the way these particular variables might be changing over various time scales, and how these changes might impact various ecosystems, including human societies. The analyses conducted in this thesis suggest that the robustness and accuracy of EMD is improved when smoothing-spline interpolation is employed as its underlying

interpolation method. In particular, Smoothing Ensemble EMD (SEEMD) is introduced, and is shown to out-perform other EMD methods in a number of important contexts. Specifically, SEEMD is shown to be more robust in the presence of noise and is able to extract more meaningful features from the environmental times series. It is also found, however, that all EMD-based methods perform poorly when confronted with time series exhibiting abrupt non-stationarity.

"To my Mom who taught me to always be a brave, honest, and independent woman."

Acknowledgements

I want to first thank my supervisor, Jason Sharples, who supported me through this journey. His honesty in science and research made him a great role model for me. This work would not be possible without his encouragement, advice, and support.

I also want to give a special thanks to my beloved partner, Nick Robins. You have always supported and encouraged me. Also to my lovely cat, Arad, for being such a joy in my life.

I want to thank Dominique Kazan and Luice for their kindness and encouragement. They taught me to not give up and no matter what happens, keep swimming. A special thanks to Peter McIntyre for all his assistance all through my PhD journey.

To my family and friends, Nastaran Nemati, Sanaz Farhadi, Billie Genandran, John Pye who have always been with me through good and bad.

To my supervisor Xiao (Hua) Wang who always made sure to keep me on track. And to the people from the School of PEMS who always made my days with beautiful smiles and positive attitudes.

Publications

The following publications correspond to work within this thesis:

- **F.M. Ziaeyan Bahri**, J.J. Sharples, *Sensitivity of the Hilbert-Huang Transform to interpolation methodology*, MODSIM2015 Conference, Gold Coast. *http://www.mssanz.org.au/modsim2015/F8/ziaeyanbahri.pdf*

- **F.M. Ziaeyan Bahri**, J.J. Sharples , X. H. Wang, Y-J. Sun, *Modelling Sydney Sea Level and East Australian Current co-variability using Hilbert Huang Transform* , MODSIM2015 Conference, Gold Coast.
http://www.mssanz.org.au/modsim2015/F2/ziaeyanbahri.pdf

- **F.M. Ziaeyan Bahri**, J.J. Sharples, *Empirical Mode Decomposition and its frequency resolution*, Submitted to MODSIM2017 Conference, Hobart

- **F.M. Ziaeyan Bahri**, J.J. Sharples , *Decadal and multi-decadal variability in fire danger and its contributors* Submitted to MODSIM2017 Conference, Hobart

- **F.M. Ziaeyan Bahri**, J.J. Sharples , *Sensitivity of the Empirical Mode Decomposition to Interpolation Methodology and Data Non-stationarity,* Ready to submit to Environmental Modeling and Software Journal

- **F.M. Ziaeyan Bahri**, J.J. Sharples , *The Empirical Mode Decomposition Method and its Application on Temperature and Sea-Level Data,* In preparation

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Context of this study

In order to understand the physical world we live in better, it is necessary to obtain quality data and to interpret them correctly. Having quality data means having appropriate measuring instruments and knowing their limitations. Another challenge is interpreting the data. This is usually achieved by using a method appropriate to the data. To gain reliable interpretations of the data, it is necessary to know the limitations and sensitivities of the chosen method.

Analytical methods usually decompose the data (time series or some other signal) into its primitive or fundamental components (modes). The use of these fundamental components to describe the original data arises in a number of applications. For example, in electronics signal processing and in certain geophysical applications, insights can be gained by decomposing a complex signal into components with distinct frequency characteristics, which can then be analysed to identify particular driving mechanisms. The Fourier Transform is the pervasive example in this respect, though other examples such as Principal Components Analysis and Wavelet Analysis are also widely employed. While methods like the Fourier Transform are key theoretical tools in signal decomposition, they suffer from certain shortcomings, such as being poorly suited to the analysis of nonlinear and non-stationary data.

Unlike global methods such as the Fourier Transform, the Empirical Mode Decomposition (EMD) procedure, [Huang et al., 1998a] expresses the input time series in terms of an *a posteriori* defined basis, which is derived from the data themselves. When combined with the Hilbert transform [Hahn, 1996], which delivers information about the frequency/power spectra of the intrinsic modes, EMD constitutes a useful data-analysis technique that can be applied to nonlinear and non-stationary time series. The combination of EMD and Hilbert spectral analysis is referred to collectively as the Hilbert-Huang Transform (HHT).

The HHT provides a method more suitable for non-stationary and nonlinear time series than the traditional methods such as Fourier Analysis [Rai and Mohanty, 2007, Peng et al., 2005, Du and Yang, 2007, Donnelly, 2006]. This method has been used in a number of areas in which non-stationary and nonlinear data naturally arise such as: biomedicine [Huang et al., 1998b, Pachori, 2008]; neuroscience [Pigorini et al., 2011]; chemical engineering [Phillips et al., 2003]; finance [Huang et al., 2003]; atmospheric science [Hong et al., 2010, Salisbury and Wimbush, 2002]; seismology [Changwei et al., 2012]; and ocean dynamics [Schlurmann, 2002, Ezer et al., 2013]. Given the broad application of EMD/HHT, 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 [Deering and Kaiser, 2005, Huang and Wu, 2008, Torres et al., 2011], end effects [Jin-Ping, 2001, Dätig and Schlurmann, 2004, Wu and Qu, 2008] and sensitivity to interpolation method [Pegram et al., 2008, Peel et al., 2009]. Despite all these studies, the mathematical theory behind the EMD/HHT method is still not well understood and deserves further consideration.

1.2 Aim and scope

Non-stationary and nonlinear data are two main sources of difficulty in data analysis. Time-series data are non-stationary if the mean, variance or both vary with time [Priestley, 1988, Parzen, 1999]. As discussed previously, EMD was designed to facilitate spectral analysis of non-stationary and nonlinear data Huang et al. [1998a]. 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 in 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. Transformation of the IMFs using the Hilbert Transform (HT) then delivers spectral information relating to the data [Huang et al., 1998a].

In the original presentation of the EMD method, cubic-spline interpolation was employed to construct the upper and lower envelopes [Huang et al., 1998a]. 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 splines, such as overshooting and undershooting, can limit the accuracy of the overall EMD process [Pegram et al., 2008]. 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 [Duffy, 2005, Chen et al., 2006, Peel et al., 2009].

In addition to the methodological limitations and sensitivities of EMD discussed above, there are also questions about how the nature of the nonlinearity and non-stationarity of the time series might affect the efficacy of EMD. Environmental time series, like those to which EMD has been applied, can possess a variety of non-stationary and nonlinear characteristics, and so it is important to understand the sensitivity of EMD to these aspects of the data. For example, Duffy [2005] found that EMD performed well for periodic signals but could perform poorly for nonlinear aperiodic time series, such as most meteorological datasets. The sensitivity of EMD to data non-stationarity has received little attention in the literature, particularly for cases in which the non-stationarity arises due to time-dependent variance. Peel et al. [2011] investigated the application of rational-spline EEMD to synthetic time series with a number of trend patterns, while this appears to be one of the limitted studies that considered time-dependent variance, Huang and Pan [2006] applied EMD to speech signals with time-dependent variance.

In this dissertation, the sensitivity of the EMD procedure to the interpolation method and data non-stationarity is investigated using sets of synthetic data constructed to cover the different types of non-stationarity, and specifically the possible ways that environmental data can be non-stationary. For example, with global warming, the mean and the variance of environmental data may be changing over time. Using the synthetic data assists us to systematically investigate the combined sensitivity of EMD to data non-stationarity and the interpolation method used. EMD methods based on three different interpolation methods, piecewise-linear interpolating splines, cubic interpolating splines and smoothing splines, are applied to time series with time-dependent mean and/or time-dependent variance, and their ability to reproduce the known signal components as IMFs assessed.

Then EMD is applied to environmental data (temperature and sea level) to determine its sensitivity in these cases to the different interpolation methods and the data non-stationarity. Smoothing-spline interpolation is found to give the most robust and accurate results, and its use is suggested as an improvement to the EMD method. Finally, EMD with smoothing-spline interpolation is applied to data on the forest fire danger index to investigate the long-term variability in fire danger in southeast Australia.

1.3 Significance of the study

In this dissertation the sensitivity of the EMD method to data non-stationarity is examined. To do this, 20 synthetic datasets based on different types of non-stationarity are constructed. These synthetic data are specifically constructed to resemble possible patterns in environmental data. Then the EMD method is used to analyse these data. The results show a clear sensitivity of the method to abrupt changes in the mean or variance of the data. This will assist scientists to know for which type of data they can use EMD as an analysis tool.

As mentioned in the previous section, interpolation is used in a critical step of the EMD method to find the upper and lower envelopes; the interpolation is used multiple times during the EMD process. Therefore, if there are any problems in the interpolation or if the interpolation method used is not suitable, then the errors are propagated and magnified through the whole process. This makes understanding the sensitivity of the EMD to interpolation method important. In this thesis, it is shown for which types of data the method is more likely to be sensitive to the interpolation method and which interpolation method most suitable.

Three different interpolation methods are used to show the sensitivity of the EMD to the interpolation method and determine if there is a way to reduce this sensitivity. The results in this study show that more sophisticated interpolation methods such as cubic spline and smoothing spline produce better results. Although cubic-spline interpolation is not the worst method to use in EMD, smoothing-spline interpolation reduces the end effects and mode-mixing issues and produces more robust results. Therefore the author suggests that using smoothing splines is more suitable for EMD analysis.

Although using smoothing splines improves the accuracy and robustness of the results, it adds a degree of freedom in the model: a smoothing parameter must be specified. Choosing a suitable smoothing parameter can be challenging. However in this thesis, the author shows that it is possible to find a range of smoothing parameters that give good results. Further study needs to be done on this matter.

The two better interpolation methods (cubic-spline and smoothing-spline interpolation) are then applied to real environmental data (temperature and sea level) to examine the impact of the EMD sensitivity on real-world data and whether these sensitivities can affect the interpretation of the results. The results indicate that, although the EMD results with both interpolation methods describe the general behaviour of temperature and sea level, the smoothing-spline results show more detail that matches the actual behaviour of temperature and sea level.

The improved smoothing-spline-based EMD is then used to investigate the decadal and multi-decadal variability in fire danger and its contributors in south-east Australia. This is the first study to examine the fire danger variability at these time scales. The study indicates that the main drivers of long-term variability can be defined over different time scales.

In summary, this thesis systematically examines the EMD method and enhances our knowledge of the sensitivity and limitations of the method when applied to environmental data. It also introduces a new variant of EMD which is shown to be more accurate, more robust and more able to produce physically meaningful results.

1.4 Structure of the thesis

This thesis has eight chapters, with the first chapter an introduction. Chapter 2 gives the background to and a brief overview of current data-analysis tools.

In Chapter 3, the EMD, its limitations and recent improvements are presented in detail.

In Chapter 4, the sensitivity of the EMD to the interpolation method used and to data non-stationarity is investigated using synthetic data.

In Chapters 5 and 6, the effects of using two of the interpolations on temperature and sea-level data are examined, and the results are compared with the literature.

In Chapter 7, the best interpolation method is used to analyse the forest fire danger index data and make predictions about long-term trends.

Chapter 8 presents the main conclusion of this thesis and highlights avenues for future work.

Chapter 2

Background

2.1 Introduction

The need to decompose a time series, or some other signal, into its primitive or fundamental components arises in a number of applications. The Fourier Transform is the pervasive example in this respect, but it and other commonly used methods are poorly suited to the analysis of nonlinear and non-stationary data, typical of environmental time series.

For environmental decision makers, having reliable information about the environment and its behaviour is essential. Reliable and useful information can only result from accurate analyses of the environmental data.

2.2 Characteristics of environmental data

Most environmental data are nonlinear and non-stationary. Data are nonlinear when they come from a non-linear structure, in which its elements are not organized in a sequential fashion. A data item in a nonlinear data structure could be attached to several other data elements to reflect a special relationship among them, and all the data items cannot be traversed in a single run.

Non-stationary data are data in which the mean, variance and/or covariance change over time [Priestley, 1988]. Non-stationary behaviour can include trends, cycles, white noise or a combination of all three. Later in this thesis, we give examples of nonlinear data, non-stationary data and combinations of these.

Unfortunately there are not many methods that can analyse both non-stationary and nonlinear data without first transforming them into stationary data or making linear assumptions. We look first at some of these methods and their limitations in this chapter.

2.3 Overview of analytical methods

Decomposition of a stochastic time series into its basis modes is common in many areas of application, and different methods exist to do this [Holmes et al., 1998]. The aim of all of these methods is to extract dominant features and trends from the data. Visser et al. [2015] reviewed thirty methods used to find the (long-term) trend in one specific set of environmental data: sea level at San Francisco over the period 1855–2014. They specifically compared the second derivative of the trends (sea-level acceleration) using these methods. They showed that different methods can lead to contradictory acceleration-deceleration inferences, and suggested that the reason for these differences is uncertainties in parameters in the models, as well as different model structures. They divided the thirty methods into five main categories, listed in Table 2.1.

The members of the first group, exploratory data analysis, are mainly simple low-pass filters, moving-average or linear-filter methods. This group has several problems, such as they are not based on a statistical formulation and thus provide no uncertainty bounds for trend estimates. In addition, they depend on the choice of a specific window size, which results in missing trend estimates at the end of the sample period.

The second group is parametric trend estimation. In these methods a linear or second-order polynomial trend is fitted to the data. Therefore, the shape of the trend is fixed or deterministic. One of the main drawbacks of this group of methods is finding a reliable uncertainty.

The third group is called non-parametric models. In this group the shape of the trend is not defined in advance and is data driven. These models do not assume fixed trend shapes. The main issue with these methods is that their interpretation is more complex. They also have the same issue as the second group with their uncertainty.

The fourth group is stochastic models, with the trend assumed to originate from the underlying noise processes. The idea is that the data are a realisation of stochastic processes, and that trends in the data should be treated accordingly. These methods have the underlying assumption that the noise in the data is white noise. This might not be the case in all situations. Visser et al. [2015] considered the EMD, the subject of this thesis, to be in the fifth, miscellaneous, group but they did not study it.

Some of the traditional time series analysis methods are now explained in more detail.

2.3.1 Fourier transforms

One of the most common data-analysis methods is the Fourier-Transform method. This method is a linear transform that converts a time series to the sinusoidal frequencies that make it up [Bracewell, 1965]. The Fourier Transform $Y(t)$ of time-domain data are of the form

$$Y(t) = \sum_{j=0}^{N-1} a_j \exp(i\omega_j t),$$

where the a_j are amplitudes derived from the original data and ω_j the frequencies.

This method is a useful technique for solving problems in linear systems [Bracewell, 1965]. However, it is also based on an assumption of stationary data, and therefore cannot analyse non-stationary terms, which often carry important information [Hlawatsch and Auger, 2013].

If the data are not linear and stationary, the output may not have the usual physical meaning of frequencies. Kolláth and Oláh [2009] investigated solar magnetic activity, which cannot be treated as a stationary periodic process. They used a finite Fourier Transform, Wavelet Analysis and generalised time-frequency distributions. Their results showed that, although the finite Fourier Transform was more powerful than Wavelet Analysis, it was not as useful as Time-Frequency Distribution Analysis. Thayaparan and Yasotharan [2000] used non-stationary radar signals produced by accelerating targets. They applied the Fourier method to detect the moving object and estimate its velocity. Their results showed that the Fourier method was not successful in detecting the nonlinear components of the accelerating target. They suggested that a more sophisticated method was required.

2.3.2 Fourier-based methods

The **Short-Time Fourier Transform** is a way to overcome the non-stationary problem in Fourier transforms. The idea is to break the non-stationary data into sections, then assume the data in each section are stationary and use the normal Fourier Transform [Okumura, 2011]. The size of each section is chosen by the user. Choice of the section size is critical; for example, a very small section size will lose the low-frequency modes, a very large section size will lose the high-frequency modes. This method also cannot be used with non-stationary data [Okumura, 2011].

Table 2.1: Categorisation of analytical methods into five groups from Visser et al. [2015]. All references were studying sea level.

Group	Methods	References
Exploratory data analysis	Moving/median averages	Houston and Dean [2013], Douglas [1991]
	Low-pass filter method	White et al. [2014], Woodworth et al. [2009]
	Test for monotonic trends	Chandler and Scott [2011]
	Expert judgment or visual inspection	[Metz, 2001]
Parametric trend estimation	Linear/piecewise-linear trend	Church and White [2011], Olivieri and Spada [2013]
	Linear with cycle/trend; cycle based on non-linear optimisation	Chambers et al. [2012]
	Second-order polynomial trend	Woodworth et al. [2011]
	Second-order trend and cycle based on nonlinear optimisation	Houston and Dean [2011]
	Multiple regression model	Zhang and Church [2012]
	Exponential-growth curve fitting	Parker [2013]
	Non-stationary extreme-value trend models	Mudelsee [2013]
Nonparametric trend estimation	Generalised additive models (GAMs)	White et al. [2014]
	Lowess-smoothing nonparametric regression technique	Chandler and Scott [2011]
	Quantile regression	Chandler and Scott [2011]
	Wavelets	Chandler and Scott [2011]
Stochastic trend models	Auto-regressive integrated moving average models (ARIMA)	Mudelsee [2013]
	Structural time series models (STM)	Chandler and Scott [2011]
	Long-memory models (ARFIMA)	Mudelsee [2013]
Miscellaneous models	Linear with shifting windows	Holgate [2007]
	Quadratic with a variety of windows	Jevrejeva et al. [2008]
	MSDA diagrams and second-order polynomial trends	Scafetta [2014]
	Rate differences (the differences between two linear trend slopes)	Sallenger Jr et al. [2012]
	Empirical mode decomposition	Ezer et al. [2013]
	Singular spectrum analysis	Jevrejeva et al. [2006]
	Multivariate adaptive regression splines (MARS)	Feng et al. [2004]
	Artificial neural networks (ANN)	Imani et al. [2014]
	Single and multitaper spectral analysis	Mudelsee [2013]

The **Spectrogram** is a Fourier Transform with a restricted time window [Bastiaans, 1981]. It is easy to implement as a Fast Fourier Transform. It assumes that the data are piecewise stationary; this is not always the case even with small time windows [Bastiaans, 1981].

The **Wigner-Ville distribution** is the Fourier transform of the central covariance function $C_c(\tau, t)$,

$$V(\omega, t) = \int_{-\infty}^{+\infty} \int C_c(\tau, t) e^{-\omega\tau}\, d\tau,$$

where $C_c(\tau, t) = D\left(t - \frac{1}{2}\tau\right) D^*\left(\frac{1}{2}\tau\right)$ and $D(t)$ are the data. This method is powerful and covers a variety of problems, although this windowed Fourier analysis suffers from all the Fourier limitations and requires a lot of judgment to apply it to complicated data.

The **Evolutionary Spectrum** method uses a family of orthogonal basis functions $\Phi(\omega, t)$ instead of the sine and cosine functions in the Fourier Transform.

$$X(t) = \int_{-\infty}^{+\infty} \Phi(\omega, t)\, dA(\omega, t),$$

where $\Phi(\omega, t)$ is the basis function. The difficulty is to define the basis functions $\Phi(\omega, t)$. The basis is a *posteriori*, and no systematic way has been found.

2.3.3 Wavelet Analysis

This method shows the similarity between a signal and the predetermined wavelet at scale a at time b. To cover all high and low frequencies, wavelets of different size (frequency) scales are used:

$$W(a, b; f, \psi) = |a|^{-\frac{1}{2}} \int_{-\infty}^{+\infty} f(t) \Psi^* \left(\frac{t - b}{a}\right) dt. \tag{2.1}$$

Here, a is a dilation factor, b is a translation of the origin (temporal location), Ψ is the basis wavelet and $1/a$ is a frequency. W is the energy of f of scale a at $t = b$.

Advantages of using this method are: the basic wavelet Ψ can be modified according to need; and it is very useful in analysing data with a slowly changing frequency. However, we have to define the Ψ in advance for the whole calculation. It is also a linear analysis, and hence has all the inherent problems of linear analysis. It is limited by the wavelet function, has poor resolution, a counter-intuitive interpretation, a non-adaptive nature, and the physical meaning is the same as for linear phenomena [Huang et al., 1998a, Duffy, 2005].

2.3.4 The Empirical Orthogonal Function expansion

The Empirical Orthogonal Function (EOF) expansion is also known as Principal-Components Analysis, Singular-Value Decomposition or Singular Spectrum Analysis. For a scalar variable, the EOF method finds spatial patterns of variability and their time variation, and gives a measure of the importance of each pattern. This method is an alternative approach to analysing a non-stationary and nonlinear time series. To define the basis functions, some prior determination has to be applied. In the EOF method, an orthogonality assumption is also made [Ghil et al., 2002]. This method is sensitive to the selection of the temporal and spatial domains.

This method is basically a decomposition of a signal or dataset in terms of orthogonal basis functions (eigenvectors), which are different from the data. It finds both time and spatial variations; the modes are directly extracted from the data.

However, the method only gives frequencies independent of time, and so cannot be used for non-stationary data [Huang et al., 1998a, Duffy, 2005]. Furthermore, the eigenvectors are not necessarily linear or stationary, and so are difficult to analyse. Moreover, the distribution of eigenvalues does not yield characteristic time or frequency scales [Duffy, 2005].

None of the above methods is suitable to analyse nonlinear/non-stationary data. The Empirical Mode Decomposition method, which was pioneered in the late 1990s, is known to deal with nonlinearity and non-stationarity in the data. In the next chapter, this method is explained in more detail.

Chapter 3

Empirical Mode Decomposition/Hilbert-Huang Transform Method

The Empirical Mode Decomposition / Hilbert-Huang Transform, proposed by Huang et al. [1998a], consists of two main steps: a decomposition algorithm, the empirical mode decomposition (EMD); and a spectral analysis tool called the Hilbert-Huang Transform (HHT).

3.1 Empirical Mode Decomposition

The Empirical Mode Decomposition decomposes a dataset into its intrinsic mode functions (IMFs) [Huang and Wu, 2008]. For a time series $y(t)$, the IMFs are determined through the EMD process according to the algorithm below.

Algorithm Empirical Mode Decomposition

1. Set *j=1*.
2. Identify all the local extrema in the signal $y(t)$.
3. Interpolate the maxima and minima to define upper and lower envelopes of $y(t)$.
4. Calculate the mean of the upper and lower envelopes $m_j(t)$.
5. Protomode $h_j(t) = y(t) - m_j(t)$.
   If $j < 3$ **Then** $y(t) = h_j(t)j + 1 \rightarrow j$ **Go to 2**
6. **If** $h_j = h_{j-1} = h_{j-2}{}^*$ **Then** IMF 1 is $c_1(t) = h_j(t)$ **Go to 7**
   Else $y(t) = h_j(t)$ **Go to 2**.
7. Find the residual time series $r_1(t) = y(t) - c_1(t)$.
8. For the next IMF $c_2(t)$, set $y(t) = r_1(t)$ **Go to 1**. 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.

* within the specified tolerance. Steps 2–4 in the EMD method are quite important.

These steps are illustrated in Figure 3.1. In this method, given the input dataset (Figure 3.1a), first the local extrema are found (Figure 3.1b); then the upper and lower envelopes are plotted using an interpolation method (Figure 3.1c). The mean of the two envelopes $m_1(t)$ is then calculated (Figure 3.1d), and finally, the difference between the mean and the data is calculated:

$$h_1(t) = y(t) - m_1(t). \tag{3.1}$$

Setting the new $y(t) = h_1(t)$, the same steps are repeated to give $h_2, h_3, \ldots$ until

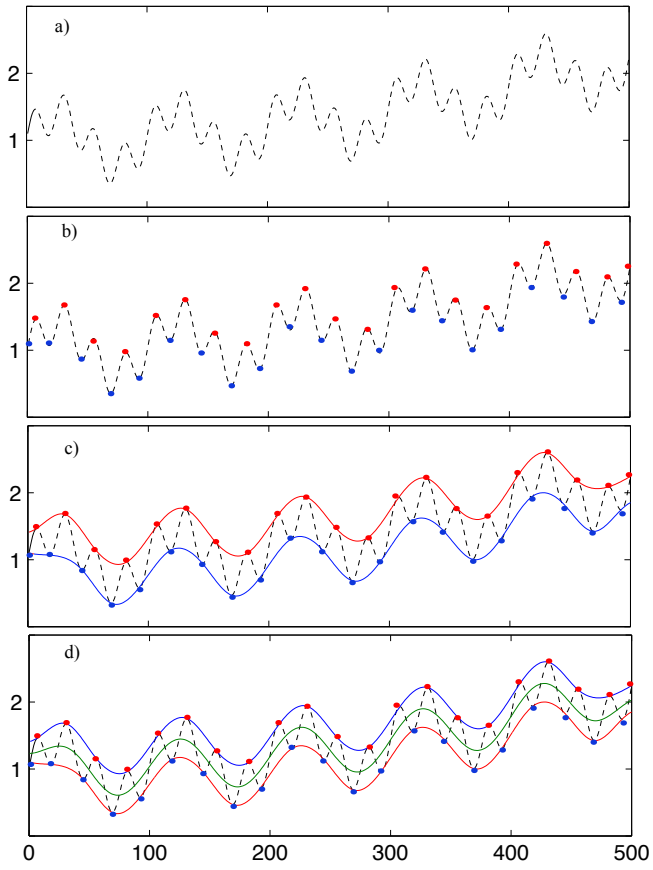

Figure 3.1: Empirical Mode Decomposition upper and lower envelopes.

$h_j = h_{j-1} = h_{j-2}$ to within some given tolerance. This often is called the 'stopping criterion', according to Huang et al. [2003]. This stopping criterion determines the number of sifting steps to produce an IMF. Huang et al. [2003] mentioned that the stopping criterion provides a symmetric IMF and eliminates riding waves. Huang proposed two possibilities for choosing the stopping criteria. First, the Cauchy-type convergence criterion which is explained in detail in Huang et al. 1998a, here

the sifting is carried to the point when the difference between the successive sifted results is smaller than a preset limit. Huang.2003 mentioned the major flaw of this approach is that it is unrelated to the definition of the IMF; it only requires the successive IMFs to be approximately equal. Second, where the sifting is stopped when the number of zero crossings and extrema is the same number for successive sifting steps (S) Huang et al. 1999. Huang et al. 1999 mentioned S can not be too large as over-sifting can have a detrimental effect on the results. Huang et al. 1999 suggested a value of $3 < S < 5$ as the default stopping criterion. Rilling et al. 2003 discussed how the amplitude of the mean in comparison with the amplitude of the corresponding mode is critical, as it can be led to under- or over-iteration which could under- or over-decompose the dataset. Rilling et al. 2003 proposed a new criterion based on two thresholds which, aimed at guaranteeing globally small fluctuations in the mean while taking into account locally large excursions. Rato et al. 2008 emphasized that the stopping criterion is another source of problems with EMD due to its degree of arbitrariness, since it may not guarantee a total signal removal to obtain a true IMF. They defined a resolution factor by the ratio between the energy of the signal at the beginning of the sifting, and the energy of average of the envelopes. If this ratio grows above the allowed resolution, then the IMF computation must stop. In this study, we used the stopping criteria introduced by Huang.1999. However, as mentioned earlier, reaching absolute equality (in the equation $h_j = h_{j-1} = h_{j-2}$) is implausible, and some tolerance is applied; in this study, when the differences between h_j, h_{j-1} and h_{j-2} are less than 10^{-3}.

If $h_j(t)$ is the last component, then $c_1(t) = h_j(t)$ is the first IMF.

In the resulting IMF:

- the mean $m_j(t)$ should be identically zero to within some tolerance;
- the number of zero-crossings and extrema of the IMF should either be equal or differ by one.

The IMF is then subtracted from the original data to give the residual. This residual contains information on the longer-period components. The residual is then used as the new dataset, and the process repeated to calculate successive IMFs.

EMD is stopped when the residual is a monotonic; this monotonic residual is the (long-term) trend in the original data. The whole process in the algorithm is called sifting. Huang et al. [1999] mentioned that to guarantee physical meaning, the sifting process needs to be limited, otherwise we will end up with a constant-amplitude component and additional, unphysical modes.

The data have then been decomposed as:

$$y(t) = \sum_{j=1}^{n-1} c_j(t) + r_n(t), \tag{3.2}$$

where the $c_j(t)$ are the IMFs, n is a total number of IMFs and $r_n(t)$ is the final residual [Huang and Wu, 2008].

3.2 Hilbert-Huang Transform

By applying the Hilbert transform to individual IMFs, we obtain a full energy-frequency-time distribution which permits further physical interpretation of the data.

The Hilbert Transform $H_j(t)$, now applied to each IMF $c_j(t)$, is given by

$$H_j(t) = \frac{1}{\pi} P \int \frac{c_j(t')}{t - t'} \, dt', \tag{3.3}$$

where P indicates the Cauchy Principal Value. An analytical function Z_j is defined for IMF$_j$ is

$$Z_j(t) = c_j(t) + iH_j(t) = A_j(t) \exp(i\theta_j(t)), \tag{3.4}$$

where $A_j(t)$ and $\theta_j(t)$ are the instantaneous amplitude and phase of IMF$_j$, respectively, with

$$A_j = (c_j^2 + H_j^2)^{1/2} \qquad \theta_j(t) = \tan^{-1}\left(\frac{H_j}{c_j}\right). \tag{3.5}$$

The instantaneous frequency of IMF$_j$, ω_j is given by:

$$\omega_j(t) = \frac{d\theta_j(t)}{dt}. \tag{3.6}$$

A limitation of this transform is that it cannot be used on all datasets [Huang et al., 1998a]. The data must be mono-component, and therefore need to be sifted and decomposed first. By first using EMD, we separate the data into IMFs based on their time scale. Then, using a Hilbert Transform on each IMF will lead us to a physical interpretation of the data. IMFs and their energy distribution are crucial in analysing and interpreting the data.

3.3 Comparison of classical time-series analyses with the EMD/HHT method

In this section, the studies that compare the EMD/HHT method with other methods are reviewed. The aim of all of these methods is to analyse experimental data with a view to extracting dominant features and trends: coherent structures. Although all of these methods deal with nonlinear problems, other than EMD/HHT, they are linear procedures.

One of the problems with the classical methods is that sometimes, although the results are mathematically meaningful, they do not give physically meaningful results. Huang et al. [1998a] and Duffy [2005] showed that, due to the nature of the Fourier Transform, it sets the frequency globally for the whole dataset. The frequency spectrum cannot represent the true energy density in frequency space. Thus, while the Fourier Transform can deliver mathematical representations of the frequency spectrum, it may have no relation to physical reality. Moreover, Fourier spectral analysis uses harmonic components to decompose the wave profile but sometimes the deformation of the wave profile is a cause of non-linearity. Therefore it is difficult to study non-stationary and nonlinear data without making linearity assumptions.

A more detailed example is presented by Donnelly [2006], who applied the Fast Fourier Transform (FFT) and Hilbert-Huang Transform (HHT) to a sum of two sine waves of different frequencies ($f_1 = 1, f_2 = 1.08$). The FFT returned the two original frequencies whereas the HHT generated only a single frequency, $f = 1.04$, the average of the two frequencies. Donnelly [2006] suggested that the reason that the HHT just gave one frequency is the small frequency difference between the two signals. Although the results of the FFT are in principle more accurate, two modes at frequencies so close together may not be physically meaningful; the HHT method in this case may provide a more meaningful result.

Wu et al. [2011] also demonstrated that Fourier-Transform results may not have physical meaning. They used two water-wave generators in a tank, and applied a Fourier Transform to predict the frequency of the generated waves. The Fourier Transform extracted ten components instead of two. They claimed this happened because orthogonality is a key constraint in Fourier analysis: orthogonality of each pair of Fourier components to some degree implies that each wave of a particular frequency is not correlated with any other wave, resulting in the misinterpretation [Wu et al., 2011].

Some studies claim that the EMD/HHT results are more effective and accurate than other methods. Schlurmann [2002] used a physical ocean model to produce two

different types of water-wave data, a monochromatic wave and a transient wave. He then applied a Fourier Transform and the Hilbert-Huang Transform to investigate the accuracy and reliability of the HHT method. The EMD/HHT method provided the more accurate and reliable results. Schlurmann [2002] also applied Wavelet Analysis and HHT to investigate the accuracy and reliability of the HHT method. Their results showed that the EMD/HHT method provided more accurate and more reliable results than Wavelet Analysis.

Pines and Salvino [2002] used a physical model in evaluating the performance of EMD in the detection of structural damage to a building. They shook the model using an hydraulic system, and recorded the acceleration before and after damage. They found the EMD/HHT results promising, as they allowed both the location and the amount of damage to be determined.

In another study, Zhang et al. [2003] used two different time series. First, they applied the EMD method to a time series $y(t)$ (non-stationary and nonlinear without trend), defined as $y(t) = y_1(t) + y_2(t)$, with $y_1(t) = \cos\big(2\pi t + 0.5\sin(2\pi t)\big)e^{-0.2t}$, a decaying wave, and $y_2(t) = 0.05\sin(30\pi t)$. Their results showed that the first IMF of the EMD picked up the noise in the data, the second picked up the signal; the rest of the IMFs, with negligible amplitude, represented the numerical error in the EMD process. Second, they applied EMD to the record of the 1994 Northridge (Los Angeles) earthquake, and compared their results with a Fourier-based analysis. Both methods produced similar component signals but the EMD results showed more detail of the earthquake rupture process. In addition, the EMD method was able to distinguish some related phenomena such as the predominant wave motion generated near the hypocenter, a large stress drop associated with the initiation of the earthquake and the components associated with longer-period signals as the rupture propagated. However, Zhang et al. [2003] did not say if all the EMD components were related to physical phenomena associated with the earthquake.

Vincent et al. [2010] found some limitations in the HHT method compared with the Fourier-Transform (FT) method. Their study showed that FT contains information up to $2\Delta t$ (20 minutes), whereas the Hilbert-Huang marginal spectrum required $4\Delta t$ (40 minutes) to resolve spectral information. This issue has been acknowledged by Huang et al. [1998a], who suggested that an interpolation method with more points between the available data would decrease the time step. They suggested that using oversampled data would also help.

Peng et al. [2005] compared their improved EMD/HHT method with Wavelet Analysis using three sets of synthetic data and a set of rolling-bearing data. Their results showed that: the improved EMD/HHT method had better resolution in time

and frequency than Wavelet Analysis; the improved EMD/HHT method had better computing efficiency than Wavelet Analysis, which means it is more suitable for large-scale signal analysis; and, the HHT spectrum often had a ripple phenomenon in its estimated frequencies, which may cause problems in the analysis.

Boudraa and Cexus [2007] applied wavelet analysis and EMD to synthetic data. They called their synthetic data: Doppler signals, which start with a high frequency, but the frequency decreases with time; Block and Bump signals, which are combinations of square steps and gentle peaks, respectively; and Heavy sine waves, which are low-frequency sine waves. Their results showed that, for some data such as their electrocardiogram and 'Heavy sine' data, the EMD results were more accurate, but for datasets such as their 'Doppler signals' and 'Bumps signals', Wavelet Analysis gave better results. Both EMD and wavelet methods did equally well in analysing their 'Block signals'.

Rai and Mohanty [2007] also investigated the effectiveness of the EMD/HHT method in bearing-fault diagnosis. To determine the characteristics of the frequency of defects in rolling-element bearings, they applied the fast Fourier Transform and the EMD/HHT process. Their results indicated that the EMD/HHT produced better results and worked well for bearing-fault diagnoses.

Vincent et al. [2010] applied the Fourier Transform (FT) and the Hanning window function to provide an estimate of the spectrum for wind-speed data. Although this method is an improvement on the standard Fourier-Transform analysis, it is still based on a stationary assumption [Madsen, 2007]. Both the FT and EMD/HHT results identified the diurnal cycle and both spectra had a similar shape. However, the Hilbert Transform had a smoother spectrum that resulted in better accuracy.

One of the significant differences between classical methods and EMD/HHT is that EMD uses intrinsic mode functions that often happen to be approximately orthogonal, rather than relying on orthogonal basis functions [Holmes et al., 1998]. Further, using a global domain in Fourier analysis and orthogonality assumes forever-repetitive datasets, which is not always the case in a physical system. Data observed in different time domains are not necessarily repetitive [Wu et al., 2011].

3.4 EMD limitations and improvements

In this section, some of the limitations and recent improvements of the EMD method are discussed in more detail.

3.4.1 EMD method limitations

Although the EMD/HHT is a useful tool for time-series analysis, it has some disadvantages as mentioned previously. Some of these limitations have been studied, such as endpoint effects, noise handling and mode-mixing problems, and the effects of different interpolation methods, and some improvements have been made. Some limitations are still unknown, such as the sensitivity of the method to the nature of the data. One of the big challenges in determining the EMD sensitivity is that the method, A its name suggests, EEMD is essentially an empirical method. While it does have a mathematical foundation, this is only algorithmic in nature, and does not provide for a particular set of basis functions that could be used to support more rigorous mathematical analysis, such as can be applied to Fourier series, for example; it is hard to check the accuracy of the results in general. Different studies have used different ways to assess the accuracy of EMD, for example by comparing EMD results with the results of other methods [Schlurmann, 2002], or reconstructing the data by plotting the sum of the modes together with the original data [Huang et al., 2003]. However, there is no guarantee about the accuracy of these results. All in all, because the method is relatively new, there is still a lot of scope for further research.

One of the most common limitations of the EMD method is mode mixing. Mode mixing occurs when a single IMF consists of signals of widely disparate time scales, or when signals of similar time scales reside in different IMF components. The presence of noise in the input signal is known to cause problems with mode mixing in the EMD method [Deering and Kaiser, 2005]. To overcome this problem, Huang and Wu [2008] introduced the Ensemble EMD (EEMD) procedure, which is a version of EMD specifically to deal with noise. Flandrin et al. [2004] found this method quite effective for analysing noisy data. However, they mentioned the lack of theory and applications of the method to real-world situations.

As discussed in Chapter 1, when EMD was presented for the first time, cubic-spline interpolation was used to interpolate between the local maxima and minima and find the upper and lower envelopes [Huang et al., 1998a]. Later Dätig and Schlurmann [2004] found that cubic-spline interpolation can cause overshooting and undershooting at the endpoints. This impacts the accuracy and overall performance of the EMD [Pegram et al., 2008]. Duffy [2005], Chen et al. [2006] and Peel et al. [2009] showed that the choice of interpolation in EMD is quite critical.

The interpolation method also plays a critical role at the beginning and end of the time series. 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. The way these additional points are defined will affect the IMFs that are

produced and, in the worst cases, can produce highly undesirable end effects [Dätig and Schlurmann, 2004]. Moreover, these end effects can propagate inwards and corrupt the whole sifting process, resulting in inaccurate IMFs.

In addition to the methodological limitations and sensitivities of EMD discussed above, there are also questions about how the nature of the nonlinearity and non-stationarity of the time series might affect the efficacy of EMD. Environmental time series, like those to which EMD has been applied, can possess a variety of non-stationary and nonlinear characteristics, and so it is important to understand the sensitivity of EMD to these aspects of the data. For example, Duffy [2005] found that EMD performed well for periodic signals but could perform poorly for nonlinear aperiodic time series, such as most meteorological datasets. The sensitivity of EMD to data non-stationarity has received little attention in the literature, particularly for cases in which the non-stationarity arises due to time-dependent variance. Peel et al. [2011] investigated the application of rational-spline EEMD to synthetic time series with a number of trend patterns, while Huang and Pan [2006] applied EMD to speech signals with time-dependent variance.

3.4.2 EMD method improvements

Improvements in interpolation

To improve the interpolation method used in the EMD method, several studies have been done to test other interpolation methods. Deléchelle et al. [2005] suggested an approach based on parabolic partial differential equations to find the mean envelope. This approach is based on the definition of the characteristic points of a function, the minima, maxima and inflection points, and the use of piecewise-cubic polynomial curves to interpolate successive characteristic points. Deléchelle et al. [2005] used a combination of two triangular waveforms as an input to test the method. The method decomposed the data into the IMFs expected. However, they did not test the method with complex input data.

Huang and Shen [2005] introduced a new spline method based on the B-spline interpolation [De Boor et al., 1978]; they called it the B-spline EMD method. The B-Spline EMD (BS-EMD) method represents the local mean of a signal as the moving average of the extrema, which is fitted using combinations of B-splines; this avoids the use of upper and lower envelopes. The advantage of not using envelopes is the elimination of one of the original EMD problems, generating the upper and lower envelopes. The use of the BS-EMD resulted in only a slight improvement because the basis of interpolation remained the same, cubic-spline interpolation.

Chen et al. [2006] made some improvements to the previous BS-EMD method. They found that the original EMD gave fewer modes than the moving-average methods. The EMD modes also had larger amplitudes and amplitude variations. However, the results from the EMD and moving-average methods were qualitatively and quantitatively similar. Furthermore, oversifting was still an issue [Huang et al., 1999]. Liu et al. [2006] applied the BS-EMD method to vibration signal analysis for localised gearbox fault diagnosis (data with no trend). They found that the method was more effective than Fourier and wavelet methods, and that it successfully detected gearbox faults. They also found that the BS-EMD method performed much better than the original EMD method.

Meignen and Perrier [2007] calculated the mean envelope as the solution of a quadratic programming problem with equality and inequality constraints; they did not calculate the upper and lower envelopes, one of the drawbacks of the EMD method. However, they used cubic splines as the interpolation method, with all its problems of overshooting and undershooting. Kopsinis and McLaughlin [2007] modified the existing EMD method by using the extrema of the original signal as fixed interpolation points, rather than recalculating the extrema at each iteration of the method, and found that this improvement reduced the total number of sifting steps in the EMD method.

Pegram et al. [2008] suggested the use of rational splines instead of cubic splines in the original EMD method. Rational splines are of the form [Späth, 1995]:

$$s_k(x) = A_k u + B_k t + \frac{C_k u^3}{1+pt} + \frac{D_k t^3}{1+pu},$$

where

$$t = \frac{x - x_k}{x_{k+1} - x_k},\ u = 1 - t.$$

The pole parameter p, controls the spline tension. When $p = -1$, $s_k(x)$ is a quadratic-spline interpolation; when $p = 0$, $s_k(x)$ is a cubic-spline interpolation; and for $p \to \infty$, a linear interpolation [Gregory and Sarfraz, 1990]. The ability to vary the pole parameter permits consideration of the interplay between spline tension and the resulting IMFs. The performance of a rational-spline EMD was tested using synthetic time series by Peel et al. [2011]. They used $0 < p < 5$, and found more sifting and more IMFs than with cubic splines. However, oversifting can lead to non-physically meaningful results [Huang et al., 1999]. Rational-spline EMD was found to perform well when applied to environmental time series [Pegram et al., 2008, Peel et al., 2009, 2011], although there remains some ambiguity surrounding the optimal choice of the tension parameter.

Some of the above studies have improved on cubic splines in more sophisticated ways. These improvements however, have either added new challenges to the method or they were not tested sufficiently on different nonlinear and non-stationary datasets.

Bahri and Sharples [2015] examined the performance of EMD based on piecewise-linear interpolation (i.e. the simplest form of interpolation) as applied to synthetic times series and to an actual geophysical time series. For the synthetic time series, Bahri and Sharples [2015] found that, on average, EMD based on linear interpolation was able to reproduce the known signal components very well, with IMFs very similar to those obtained from the original EMD methods using cubic splines. Overall, the linear-interpolation-based methods were able to better capture the trend in the data, while the original method provided better estimates of both the high- and low-frequency oscillatory components. For the geophysical time series (sea level), the linear-interpolation-based EMD was found to produce a different estimate of the long-term trend in the data than that obtained from classical EMD, which raised a number of questions about the interpretation of EMD output arising from actual environmental data. For example, which trend will give the most reliable information for decision making?

Endpoint considerations

The behaviour of the time series at the endpoints and appropriate boundary conditions are related critical issues that, together with the interpolation method, can affect the IMFs resulting from EMD. One of the first studies into this issue was conducted by Jin-Ping [2001]. They pioneered the mirror-extension method, in which the data are modified to form a closed curve. Therefore, the effects of the endpoints of the data do not arise. Figure 3.2 shows data extended using the mirror-extension technique. After the sifting process, only the output from the upper part is used. One limitation of this method is that the data are assumed to end at an extremum; there is no general reason to expect this to be the case.

Zeng and He [2004] modified the mirror-extension method by doing even and odd extensions to the data (Figure 3.3). They suggested the following steps to deal with the endpoints.

1. Construct a new series by even extension of the data (Figure 3.3b).
2. Construct a new series composed of all maxima. If the right endpoint of the extended dataset series is not a local maximum, append an additional maximum to the series as the periodic repeat of the leftmost maximum, as shown in Figure 3.3.

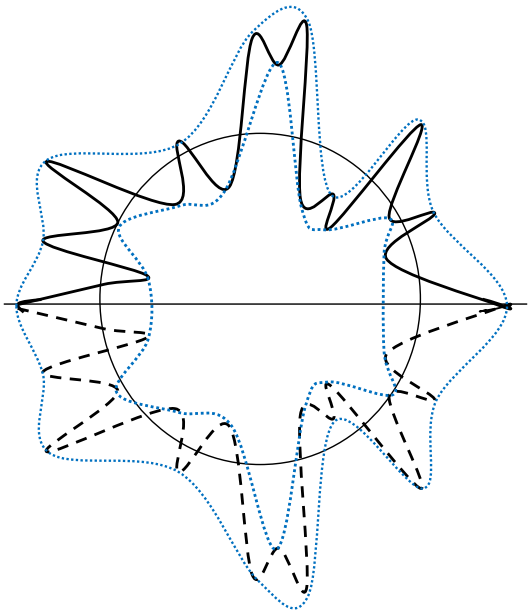

Figure 3.2: Mirror-extension technique. The data (solid line) in the upper half are reflected in the horizontal mirror (giving the dashed line) to form a closed curve. The blue dotted lines are the upper and lower envelopes.

3. Connect all the maxima in Step 2 using an interpolation method to give the upper envelope of the even-extended data series.
4. Repeat Steps 2 and 3 on the minima to construct the lower envelope of the even-extended data series.
5. Construct a new series by odd extension of the original data (Figure 3.3c).
6. Repeat Steps 2, 3 and 4 to construct the upper and lower envelopes for the odd-extended data series.
7. Subtract the mean of the four envelopes from the original data to complete one step of the sifting process.

Lin et al. [2012] combined mirror imaging and extrapolation using a regression function to reduce endpoint effects. This method consists of two steps. First, the signal is extrapolated at each endpoint using regression functions to form the primary signal. The primary signal is then further expanded by mirror extension. EMD is then performed on the resulting signal. Lin et al. [2012] tested their method using simple synthetic data (with no trend). Their results showed a significant improvement in endpoint effects.

Deng et al. [2001] applied neural networks to determine suitable boundary conditions. They tested their method using simple synthetic data, a combination of sine and cosine waves but no trend. The Hilbert Transform of the result obtained from the neural network contained more detail than the original EMD method, so

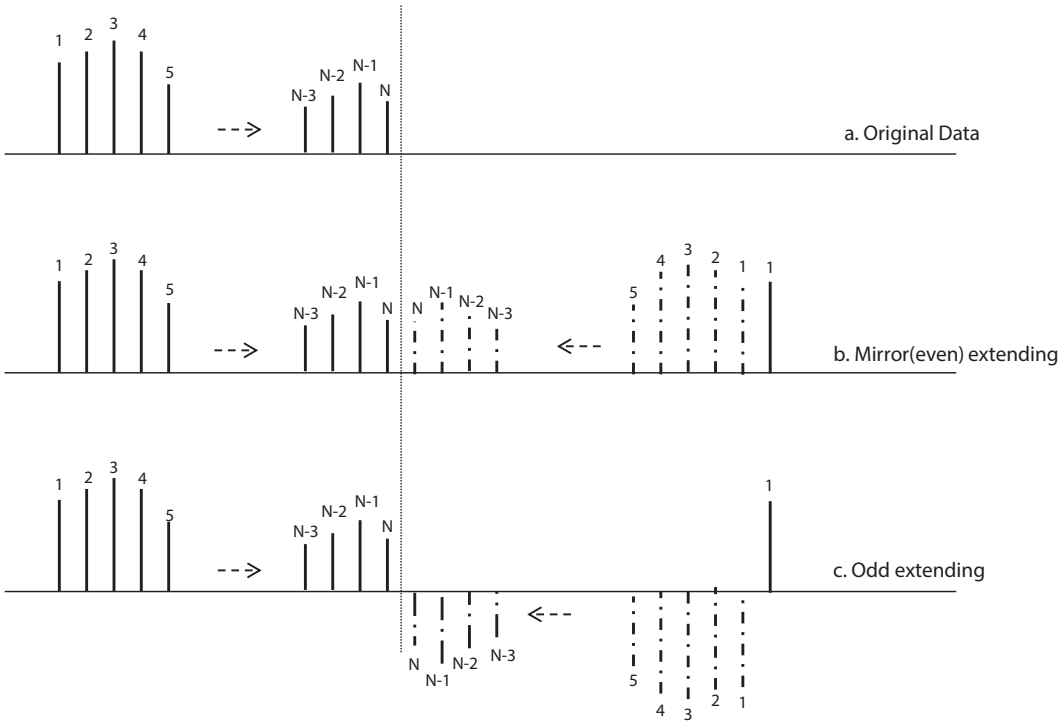

Figure 3.3: Odd and even extensions of the original N data points.

the neural network extension to the EMD improved the accuracy of the results. Zhidong and Yang [2007] introduced a method that they called the 'New Method' in an attempt to deal with endpoint effects. Essentially, they averaged the maxima and minima of all points except the first and the last point. Then the first and last points are compared with the average of the maxima and minima. If an endpoint is bigger than the average of the maxima, it is counted as a maximum point; if it is smaller than the average of the minima, it is counted as a minimum point. Zhidong and Yang [2007] then used two simple synthetic datasets (without trend), and found their results to be more accurate than the original EMD method. Ren et al. [2006] suggested the use of a window-function method, similar to Fourier Transforms. However, windowing can destroy information in the data [Dätig and Schlurmann, 2004].

Another endpoint improvement was put forward by [Dätig and Schlurmann, 2004], who discussed the importance of endpoints on the whole sifting process; the interpolation methods are sensitive to the boundary conditions, with the possibility of large swings of the interpolating functions at the endpoints, either overshooting or undershooting. These swings propagate inward, corrupting the whole data series, consequently over-riding some of the characteristics of the data.

Dätig and Schlurmann [2004] suggested a simple solution to the endpoint problem by adding a new minimum and maximum to the beginning and end of the data.

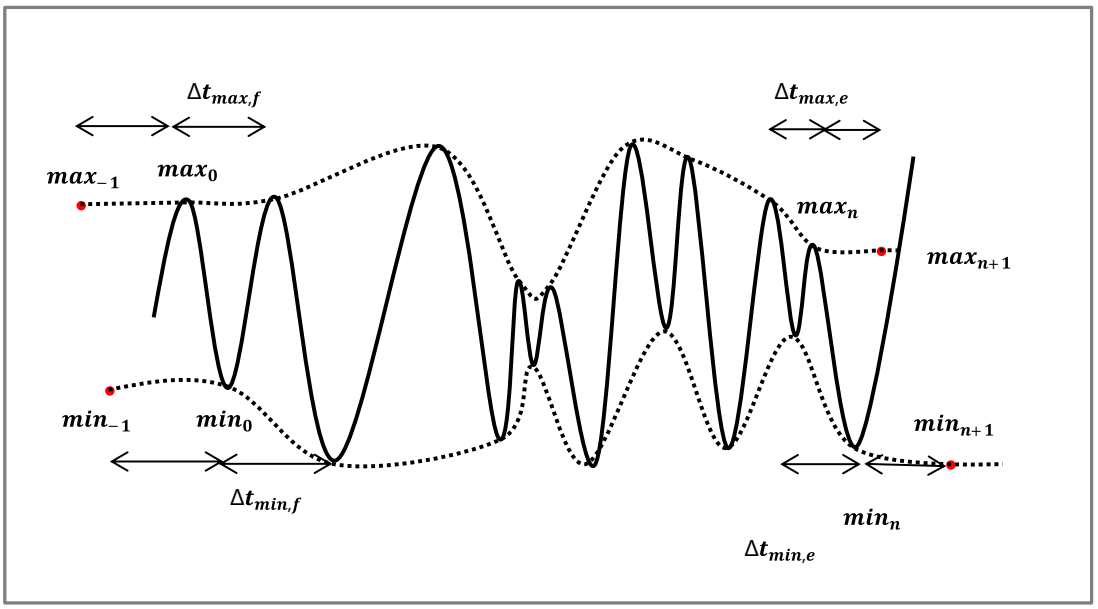

Figure 3.4: Endpoint solution according to Dätig and Schlurmann [2004].

Given a time series $\eta(t)$, if the first minimum of the dataset is at t_{min_0}, they create a new minimum at $t_{min_{-1}}$, where $\eta(t_{min_{-1}}) = \eta(t_{min_0})$. The time $t_{min_{-1}}$ is defined such that $t_{min_0} - t_{min_{-1}} = t_{min_1} - t_{min_0}$. Similarly for $t_{max_{-1}}$ and for the corresponding quantities at the end of the dataset (Figure 3.4). Dätig and Schlurmann [2004] also suggested an improvement in their method by taking the slope of the data into the account. This method is known as the Slope-Based Method (SBM) for endpoints. The slopes relevant to the start of the data are defined by:

$$\text{slope}_1 = \frac{\eta(t_{max_1}) - \eta(t_{min_0})}{t_{max_1} - t_{min_0}}, \tag{3.7}$$

$$\text{slope}_2 = \frac{\eta(t_{min_0}) - \eta(t_{max_0})}{t_{min_0} - t_{max_0}}. \tag{3.8}$$

Then,

$$\eta(t_{min_{-1}}) = (\text{slope}_1 + \text{slope}_2)\Delta t_{min} + \eta(t_{min_0}), \tag{3.9}$$

where $\Delta t_{min} = t_{min_1} - t_{min_0}$. Similar calculations are then used to find $\eta(t_{max_{-1}})$ and the corresponding quantities for the end of the dataset.

Wu and Qu [2008] discussed the SBM method and found that, although this method works most of the time for non-stationary and nonlinear data, sometimes the interpolation functions oscillate wildly at the end of the time series. They therefore developed the improved Slope-Based Method. In this method they extended the slope of two neighbouring maximum values, then compared that with the value of the endpoints. The bigger value was chosen for the upper envelope. Similarly for

the lower envelope. They used synthetic data (a combination of two cosine signals) and compared their results with those using the SBM method; their results were more robust and accurate.

The Ensemble Empirical Mode Decomposition

As mentioned in Section 3.4.1, noisy input data can cause mode mixing in the EMD method. When dealing with noisy datasets, the EMD procedure is implemented as part of an ensemble process, which is referred to as the ensemble EMD (EEMD). In the EEMD procedure, noise $\epsilon \approx N(0, \delta)$, normally distributed with zero mean and standard deviation δ, is added to the EMD; the means of the IMFs found from numerous EMD applications give the the final ensemble IMFs [Huang and Wu, 2008].

Torres et al. [2011] introduced what they called a Complete Empirical Mode Decomposition to help combat the mode-mixing problem when noise is present. In this method, they added a particular type of noise at each stage of the sifting process and obtained a unique residual for each mode. To test the result, they used a synthesised Dirac signal with 512 samples. Their results had fewer IMFs than the EEMD method. The high-frequency IMFs in both methods were similar but at lower frequencies, there appeared to be more mode mixing than using the Complete Ensemble Empirical Mode Decomposition method. Another advantage of using a Complete Empirical Mode Decomposition is that it provides an improvement in terms of computational cost. Wu and Huang [2009] used a noisy synthetic data and compared the results of EMD and EEMD. They showed that the results from EEMD were significantly better than the ones from EMD. They mentioned that although EEMD improves mode mixing effects, it does not reduce the impact of different sifting stoppage rules or end effect methods.

3.5 Methods used in this study

3.5.1 EMD or Ensemble EMD?

In this study, whenever the data are noiseless, the EMD method is used. In the presence of noise, the Ensemble EMD (EEMD) method is used, with the number of ensemble members depending on the noise in the data (noisier data need more ensemble members).

3.5.2 Endpoints

In this study the Slope-Based Method discussed in Section 3.4.2 is used for the endpoints. The effect of the endpoints, which also impact the accuracy of the IMFs found using the different interpolation methods, is clearly an area for further study.

In this study, the main focus is to investigate the sensitivity of EMD/EEMD to interpolation method. We did not investigate on endpoint effects.

3.5.3 Interpolation methods

In this study, as mentioned before, we focus on Step 3 of the EMD algorithm, which determines the upper and lower envelopes, using each of three interpolation methods. In addition to the cubic-spline interpolation originally advocated by Huang et al. [1998a], we also use piecewise-linear interpolation and a smoothing-spline interpolation of the extrema.

Linear interpolation

A piecewise-linear function is the simplest piecewise interpolating function. In piecewise-linear interpolation (just 'linear interpolation' from now on), each successive pair of data points $(x_i, y_i), (x_{i+1}, y_{i+1})$ is joined by a straight line.

A significant issue with linear interpolation is that the overall interpolating function is not differentiable or smooth. A non-differentiable function can introduce new issues into a system in a similar manner to a non-continuous function. In general, linear interpolation is almost always inappropriate unless the function is almost linear or a large error is acceptable [Khoury and Harder, 2015].

Despite all the problems with linear interpolation, it is considered because it is the simplest interpolation method, and it enables us to find the largest possible errors in the method that can occur in choosing an interpolation method. The linear interpolation function used here is the *interp1* function in MATLAB.

Cubic-spline interpolation

Interpolation with cubic splines is essentially equivalent to passing a flexible plastic curve through the data points. If one can hammer some nails into a board to represent a set of data points, the plastic curve can then be deformed to touch all the nails. Between the nails, the curve acts as the interpolating function. The cubic-spline interpolant and its first two derivatives are continuous at the data points. Let $g_i(x)$ be the cubic in the interval $x_i \leq x \leq x_{i+1}$ and $g(x)$ denote the collection of all cubics for the whole data. Then,

$$g_i(x) = \frac{g''(x_i)}{x_i - x_{i+1}} \frac{(x - x_{i+1})^3}{6} + \frac{g''(x_{i+1})}{x_{i+1} - x_i} \frac{(x - x_i)^3}{6} + C_1 x + C_2. \tag{3.10}$$

C_1 and C_2 are constants that can be obtained by matching the function at the boundaries, which leads to:

$$\begin{aligned} g_i(x) &= \frac{g''(x_i)}{6}\left[\frac{(x_{i+1}-x)^3}{\Delta_i} - \Delta_i(x_{i+1}-x)\right] \\ &\quad + \frac{g''(x_{i+1})}{6}\left[\frac{(x-x_i)^3}{\Delta_i} - \Delta_i(x-x_i)\right] \\ &\quad + f(x_i)\frac{x_{i+1}-x}{\Delta_i} + f(x_{i+1})\frac{x-x_i}{\Delta_i}, \end{aligned} \tag{3.11}$$

where $\Delta_i = x_{i+1} - x_i$, $x_i < x < x_{i+1}$ and $i = 1, 2, 3, \ldots, N-1$. The unknowns $g''(x_i)$ and $g''(x_{i+1})$ can be obtained using the continuity of the first derivative [Moin, 2010]. The cubic-spline interpolation function used here is the *spline* function in MATLAB.

Smoothing-spline interpolation

The smoothing spline (*csaps* in MATLAB) for a dataset $y(x_j)$ finds the interpolating function $f(x)$ that minimizes the quantity

$$p\sum_{j=1}^{n}\left|y(x_j) - f(x_j)\right|^2 + (1-p)\int\left|D^2 f(t)\right|^2 dt, \tag{3.12}$$

where n is the number of data points and the integral is over the smallest interval containing all the data points. p is a smoothing parameter: if $p \to 0$, f is the least-squares straight-line fit to the data; at the other extreme, if $p \to 1$, f is the variational, or 'natural' cubic-spline interpolant. Smoothing effectively averages over local variations. Due to noise, however, there will be a range of p values for which the method works well. We find that there is a range of p values that extracts the low-frequency behaviour, but misses the high frequencies, and vice versa. The smoothing parameter in our implementation of EMD/EEMD is chosen to favour retrieval of the lower-frequency IMFs and the trend (residual). We provide some discussion of how the smoothing parameter can be chosen to give good high-frequency performance.

3.5.4 Validation of the EMD/EEMD algorithm

In this study, we implement the EMD/EEMD algorithm in MATLAB following the EMD procedure given by Huang and Wu [2008] but using the Slope-Based Method for the endpoints (Section 3.4.2). In particular, we have not attempted to ameliorate further any of the known issues with endpoints and extrapolation of the interpolating functions beyond the first and last extreme points. The reader is referred to Dätig and Schlurmann [2004] and Peel et al. [2009] for further discussion of these issues.

The implementation of the EMD algorithm used in this work was validated[1] by

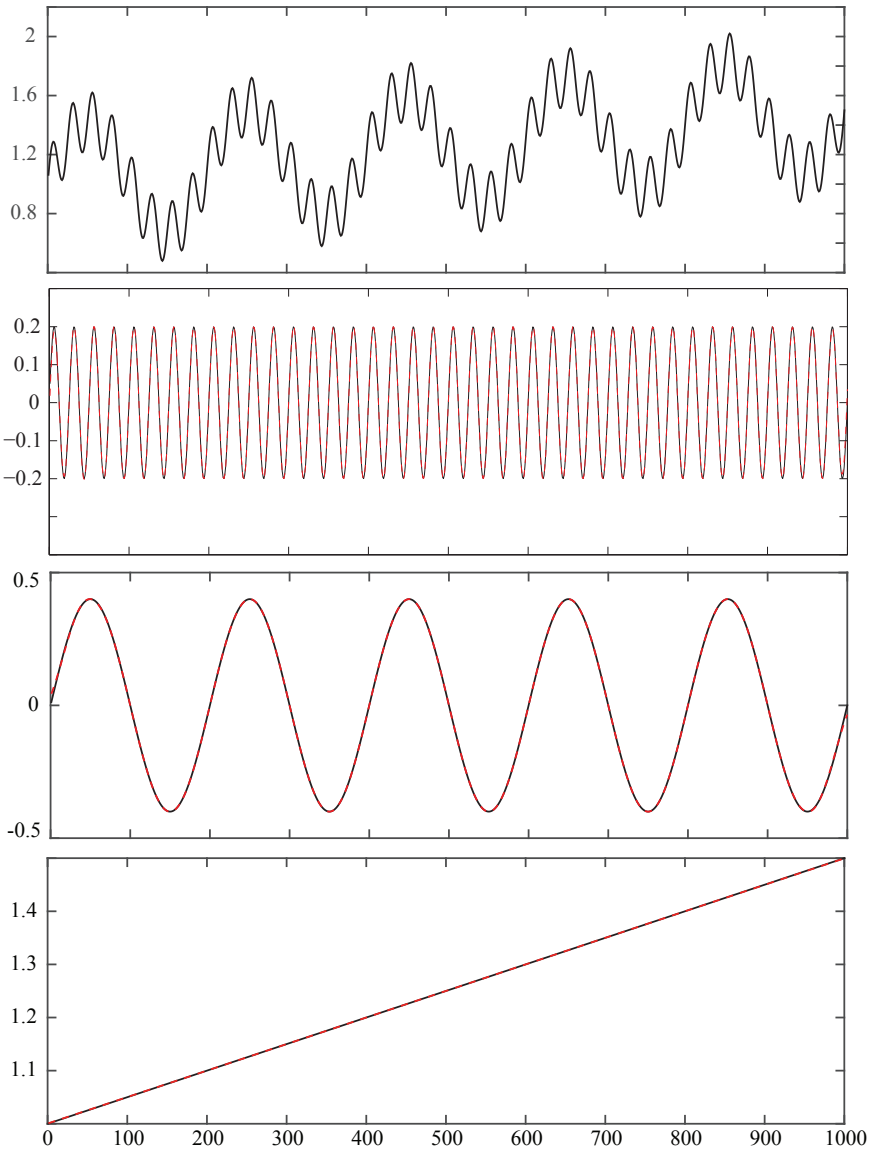

Figure 3.5: Validation of the EMD algorithm. The top panel shows the input synthetic data (see Chapter 4); the other three panels show the three components extracted by the algorithm (using cubic-spline interpolation) from the data (red dashed lines), together with the actual components (black solid lines).

applying it to the synthetic data in Figure 3.5. The EMD method, as coded for this study, extracted the three components in the data very accurately.

We note that there are number of other codes available following the same algorithm, such as Huang's code[2], Flandrin's EEMD codes[3] and the MATLAB HHT toolbox.[4]

[1]The code used in this study is available from the author
[2]http://rcada.ncu.edu.tw/intro.html
[3]http://perso.ens-lyon.fr/patrick.flandrin/emd.html
[4]http://au.mathworks.com/matlabcentral/fileexchange/19681-hilbert-huang-transform

Chapter 4

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

This chapter describes the construction and use of synthetic datasets to assess the application of alternative interpolation methods in the EMD and EEMD methods and how well they deal with non-stationary data.

4.1 Synthetic time series

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. The synthetic datasets here cover different types of non-stationarity with and without noise. In addition, due to the focus of this study on environmental data, the synthetic data are designed so that they have similar characteristics to environmental data such as temperature and sea-level data.

Global warming is causing a general increase in the mean and changing the variance of environmental data such as temperature and sea level. An aim of this study is to see if EMD can extract reliable information about the shape of the (long-term) trend. Obtaining an accurate estimate of the trend is important in environmental studies, as it shows long-term changes in the climate. Environmental scientists use the shape of the trend to calculate acceleration in the mean of the data [Church and White, 2011], which is used for environmental planning and adaptation. The frequency of significant environmental events such as El Niño and La Niña has also changed due to global warming; this appears as changes in the variance of the data [Priestley, 1988, Parzen, 1999].

Hence, to systematically assess the performance of EMD when applied to time series possessing different types of non-stationarity, separate synthetic time series are constructed with: (i) time-dependent mean; (ii) time-dependent variance; and (iii)

time-dependent mean and variance. The sensitivity of the EMD procedure to the different types of data non-stationarity is then investigated systematically.

4.1.1 Stationary synthetic time series

As a basis for constructing non-stationary synthetic time series, we use a stationary time series comprised of two sinusoidal components, a low-frequency component $L(t)$ and a high-frequency component $H(t)$:

$$x_0(t) = L(t) + H(t); \quad t = 1, 2, \ldots, 1000, \tag{4.1}$$

where

$$L(t) = 0.4 \sin\left(\frac{2\pi t}{200}\right); \quad H(t) = 0.2 \sin\left(\frac{2\pi t}{25}\right). \tag{4.2}$$

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

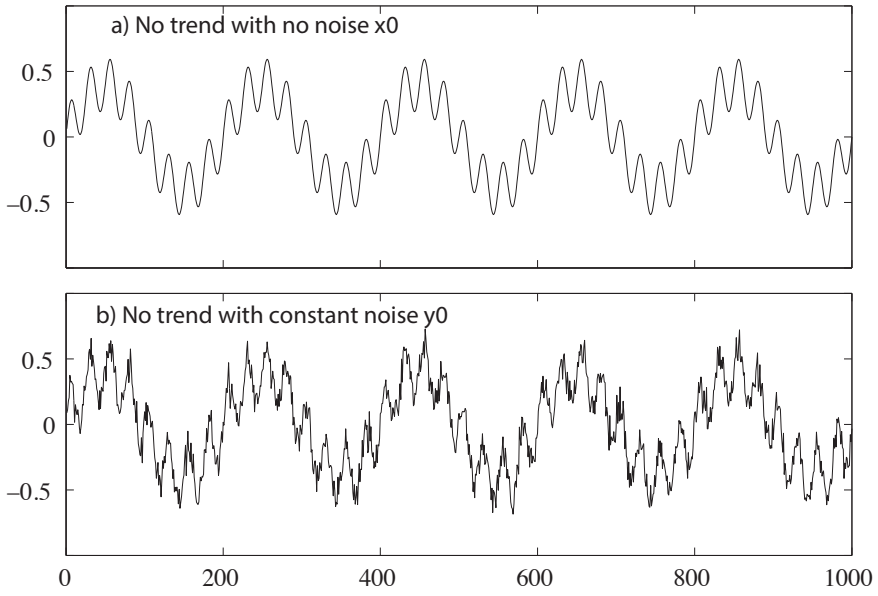

Figure 4.1: Synthetic stationary time series **(a)** $x_0(t)$ with constant mean and constant variance (Equation 4.1); **(b)** $y_0(t)$ with constant mean and constant variance with added noise (Equation 4.3).

We next define the noisy time series

$$y_0(t) = x_0(t) + \epsilon_0(t); \quad t = 1, 2, \ldots, 1000, \tag{4.3}$$

where the noise $\epsilon_0 \sim N(0, \delta_0)$, i.e. normally distributed with mean zero and constant standard deviation δ_0; δ_0 is taken as 20% of the standard deviation of x_0 (Figure 4.1b).

4.1.2 Synthetic data with time-dependent mean

Non-stationary time series with time-dependent mean are constructed by adding various trends $T_i(t)$ to $x_0(t)$. Specifically, we consider the non-stationary time series

$$x_i(t) = x_0(t) + T_i(t), \quad i = 1, \ldots, 4; \quad t = 1, 2, \ldots, 1000, \tag{4.4}$$

with trends

$$T_1(t) = 1 + \frac{t}{2000}, \tag{4.5}$$

$$T_2(t) = 1 + \frac{(t+300)^2}{200000}, \tag{4.6}$$

$$T_3(t) = 7 - \ln(1002 - t), \tag{4.7}$$

$$T_4(t) = 2\exp(-0.09(20t - 1)^2). \tag{4.8}$$

T_1 is a 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; and T_4 is a Gaussian trend, with a sharp increase followed by a sharp decrease. Figure 4.2 shows the non-stationary time series defined by Equation 4.4 with the respective trends, Equations 4.5–4.8.

We next define noisy versions of these time series as:

$$y_i(t) = x_i(t) + \epsilon_i, \quad i = 1, \ldots, 4; \quad t = 1, 2, \ldots, 1000. \tag{4.9}$$

The noise terms $\epsilon_i \sim N(0, \delta_i)$ are added to emulate actual data (e.g. environmental data). The constant variance of the noise δ_i is 20% of the standard deviation of x_i. Figure 4.3 shows the non-stationary time series defined by Equation 4.9 with the respective trends, Equations 4.5–4.8.

4.1.3 Synthetic data with time-dependent variance

Two separate time series, z_{05} and z_{06} with time-dependent variance are constructed based on $x_0(t)$:

$$z_{0j}(t) = x_0(t) + \epsilon_j(t), \quad j = 5, 6; \quad t = 1, 2, \ldots, 1000, \tag{4.10}$$

where the time-dependent noise term is $\epsilon_i(t) \sim N(0, \delta_i(t))$, with

$$\delta_5(t) = 0.0005t, \tag{4.11}$$

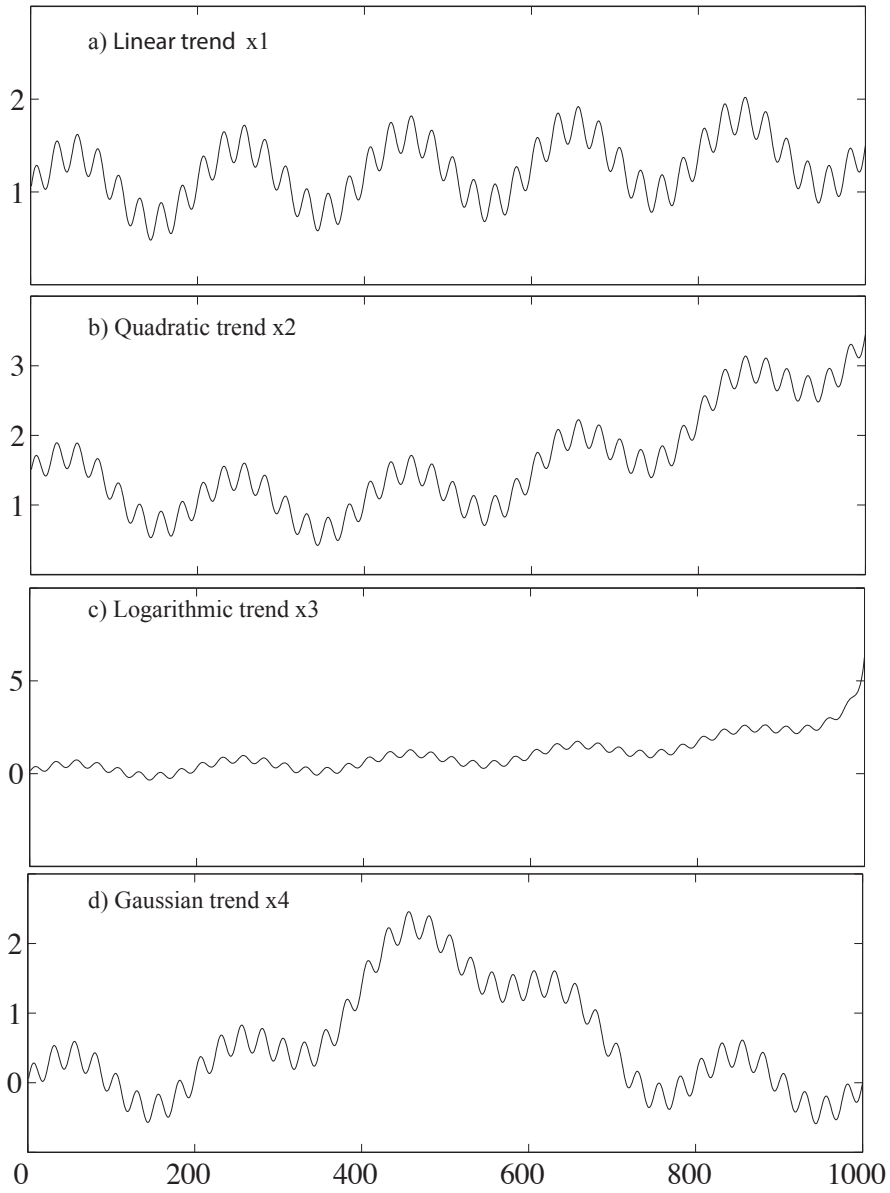

Figure 4.2: Synthetic data with time-dependent mean (Equation 4.4): **(a)** linear trend x_1; **(b)** quadratic trend x_2; **(c)** logarithmic trend x_3; **(d)** Gaussian trend x_4.

and

$$\delta_6(t) = 0.0005 + 0.5 \exp\left(-(0.006t - 3)^2\right). \tag{4.12}$$

Hence, the time series $z_{05}(t)$ represents data with a monotonically increasing standard deviation, while $z_{06}(t)$ represents data with maximum variance in the middle of the series. Figure 4.4 shows the non-stationary time series defined by Equation 4.10.

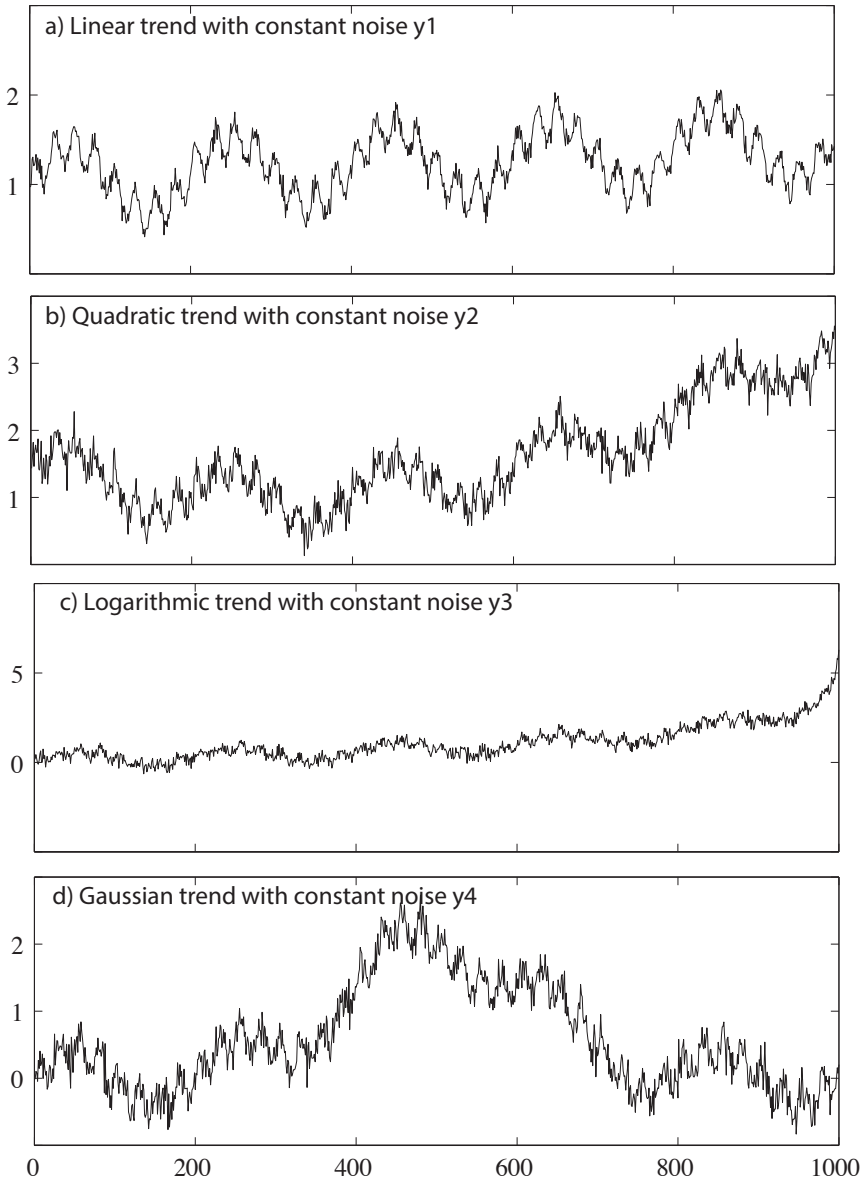

Figure 4.3: Synthetic data with constant noise ϵ_i and time-dependent mean (Equation 4.9): **(a)** linear trend y_1; **(b)** quadratic trend y_2; **(c)** logarithmic trend y_3; **(d)** Gaussian trend y_4.

4.1.4 Synthetic data with time-dependent mean and time-dependent variance

Time-series data with both a time-dependent mean and time-dependent variance are constructed as follows:

$$z_{ij}(t) = x_i(t) + \epsilon_j(t), \quad i = 1, 2, \ldots 4; \quad j = 5, 6; \quad t = 1, 2, \ldots, 1000. \tag{4.13}$$

These time series are shown in Figure 4.5 for $z_{i5}(t)$ and Figure 4.6 for $z_{i6}(t)$.

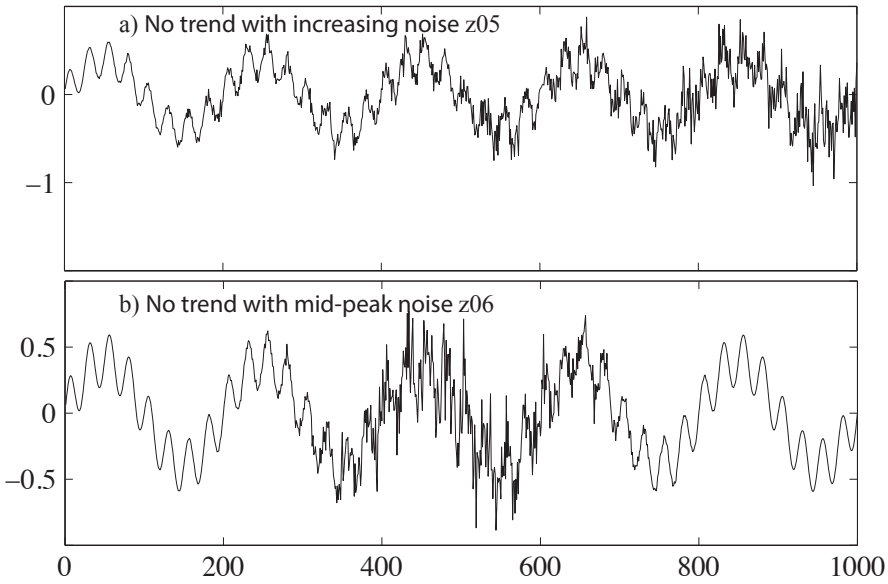

Figure 4.4: Synthetic data with time-dependent variance (Equation 4.10): (**a**) $z_{05}(t)$; (**b**) $z_{06}(t)$.

Hence, a total of twenty synthetic datasets (2 stationary and 18 non-stationary) are used to assess the performance of the EMD and EEMD procedures using the different interpolation methods. These datasets are summarised in Table 4.1

Data	**Description**	**Equation**
x_0	sum of sinusoids L, H	4.1
y_0	x_0 plus constant noise ϵ_0	4.3
x_i	x_0 plus trends T_i, $i = 1, \ldots, 4$	4.4; 4.5– 4.8
y_i	x_i plus constant noise ϵ_i; $\epsilon_i \sim N(0, \delta_i)$	4.9
z_{05}	$x_0 + \epsilon_5$; $\delta_5(t) = 0.0005t$	4.10; 4.11
z_{06}	$x_0 + \epsilon_6$; $\delta_6(t) = 0.5\big(0.001 + e^{-(0.006t-3)^2}\big)$	4.10; 4.12
z_{ij}	$x_i + \epsilon_j$, $\epsilon_j \sim N(0, \delta_j)$, $j = 5, 6$	4.13

Table 4.1: Summary of the datasets used as input to EMD/EEMD.

4.2 Sensitivity of EMD and EEMD to the interpolation method.

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 estimate the sinusoidal signals $L(t)$ and $H(t)$ from the stationary input signals x_0 (Equation 4.4) and y_0 (Equation 4.9). We then investigate their performance in recovering the sinusoidal signals and the various trends T_i (Equations 4.5 to 4.8) from the input time series with the various forms of non-stationarity.

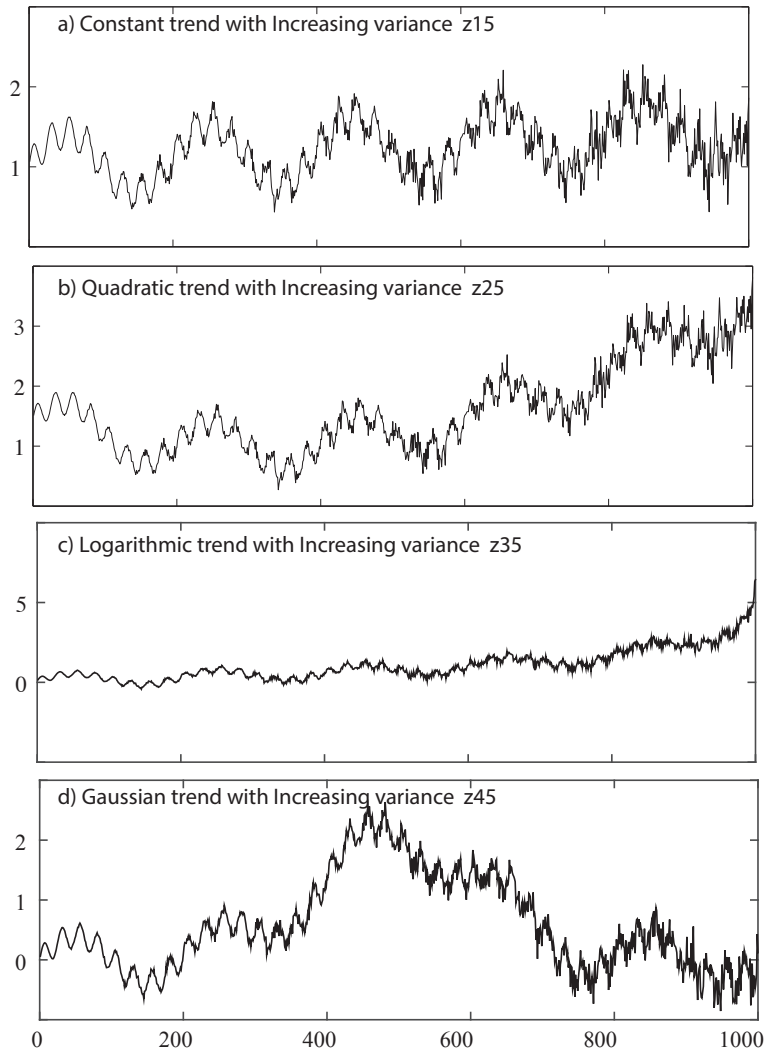

Figure 4.5: Synthetic data with time-dependent mean and variance (Equation 4.13 with $\epsilon_5(t)$): **(a)** linear trend z_{15}; **(b)** quadratic trend z_{25}; **(c)** logarithmic trend z_{35}; **(d)** Gaussian trend z_{45}.

We first show the EMD/EEMD original results for two selected datasets and then discuss how we analyse 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 are too small for the modes to carry valuable information. Therefore eliminating 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 Peng et al. [2005]. Figures 4.7 to 4.9 show the results of LEEMD (linear–EEMD), CEEMD (cubic–EEMD) and SEEMD (smoothing–EEMD) on the three datasets (x_0, y_0 and y_1).

Figure 4.7 shows the results of EMD for the x_0 stationary data. The linear EMD extracts six modes from the input data, four of which are too small. The cubic

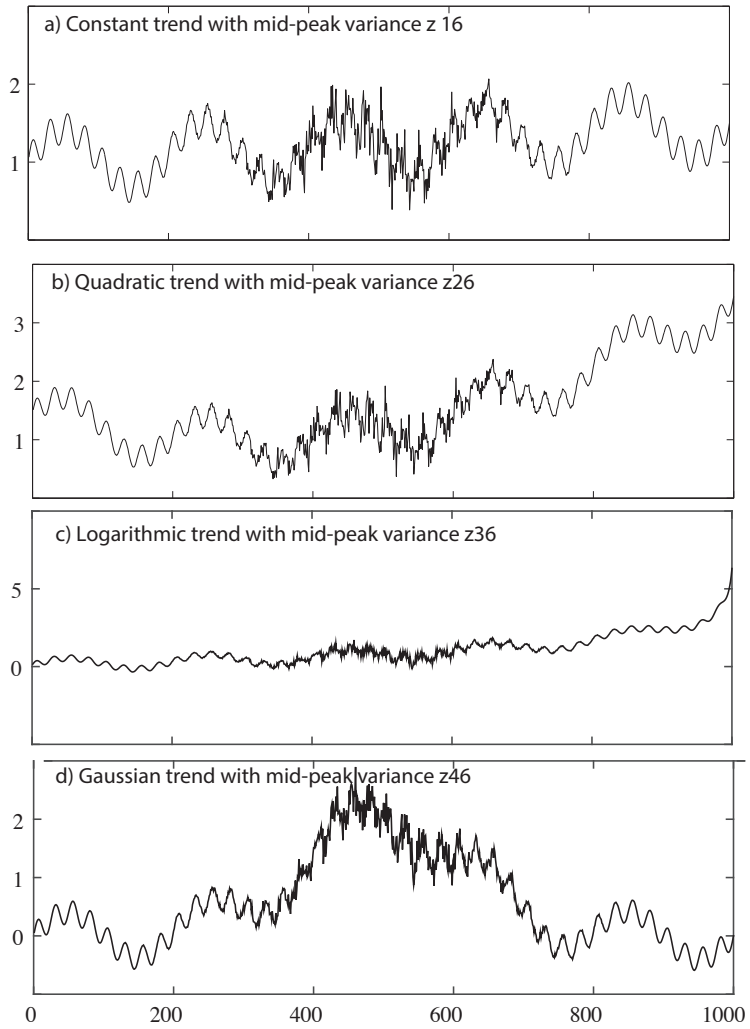

Figure 4.6: Synthetic data with time-dependent mean and variance (Equation 4.13 with $\epsilon_6(t)$): **(a)** linear trend z_{16}; **(b)** quadratic trend z_{26}; **(c)** logarithmic trend z_{36}; **(d)** Gaussian trend z_{46}.

and smoothing EMD both extract five modes from the input data of which, in both cases, three are too small. So, all interpolations produce two modes, which nearly perfectly match the original modes (correlation coefficient ≈ 1). The first mode produced is the high-frequency mode, the second the low-frequency mode. This shows that for the noiseless stationary data, the EMD method is able to extract the actual modes regardless of the type of interpolation used.

In the presence of noise, the results are not as straightforward. Figure 4.8 shows the EMD outputs for the three interpolation methods for the y_0 data (Equation 4.3). 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

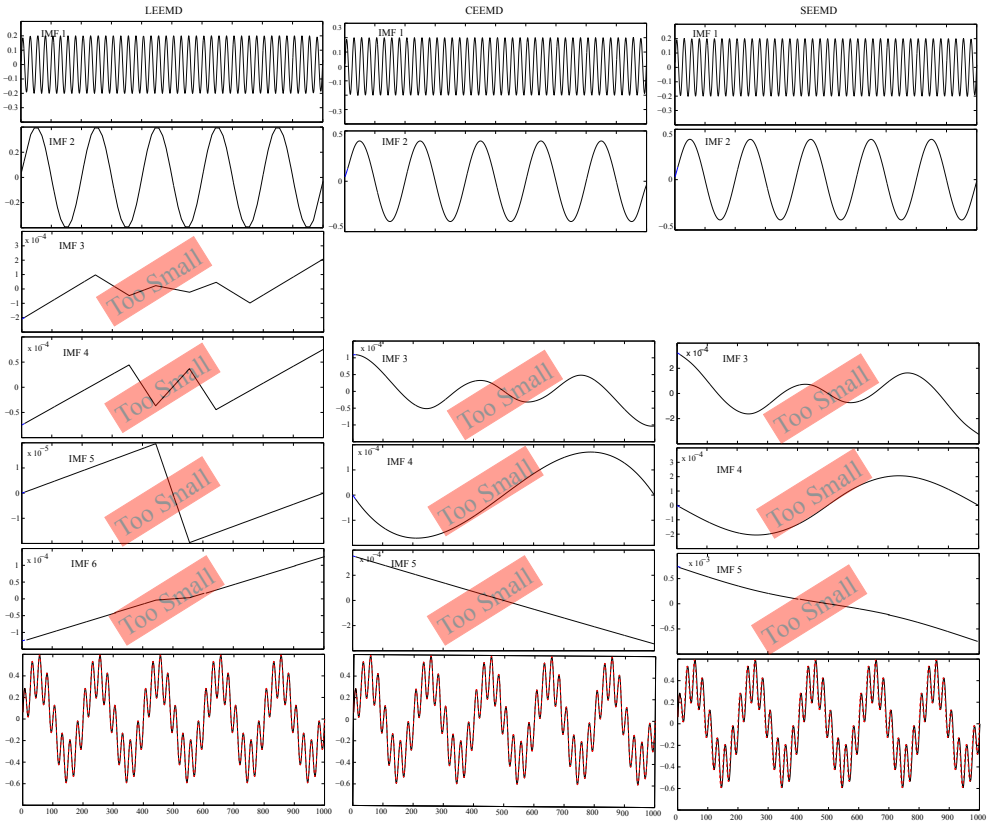

Figure 4.7: EMD results for the noiseless stationary data x_0 (Equation (4.1)) using LEEMD (left column), CEEMD (middle column) and SEEMD (right column). The top two rows shows the high- and low-frequency modes extracted by all three methods. The bottom row shows the sum of the two extracted modes (red dash line) compared with the original signal x_0 (black solid line).

.

this mode in SEEMD is underestimated due to the smoothing parameter being set to extract the low-frequency mode more accurately (see Section 4.4).

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 [Peng et al., 2005]. It is not clear which is the better method.

Adding the modes, then adding the trend does 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. Another problem is in deciding which modes to add up. This is quite challenging, especially when we deal with data with several high or low frequencies. Then, the IMFs are often a combination of two 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_0 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.

Figure 4.9 shows the EEMD results for data with constant noise and trend (y_1, Equation (4.9)). LEEMD extracts ten modes, CEEMD nine and SEEMD only six. Similar to the y_0 results, the first two modes of LEEMD and CEEMD, and the first mode of SEEMD represent noise. The high-frequency mode then appears in the third IMF from LEEMD and CEEMD, and the second from SEEMD. The next three modes from LEEMD and CEEMD show that the methods struggle to extract the low-frequency mode due to the challenges mentioned above. However, SEEMD does find the low-frequency mode accurately. This shows another advantage of using smoothing-spline interpolation in the EEMD method.

The mean absolute error (MAE) is used to quantify the differences between the IMFs/residuals from EMD/EEMD and the corresponding sinusoidal components/trends of the input data, that is how well EMD/EEMD recovers the original data:

$$\mathrm{MAE}_I = \frac{1}{N}\sum_{k=1}^{N}\left|M(t_k) - \mathrm{IMF}_I(t_k)\right|. \tag{4.14}$$

Here, $M(t)$ can be one of the sinusoidal components $L(t)$ or $H(t)$ in Equation 4.1, or one of the trends $T_i(t)$ in Equations 4.5–4.8; $\mathrm{IMF}_I(t)$ is the corresponding IMF. The subscript I denotes the interpolation method: L: linear; C: cubic; S: smoothing; and N is the number of data points.

MAE values for all the analyses below are given in Table 2 on page 58.

4.2.1 EMD/EEMD and the stationary data

We now consider the performance of the EMD procedure applied to the input noiseless stationary time series $x_0(t)$ (Equation 4.1) and that of the EEMD procedure applied to corresponding time series with noise, $y_0(t)$ (Equation 4.3). Although EMD was originally developed to deal with non-stationary data, the ability of the EMD and EEMD methods to retrieve the component signals from stationary data establishes a benchmark for the subsequent analyses using non-stationary data.

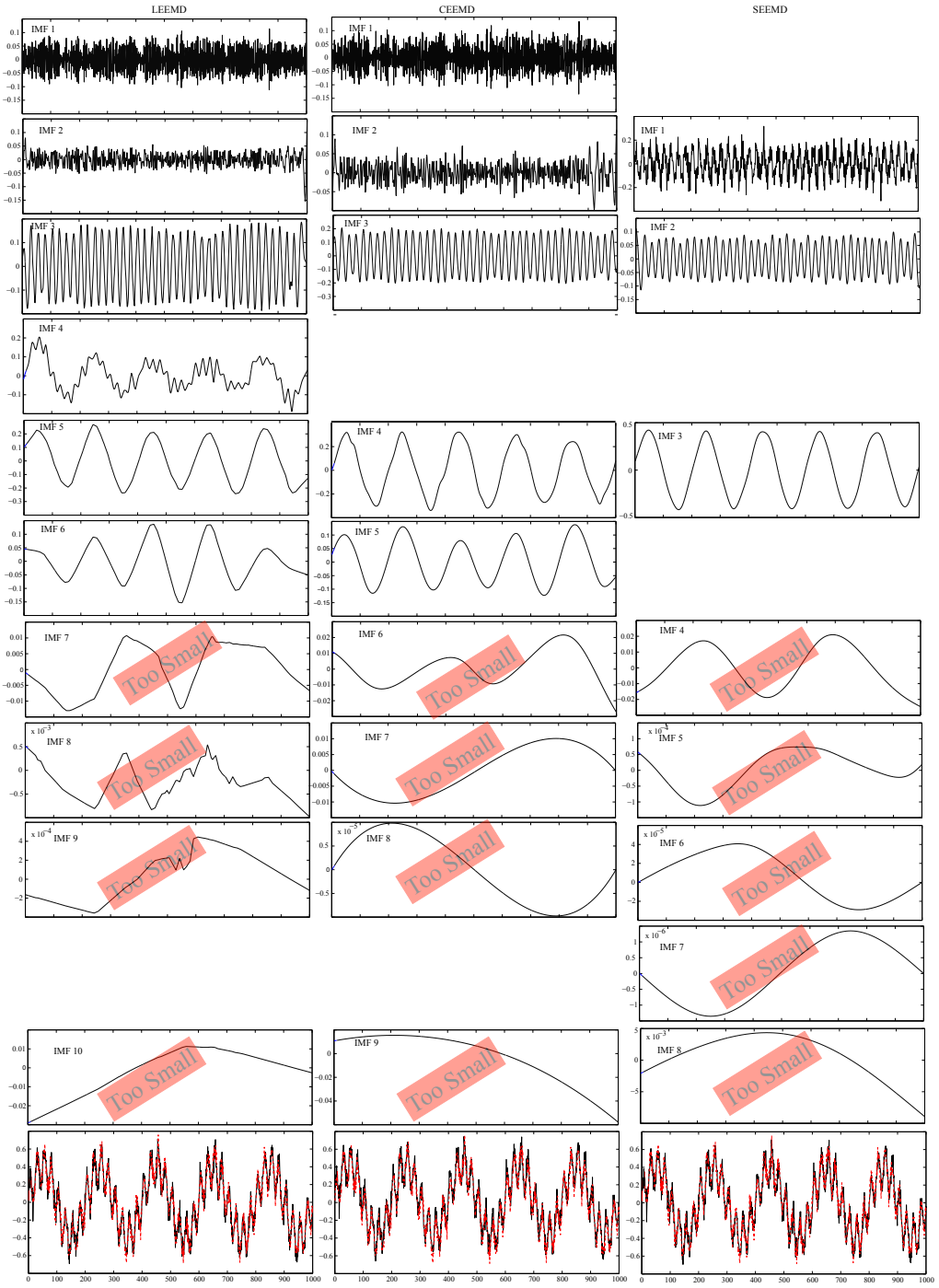

Figure 4.8: EMD results for the noisy y_0 data (Equation 4.3) using LEEMD (left column), CEEMD (middle column) and SEEMD (right column). The top two rows are noise, the third row the high-frequency modes and rows 4 to 6 low-frequency modes extracted by three methods. The bottom row shows the sum of the two extracted modes (red dashed line) compared with the original signal y_0 (black solid line).

Figure 4.10 shows that EMD applied to the noiseless time series x_0 is able to extract the high- and low- frequency sinusoidal components to a very high degree of accuracy. The cubic- and smoothing-splines with EMD performed the best, with mean absolute

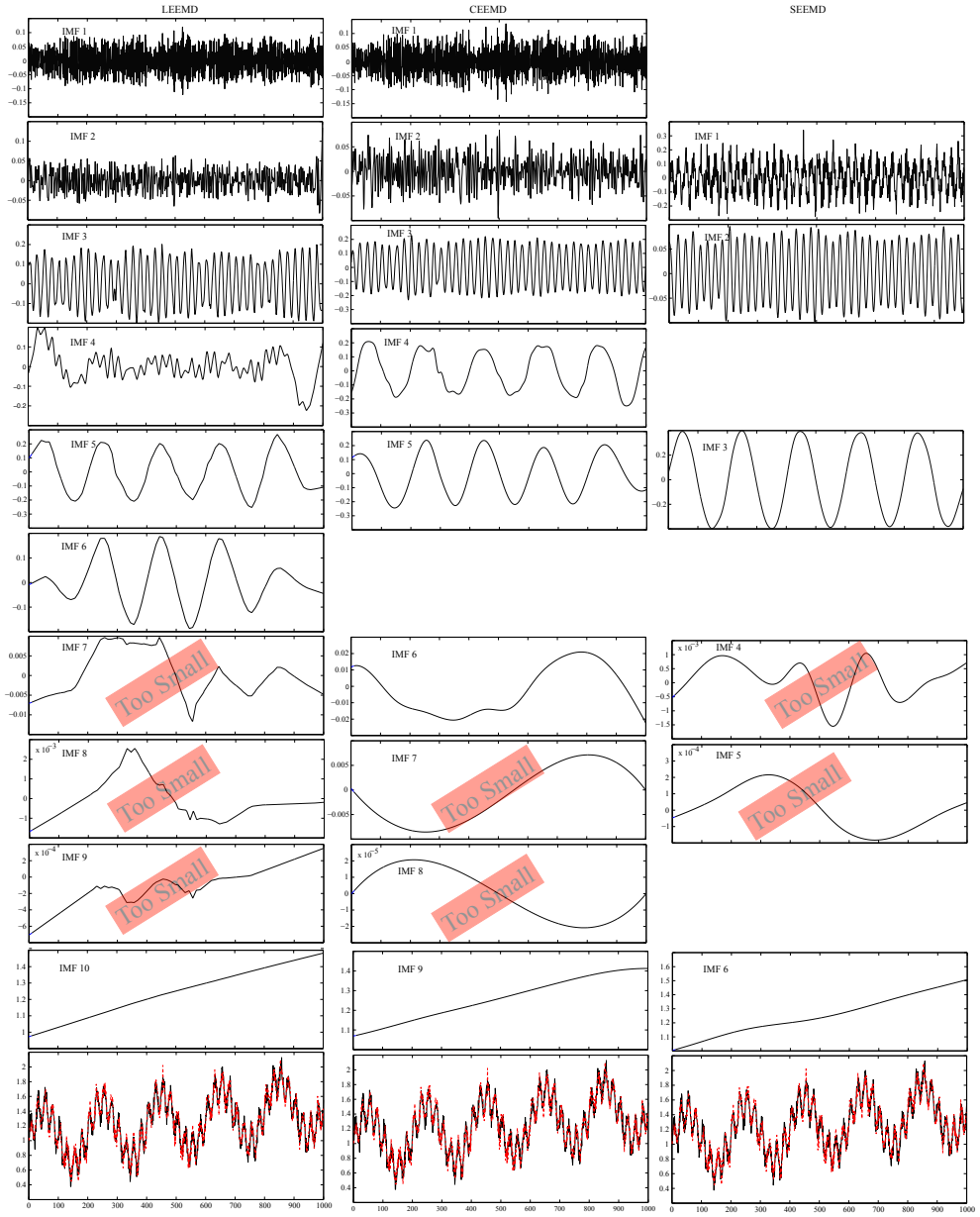

Figure 4.9: EMD results for the noisy y_1 data (Equation 4.9) using LEEMD (left column), CEEMD (middle column) and SEEMD (right column). The top two rows are noise, the third row the high-frequency modes and rows 4 to 6 low-frequency modes extracted by three methods. The 10th row is the trend and the bottom row shows the sum of the two extracted modes (red dashed line) compared with the original signal y_1 (black solid line).

errors (MAEs) of 6.68×10^{-4} and 5.22×10^{-4}, respectively, for the low-frequency component, and MAEs of 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.

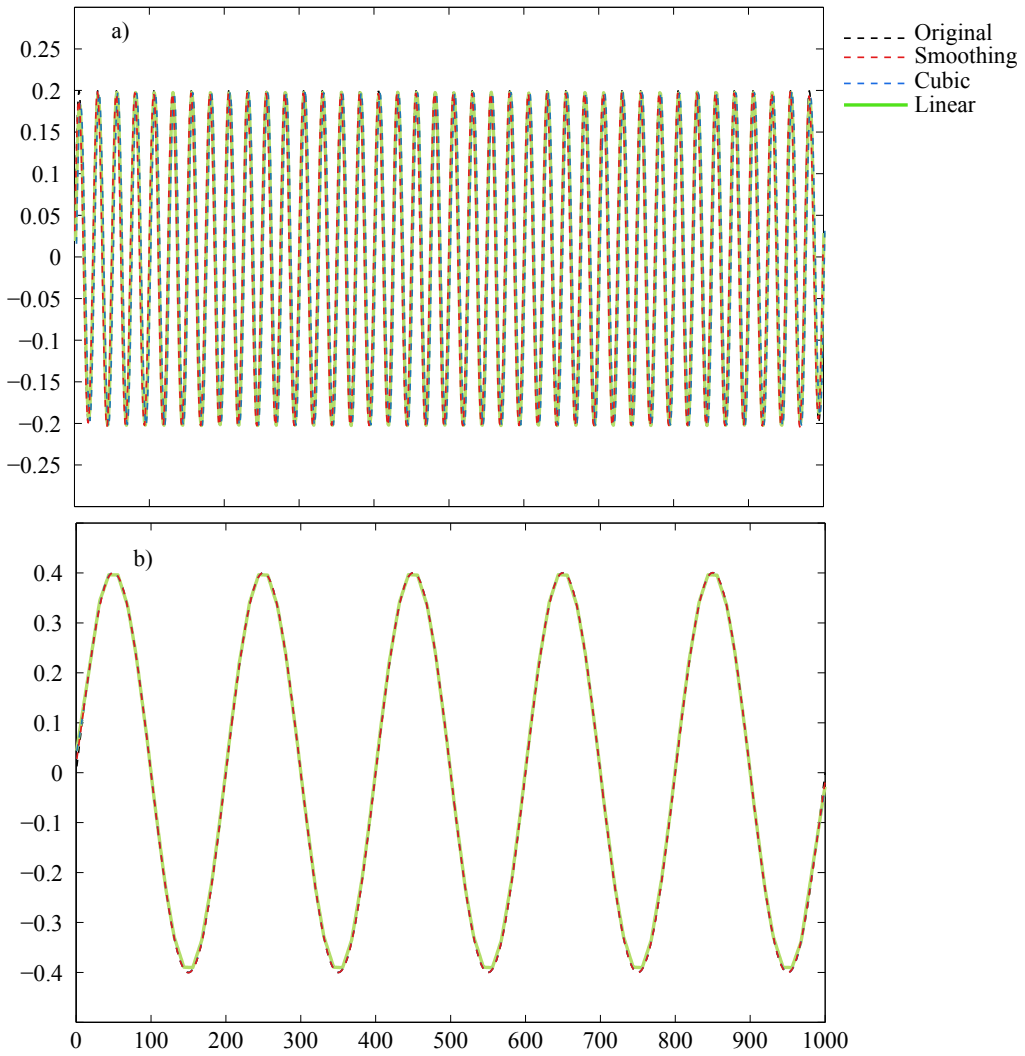

Figure 4.10: EMD results for the noiseless stationary x_0 (Equation 4.1); (**a**): high-frequency IMFs; (**b**): low-frequency IMFs. *Original* is either the high-frequency component $H(t)$ (top) or the low-frequency component $L(t)$ (bottom) of the input signal.

EEMD applied to the noisy time series y_0 (Equation 4.3) is able to accurately recover the frequencies of the sinusoidal components, but performs less well in capturing their amplitudes, as can be seen in Figure 4.11. For the high-frequency component, 'cubic EEMD' (CEEMD) performs the best (MAE $= 2.22 \times 10^{-2}$), while 'smoothing EEMD' (SEEMD) performs the worst (MAE $= 7.86 \times 10^{-2}$). 'Linear EEMD' (LEEMD) is able to capture the high-frequency component to a similar degree of accuracy as the cubic EEMD (MAE$= 3.69 \times 10^{-2}$), but produces very inaccurate results for the low-frequency component (MAE $= 1.15 \times 10^{-1}$). For the low-frequency component, smoothing EEMD performed the best (MAE $= 1.61 \times 10^{-2}$), with cubic EEMD very inaccurate (MAE $= 1.86 \times 10^{-1}$).

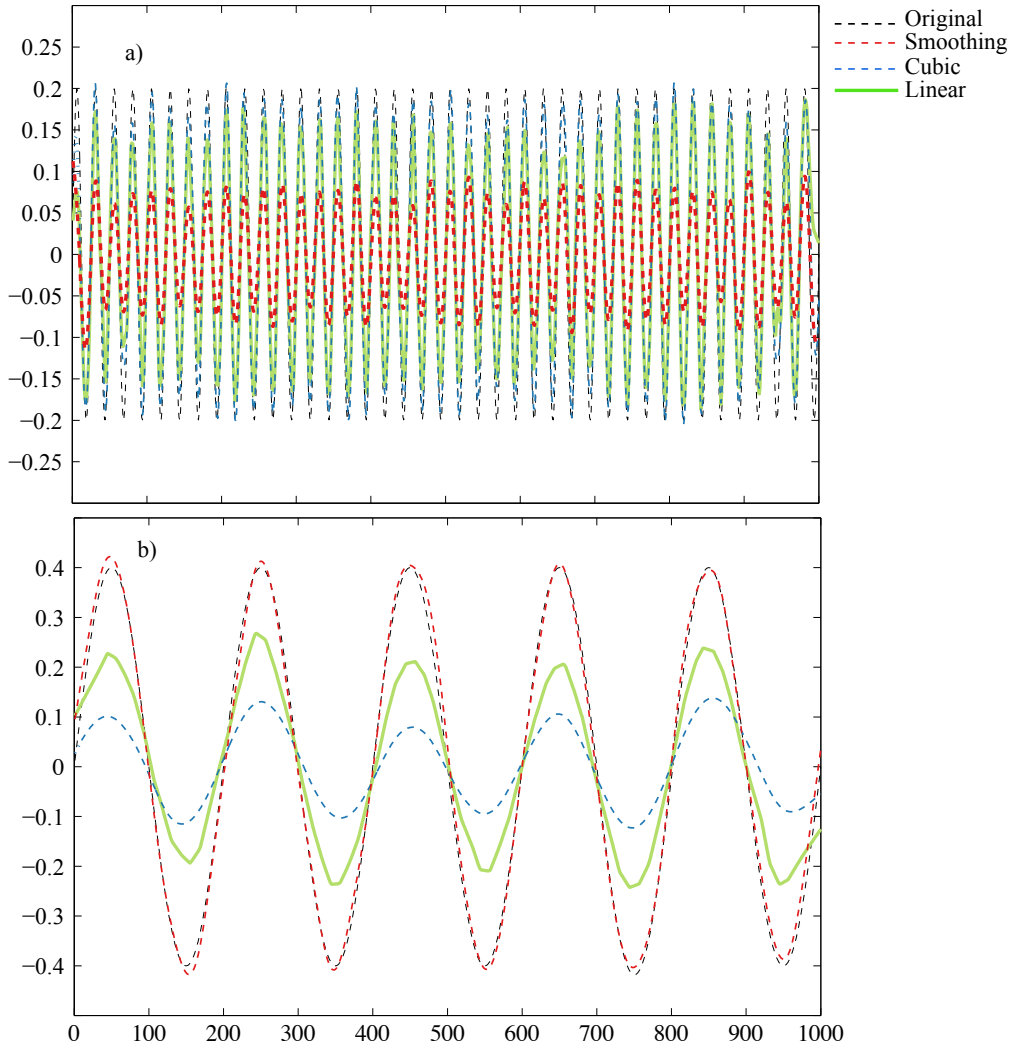

Figure 4.11: EEMD results for the corresponding data with noise y_0 (Equation (4.3)); (**a**): high-frequency IMFs; (**b**): low-frequency IMFs. *Original* is either the high-frequency component $H(t)$ (top) or the low-frequency component $L(t)$ (bottom) of the input signal.

The large error in smoothing EEMD for the high-frequency modes can be reduced by changing the smoothing parameter. This is discussed in more detail in Section 4.4.

4.2.2 EMD/EEMD and non-stationary data with time-dependent mean

EMD applied to x_1, the noiseless time series with linear trend (Equation 4.4), performs very well regardless of the interpolation method used, producing identical MAE statistics to those obtained from x_0 for the sinusoidal modes. Similarly, EMD based on each of the interpolation methods is able to accurately extract the linear trend in x_1, 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.

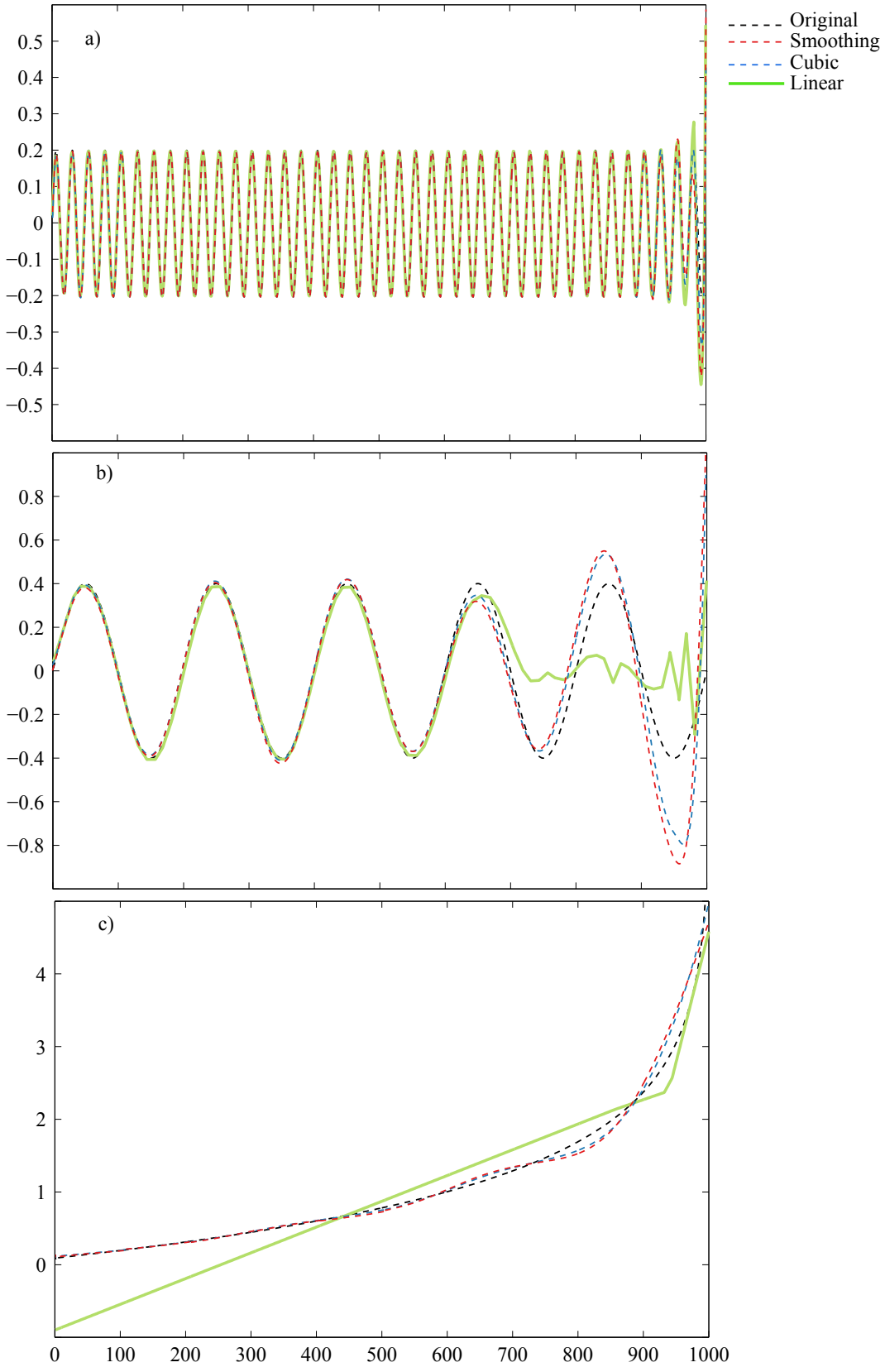

Figure 4.12: EMD results for the non-stationary data with logarithmic time-dependent mean, x_3 (Equation 4.4).

The corresponding MAE statistics for the times series x_2, x_3 and x_4 can be seen in Table 4.2 (page 58). In the absence of noise, cubic EMD and smoothing EMD provide the better estimates of the sinusoidal components and the various trend components. Linear EMD performs considerably worse in estimating the logarithmic trend component of x_3, while EMD with all the interpolation methods has difficulty in recovering the Gaussian trend component of x_4.

Figures 4.12 shows the results of the various EMD procedures applied to x_3. It is evident that the nature of the logarithmic trend results in poor EMD performance; this is particularly noticeable in the inability of the linear EMD to estimate the low-frequency sinusoidal component in the region where the trend increases abruptly, though similar issues can be seen with the cubic and smoothing EMD results (Figure 4.12). There are also problems with the estimation of the high-frequency component near the end of the time series for the all interpolation methods (Figure 4.12, top panel). Similar issues with the performance of the linear EMD are evident for the low-frequency mode in Figures 4.12 and 4.13, with the region of particularly poor performance coinciding with where the trend component exhibits the greatest variability.

The various EMD methods perform relatively poorly in extracting the Gaussian trend component from x_4 (Figure 4.13). The linear EMD method, in particular, fails to accurately extract the trend. The cubic and smoothing EMD methods are able to extract the general pattern and symmetry of the trend component, but underestimate its variability. Each of the EMD methods, however, is able to accurately reproduce the high-frequency mode of x_4.

The MAE statistics resulting from application of EEMD to the constant-noise time series y_1, y_2, y_3 and y_4 can be seen in Table 4.2 (page 58). For the linear-trend time series y_1, the smoothing EEMD results in a more accurate recovery of the three signal components, in particular the linear trend and the low-frequency sinusoidal components, much better that the linear and cubic EEMD. Indeed, the MAE for the low-frequency component is an order of magnitude smaller with smoothing EEMD. Smoothing EEMD also deals with noise but it is important to set the right smoothing parameter (see Section 4.4).

For the time series y_2 with quadratic trend, the smoothing EEMD again provides the lowest MAE values for the low-frequency component, although those from the cubic EEMD are only slightly higher. Similarly, for the time series y_4 with Gaussian trend, smoothing EEMD yields the lowest MAE for the low-frequency component. The exception is the logarithmic trend in y_3, for which cubic EEMD yields a lower MAE than the smoothing EEMD (see Table 4.2, page 58). As can be seen in Figure

4.14, the linear EEMD has considerable difficulty in accurately reproducing the amplitudes of both sinusoidal components.

4.2.3 EEMD on non-stationary data with time-dependent variance

The results of applying the various EEMD methods to the non-stationary time series z_{05} and z_{06} are presented in Figures 4.15 and 4.16. Figures 4.15 and 4.16 clearly show that the ability of linear EEMD to recover the high-frequency sinusoidal component is considerably affected by the noise, a result which is consistent with those relating to the application of linear EEMD to the time series y_1, y_2, y_3 and y_4. Again, it is the amplitudes of the estimated components that are most sensitive to the presence of noise, whereas their frequencies appear to be far more robust. The regions of poorest performance of linear EEMD coincide with the regions where the amplitude of the noise is greatest, near the end of the time series for z_{05} and in the middle of the time series for z_{06}. For z_{05}, linear EEMD yields an MAE for the high-frequency component of 5.02×10^{-2}, while cubic and smoothing EEMD produce MAEs of 4.514×10^{-2} and 7.73×10^{-2}, respectively. For the case of z_{06}, in which the amplitude of the noise is greatest in the middle of the time series, linear EEMD produces a MAE of 4.66×10^{-2}, while cubic and smoothing EEMD produce 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 modes. Unfortunately, that means the high-frequency IMFs of SEEMD 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 4.4.

For the low-frequency sinusoidal component, Figures 4.15 and 4.16 show that the 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_{05}, linear EEMD yields an MAE for the low-frequency component of 1.67×10^{-1}, while cubic and smoothing EEMD produces MAEs of 9.41×10^{-1} and 1.65×10^{-2}, respectively. For z_{06}, linear EEMD

produce a MAE of 1.60×10^{-1}, while cubic and smoothing EEMD produce 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 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_{06} case (Figures 4.15 and 4.16). This shows that smoothing EEMD with a suitable smoothing parameter is a good alternative to the standard (cubic interpolation) EEMD, with greater accuracy and less mode mixing.

With smoothing splines, the best accuracy is achieved for the low-frequency modes because the smoothing parameter is set in favour of the low-frequency modes.

4.2.4 EEMD on non-stationary data with time-dependent mean and variance

Here we consider the results of applying the various EEMD methods to the time series z_{ij} with $i \geq 1$ (Equation 4.13). Table 4.2 (page 58) details the resulting MAE statistics for the various cases, while Figures 4.17–4.20 show some selected examples of EEMD performance.

Consistent with the results in the previous section, linear EEMD is highly sensitive to the presence of noise in the input time series. This is clearer when we consider the MAEs for the x_i $(3.2 - 6.8)$ and the y_i $(36.9 - 81.2)$ in Table 4.2 (page 58). Again, it is the amplitude of the estimated signals that is most badly affected (e.g. see Figure 4.17). Interestingly however, linear EEMD is better able to accurately reproduce the trend component of the input signals, in contrast to its ability to reproduce their sinusoidal components. Table 4.2 (page 58) 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 mode. Figures 4.18 and 4.19 show that cubic EEMD is also prone to poor performance, particularly in its ability to reproduce the low-frequency sinusoidal component.

Inspection of Table 4.2 indicates that, overall, smoothing EEMD provides the most accurate estimation of the original signal components. Indeed, smoothing EEMD produces the smallest MAE values in all but two of the cases considered, namely z_{16} and z_{25}, for which cubic EEMD yields the smallest MAE values. However, in these cases the smoothing EEMD mean absolute errors are only slightly larger. For the sinusoidal components, smoothing EEMD consistently produces the smallest MAE values. For the low-frequency components, smoothing EEMD produces MAE values an order of magnitude smaller than those obtained from linear and cubic EEMD.

It should be noted that in these cases the smoothing EEMD MAE values are only slightly larger: 2.62×10^{-2} as opposed to 1.65×10^{-2} for the z_{16} case; and 1.62×10^{-2} as opposed to 1.21×10^{-2} for the z_{25} case. It is also worth noting that these two cases involve estimation of the trend component.

In terms of estimating the trend component of the input signal, it is the logarithmic and Gaussian trends that most confound the various EEMD procedures. Figures 4.19 and 4.20 show that the abrupt changes represented by these trend components result in poor EEMD performance. For the logarithmic trend cases, this is particularly evident near the end of the time series, while for the Gaussian trend cases the EEMD methods fail to accurately estimate the full extent of the trend, similar to that discussed above in the case of y_4. Also, as was the case with y_4, it is worth noting that despite its relatively poor performance, smoothing EEMD better captures the variability and symmetry of the Gaussian trend component for the two cases z_{45} and z_{46} (see Appendix).

4.3 Robustness of the EEMD ensemble members

In this section, we consider the consistency of the various interpolation-based EMD procedures that are employed in EEMD to estimate the trend component of the various input time series. To do this, we construct confidence intervals based on the full spectrum of the EEMD ensemble members. Figures 4.21 and 4.22 show EEMD trend components along 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 lay, while the upper confidence limit defines the value above which 2.5% of the ensemble member values lay.

Figure 4.21 shows the trend components estimated by applying the various EMD methods to the time series x_1, x_2, x_3 and x_4. The confidence intervals in Figure 4.21a indicate that there is a considerable degree of variability in the linear trends determined in each of linear EMD ensemble runs. In particular, the confidence intervals indicate that linear EMD produces linear trend estimates with gradients ranging from 0.58 to 1.30 (width of trend estimation = 0.72). Figure 4.21a shows that cubic EMD also produces 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 EMD are 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 Figures 4.21b, 4.21c and 4.21d, respectively. Linear EMD produces the broadest range of trend estimates, while smoothing EMD exhibits considerable robustness, producing consistent estimates of the various trends across the 3000 ensemble runs. In the case of the Gaussian trend in Figure 4.21d, the confidence intervals associated with linear and cubic EEMD indicate that they are far less consistent than smoothing EMD in their ability to estimate the trend. However, it should be noted that the linear EEMD confidence intervals are the only ones that encompass the actual trend component.

Figure 4.22 shows that similar patterns in the confidence intervals arise when considering time series data with time-dependent mean and time-dependent variance (z_{i5}, $i \geq 1$). The broadest confidence intervals arise in connection with linear EMD, while smoothing EMD exhibits a considerable degree of robustness across the ensemble members.

4.4 Varying the smoothing parameter in SEEMD

In the previous sections, smoothing EEMD was implemented assuming a smoothing parameter of $p = 0.0015$. 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 more accurate estimation of the high-frequency component, and what effect it has on the accuracy of estimates of the trend and low-frequency component. Given that as $p \to 1$, smoothing-spline interpolation approaches cubic interpolation, 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 components used previously:

$$Y(t) = S(t) + T_1(t) + \epsilon_0, \tag{4.15}$$

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

and:

$$S(t) = 0.6 \sin\left(\frac{2\pi t}{400}\right) + 0.4 \sin\left(\frac{2\pi t}{200}\right) + 0.4 \sin\left(\frac{2\pi t}{100}\right) + 0.6 \sin\left(\frac{2\pi t}{50}\right); t = 1, 2, \ldots, 1000. \tag{4.16}$$

These data are shown in Figure 4.23.

SEEMD is used on these data with the smoothing parameter varying in the range $10^{-6} < p < 1$. The IMFs for each p value are compared with the corresponding sinusoidal components in Equation 4.16; the MAEs for each component are shown in Figure 4.24.

For $p > 10^{-1}$ and $p < 10^{-3}$, the MAEs for all IMFs fluctuate considerably and are relatively high: some values of p give good results for the higher-frequency IMFs but not for the lower-frequency IMFs, and vice versa (Figure 4.24). However, for $10^{-3} < p < 10^{-1}$, the MAEs for all four IMFs are low, suggesting that this range contains suitable values for the smoothing parameter. However, this is for the synthetic data presented here; the range may be different for different data. Finding suitable values for the smoothing parameter needs further study. An obvious approach would be to use generalised cross validation [Wahba and Wang, 1995] to

determine an optimal value of p at each stage of EMD. However, this would increase the computational demands of the procedure considerably.

4.5 Discussion and conclusions

The lack of a complete formal mathematical framework underpinning EMD means that theoretical analysis of the sensitivity of the EMD method is not possible. Studies like this one, which 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 stationary data and EEMD to non-stationary data have been investigated using different types of non-stationarity (time-dependent mean, time-dependent variance or both). 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 underestimates the amplitude of the sinusoidal signal components. Linear interpolation did however provide reasonable estimates for the trend components (except for data with abruptly changing 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 was set to favour 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, at least for the datasets considered in this study. There is a need for further study on how to find suitable smoothing parameters for data in general, without inflating the computational demands of the method too much.

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 the non-stationarity of the input data. In all cases, there is much less variation in IMF ensemble members using SEEMD.

We have also shown that EEMD performance depends on the nature of the input data. It is sensitive to abrupt changes in the mean, 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 mean, EEMD is sensitive to the choice of interpolation method.

Estimation of the frequency of the sinusoidal components by the IMFs appears less sensitive to the data than estimation of the amplitude. 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 analysing long-term changes, and critically important to the accurate study of environmental data [Ezer et al., 2013].

For completeness and future reference, the results of our study for each type of non-stationary data are summarised in Table 4.2. Full details are in Appendix 1.

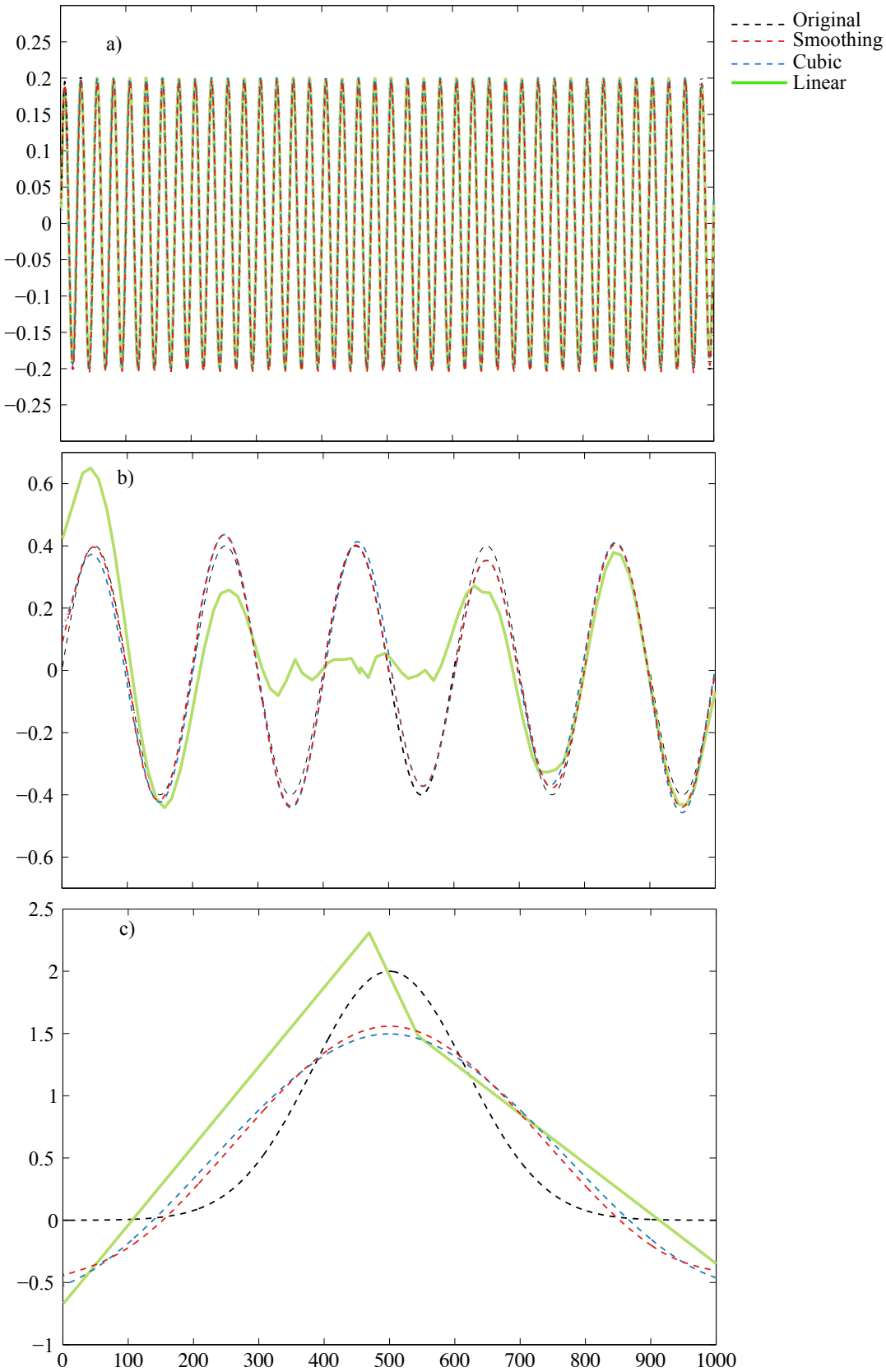

Figure 4.13: EMD results for the non-stationary data with Gaussian time-dependent mean, x_4 (Equation 4.4).

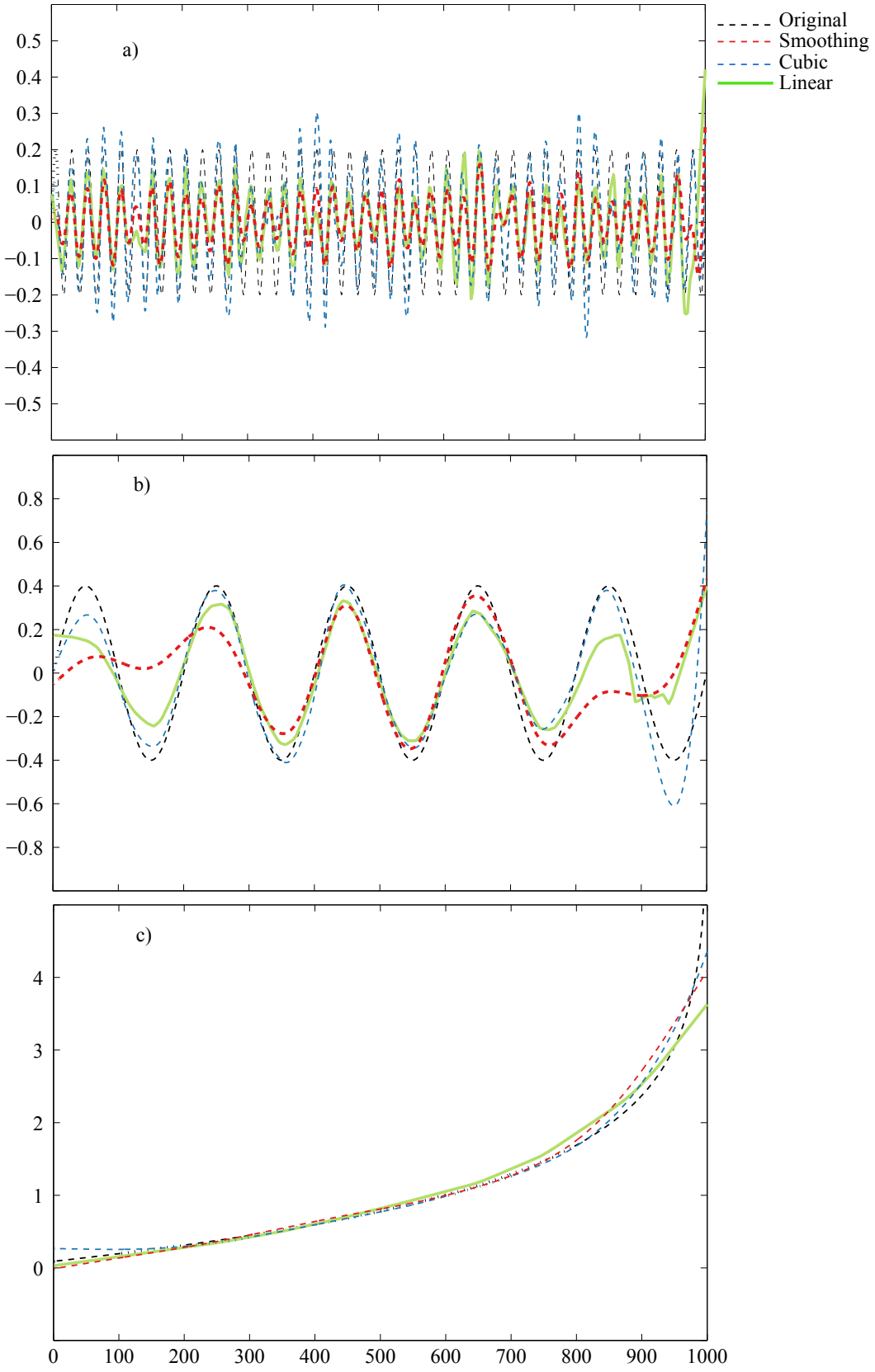

Figure 4.14: EEMD results for the non-stationary data with time-dependent mean and constant noise, y_3 (Equation 4.9).

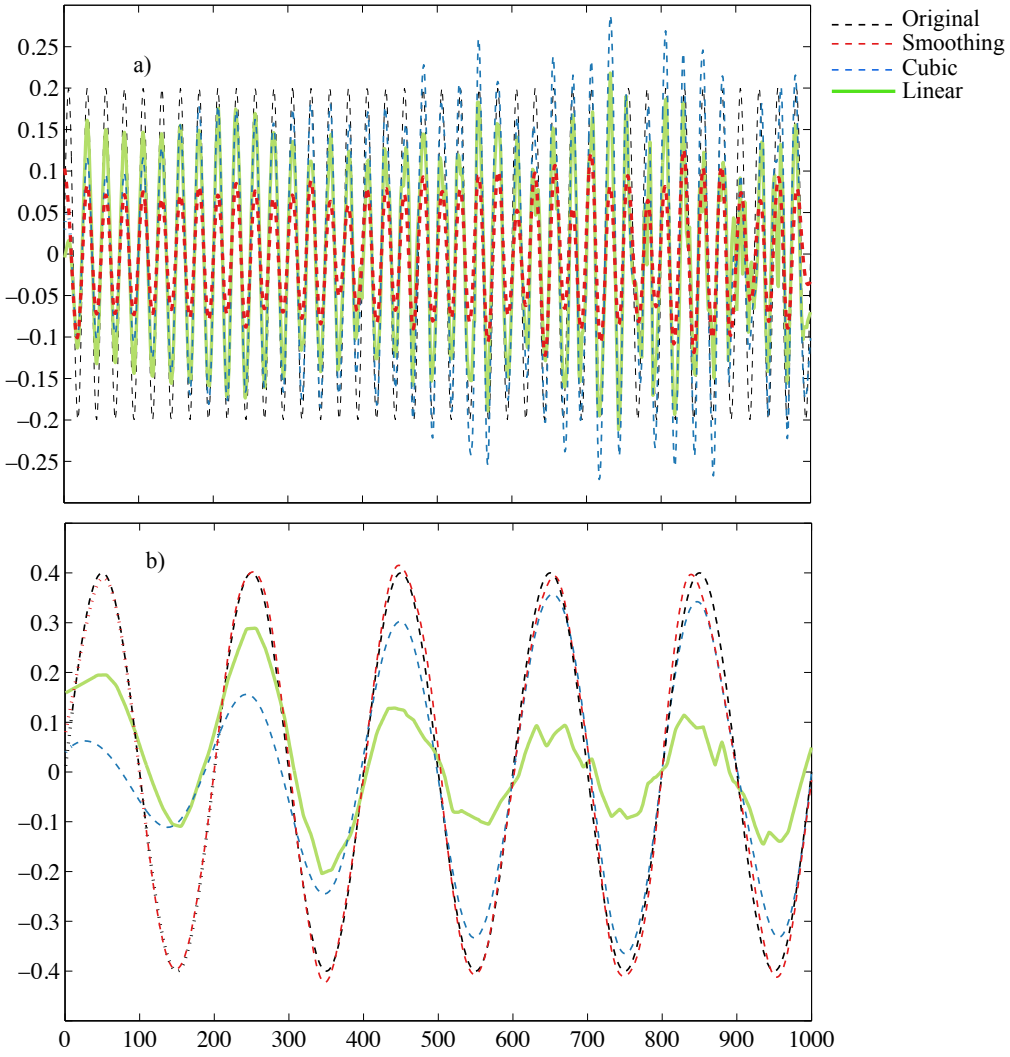

Figure 4.15: EEMD results for the non-stationary data with time-dependent variance, z_{05} (Equation 4.10).

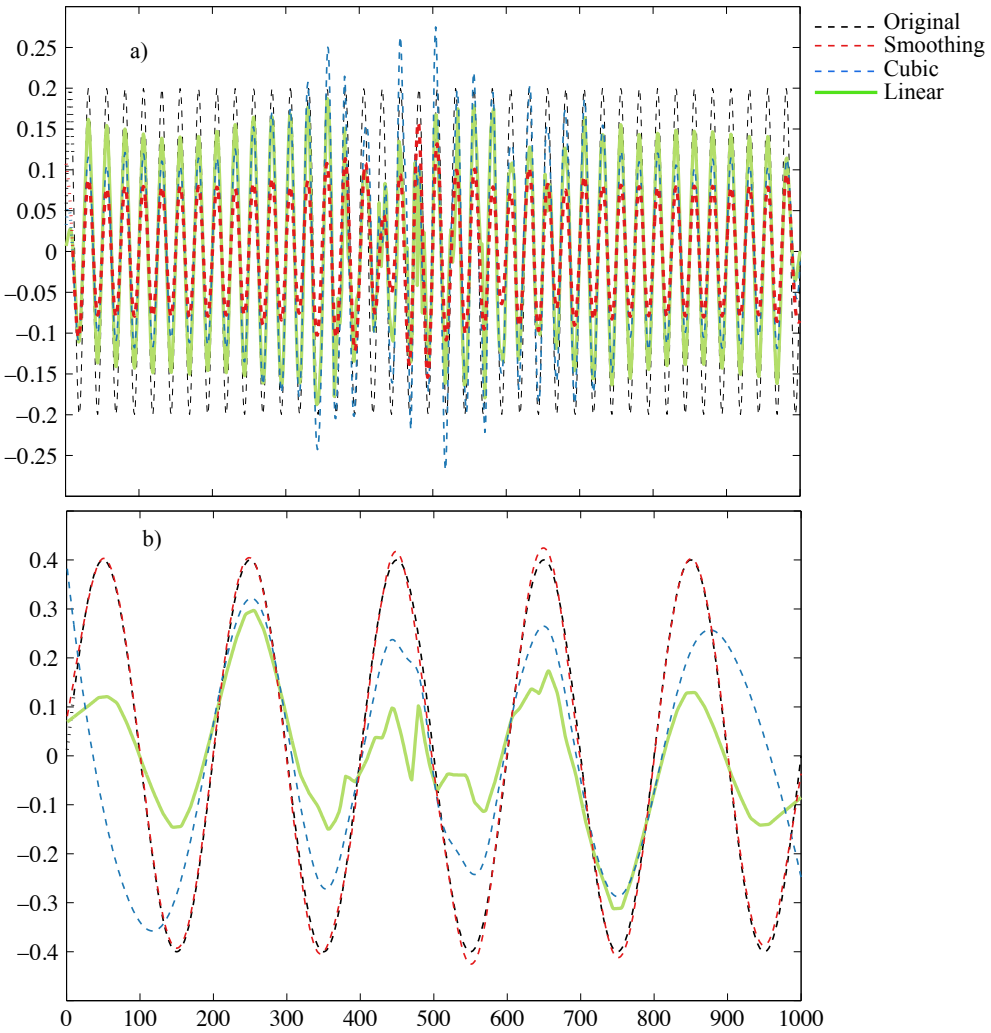

Figure 4.16: EEMD results for the non-stationary data with time-dependent variance, z_{06} (Equation 4.10).

	High-frequency mode			Low-frequency mode			Trend		
Data	**LEEMD**	**CEEMD**	**SEEMD**	**LEEMD**	**CEEMD**	**SEEMD**	**LEEMD**	**CEEMD**	**SEEMD**
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	36.4	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_{05}	50.2	41.4	77.3	167.8	941.0	16.5	NA	NA	NA
z_{15}	54.1	46.4	75.9	158.4	96.5	19	32.65	19.97	12.51
z_{25}	53.1	47.2	79.5	143.8	108.0	72.8	84.78	16.48	26.17
z_{35}	58.0	47.8	79.6	178.1	118.1	55.5	389.9	419.8	283.76
z_{45}	60.6	50.7	79.4	174.2	116.4	35.4	90.99	224.74	79.27
z_{06}	46.6	48.7	73.2	160.3	147.3	10.4	NA	NA	NA
z_{16}	46.9	48.5	74.8	146.9	150.4	12.0	45.9	12.08	16.2
z_{26}	51.4	46.2	76.2	162.3	135.6	21.9	104.102	36.109	23.78
z_{36}	48.1	49.0	77.9	171.2	193.8	75.7	319.77	388.4	189.25
z_{46}	50.0	49.2	76.4	184.7	177.6	25.5	124.8	99.6	74.8

Table 4.2: Mean absolute error (MAE) of the EMD (x_1–x_4) or EEMD IMFs ($\times 10^{-3}$) for all the data inputs. 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.

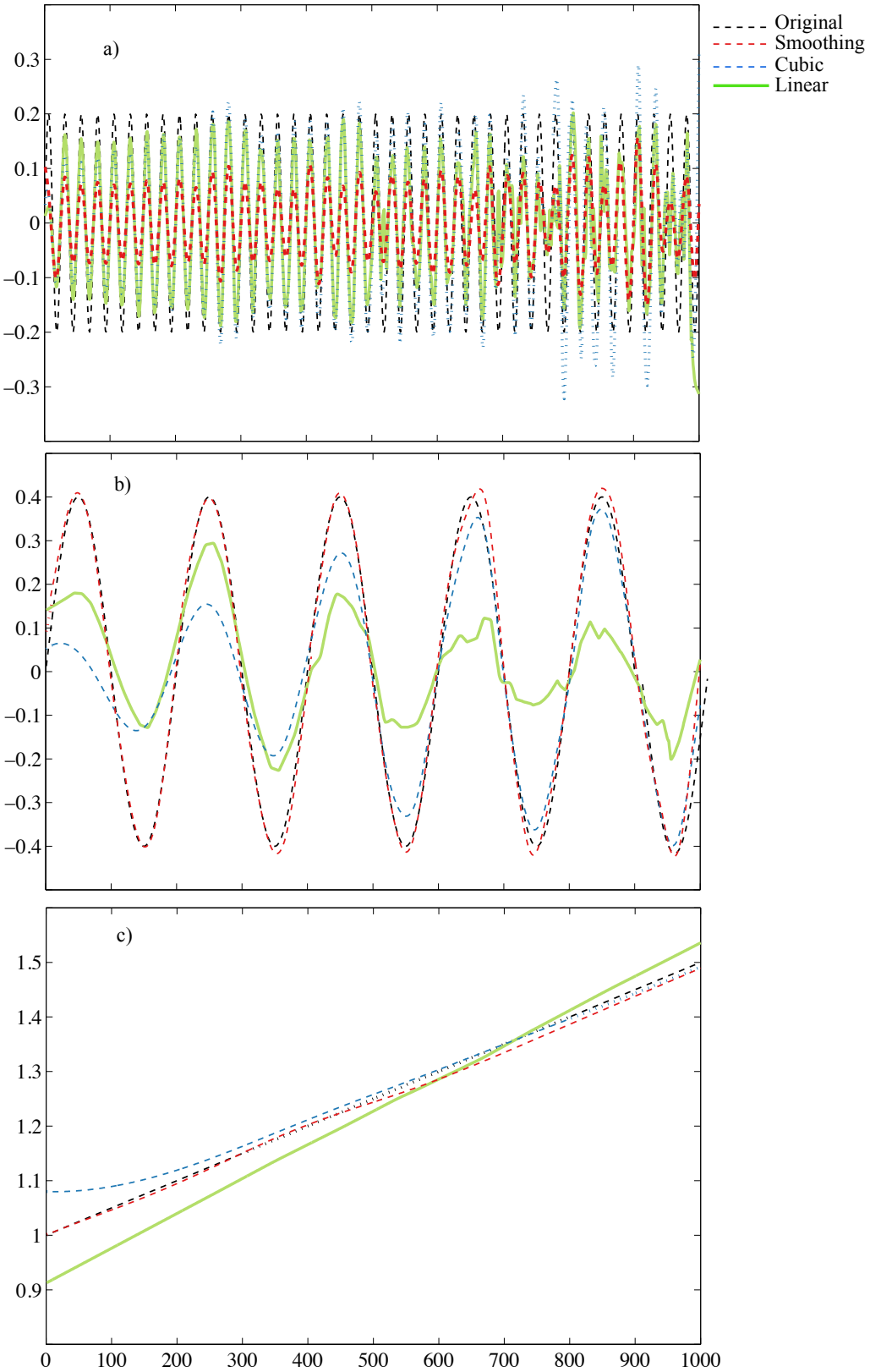

Figure 4.17: EEMD results for the non-stationary data with time-dependent mean and variance, z_{15} (Equation 4.13).

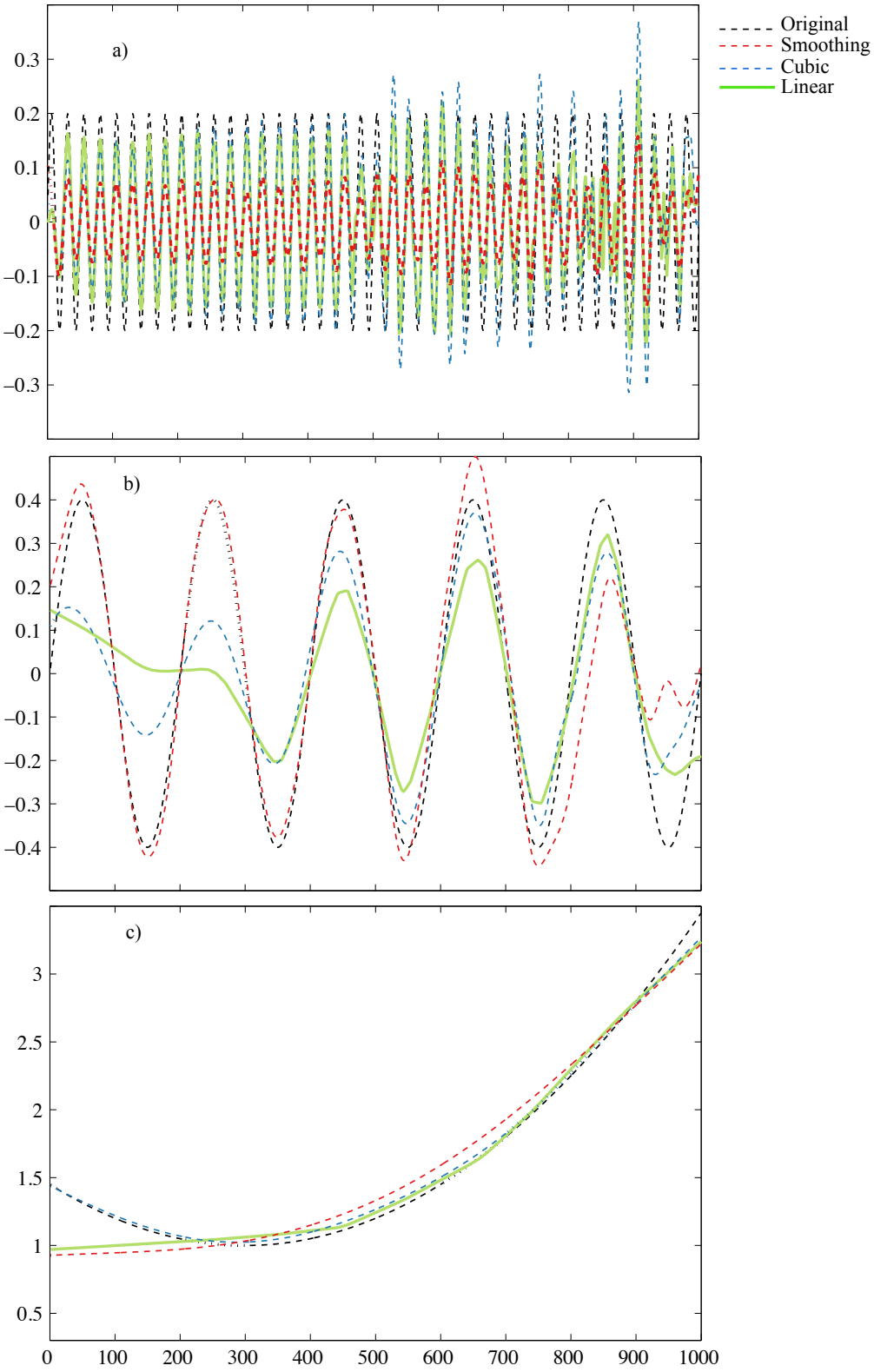

Figure 4.18: EEMD results for the non-stationary data with time-dependent mean and variance, z_{25} (Equation 4.13).

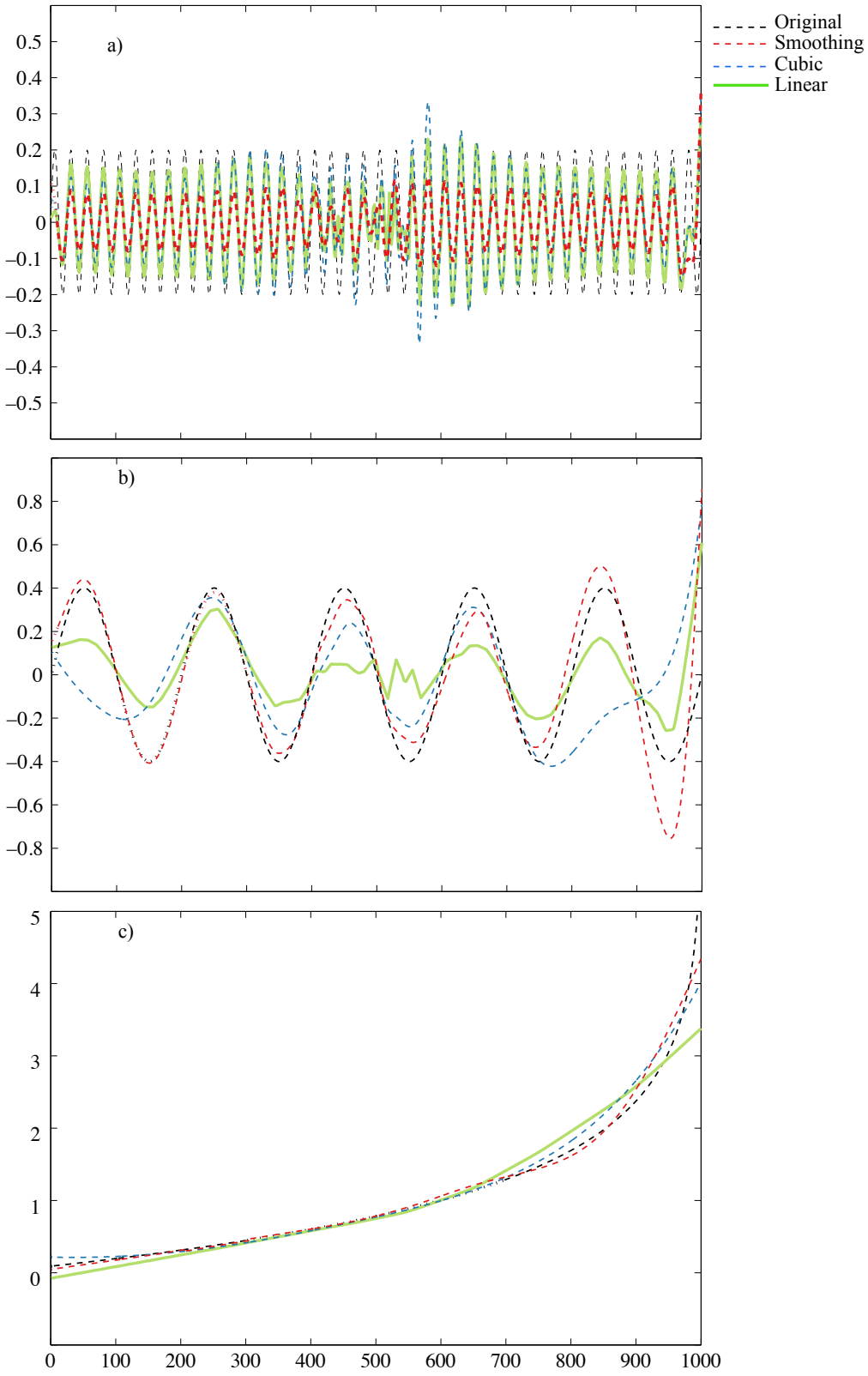

Figure 4.19: EEMD results for the non-stationary data with time-dependent mean and variance, z_{36} (Equation 4.13).

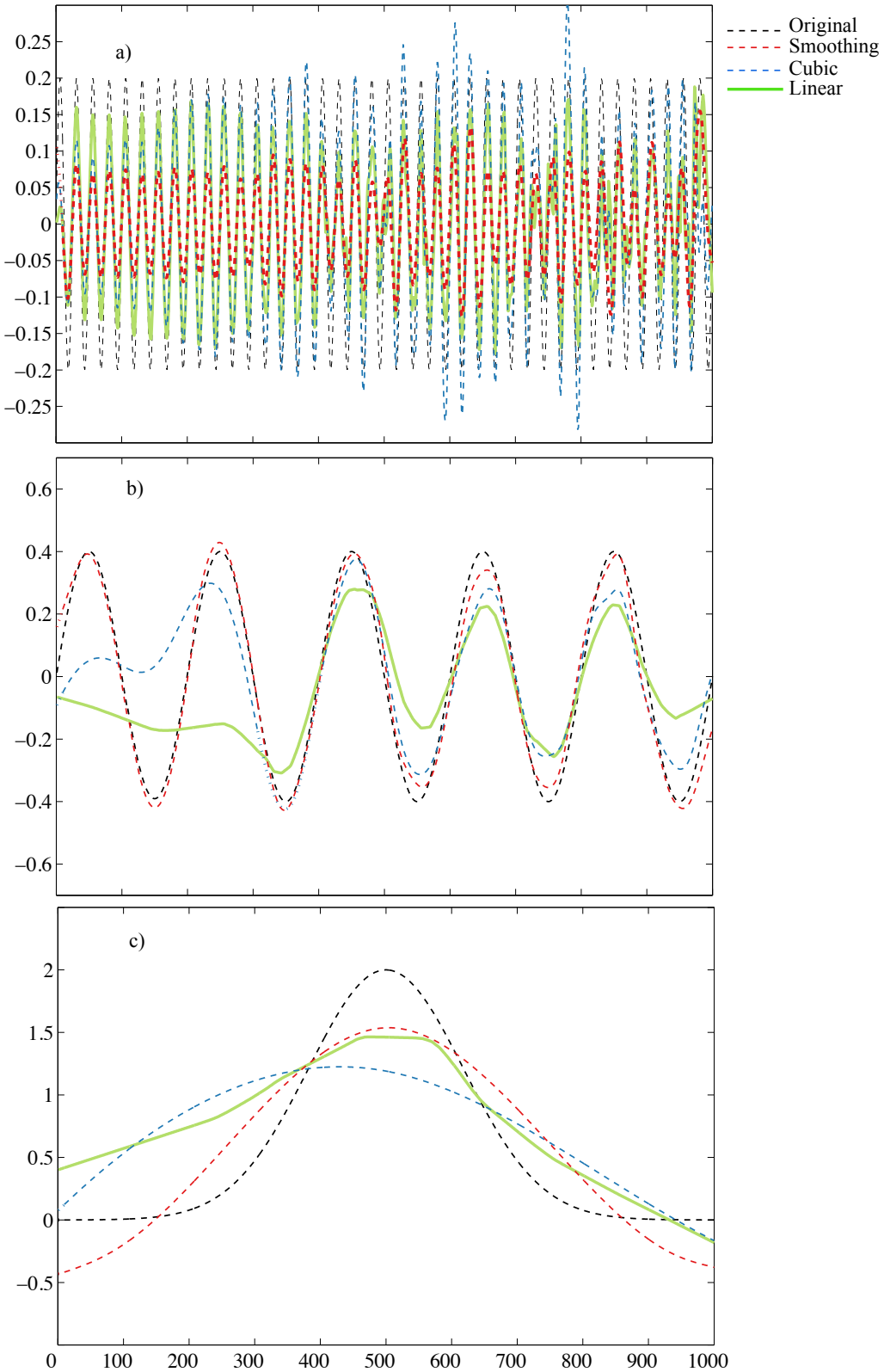

Figure 4.20: EEMD results for the non-stationary data with time-dependent mean and variance, z_{45} (Equation 4.13).

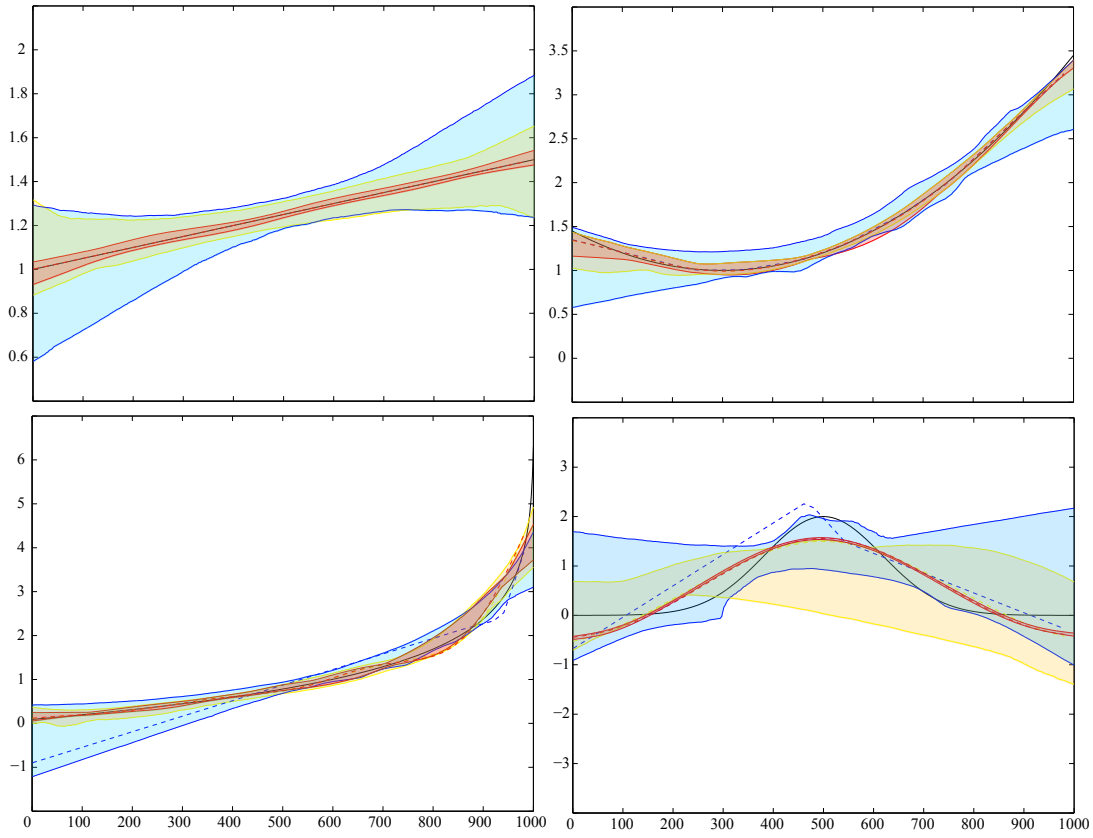

Figure 4.21: 95% confidence intervals for the trend results for the x_1 (top left panel), x_2 (top right panel), x_3 (bottom left panel) and x_4 (bottom right panel) data. Blue shading: linear EEMD; red shading: smoothing EEMD; and yellow shading: cubic EEMD.

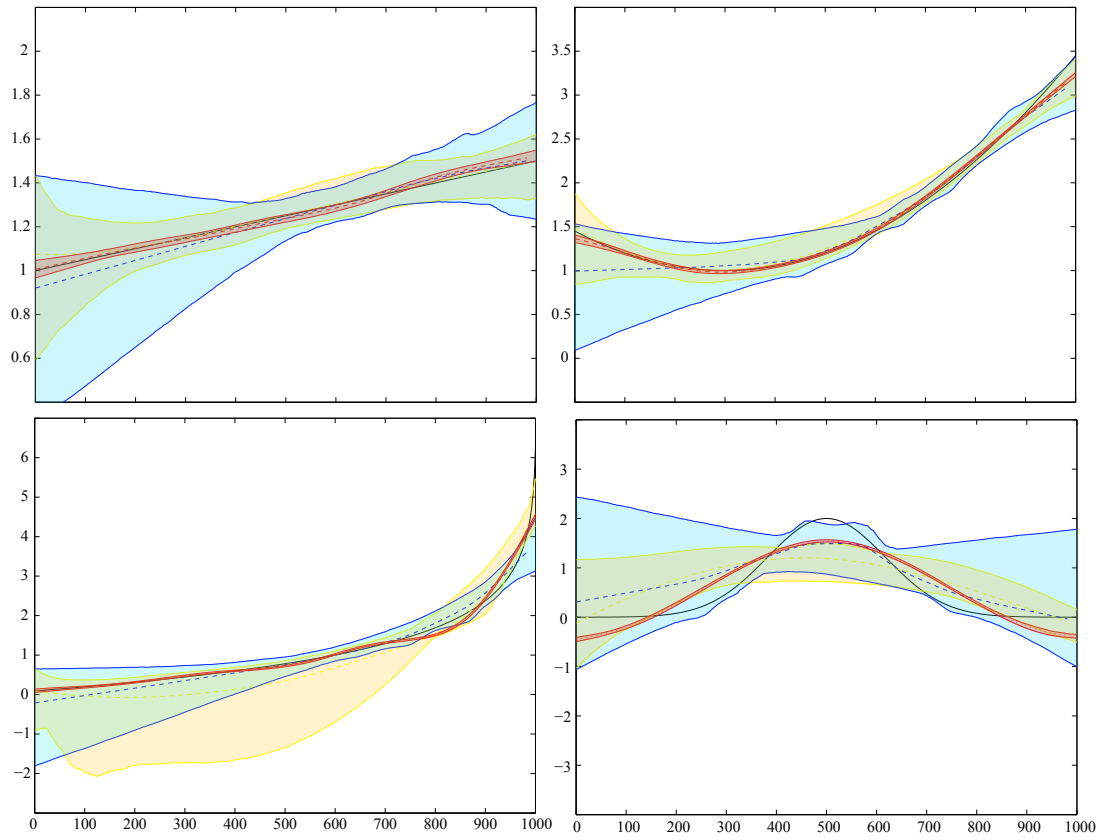

Figure 4.22: As for Figure 4.21 but for the trend results for z_{i5}, $i = 1, 2, 3, 4$. Blue shading: linear EEMD; red shading: smoothing EEMD; and yellow shading: cubic EEMD.

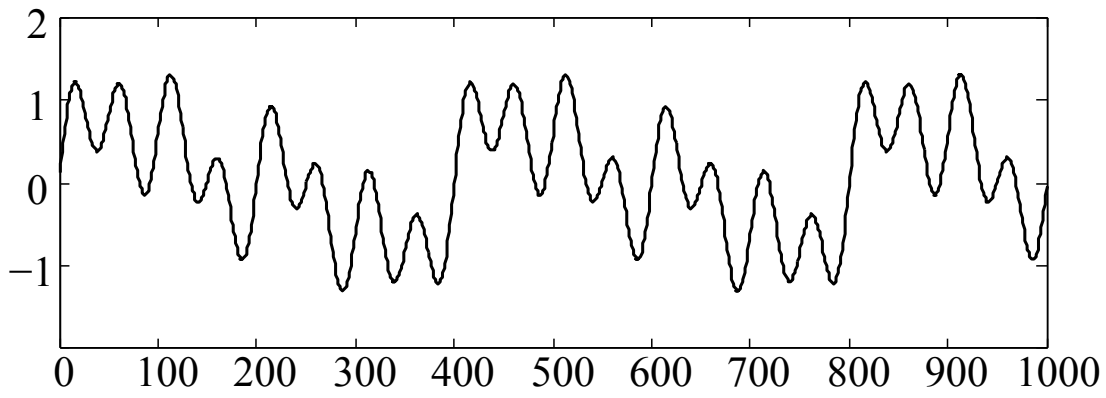

Figure 4.23: Input data $Y(t)$ (Equation 4.15).

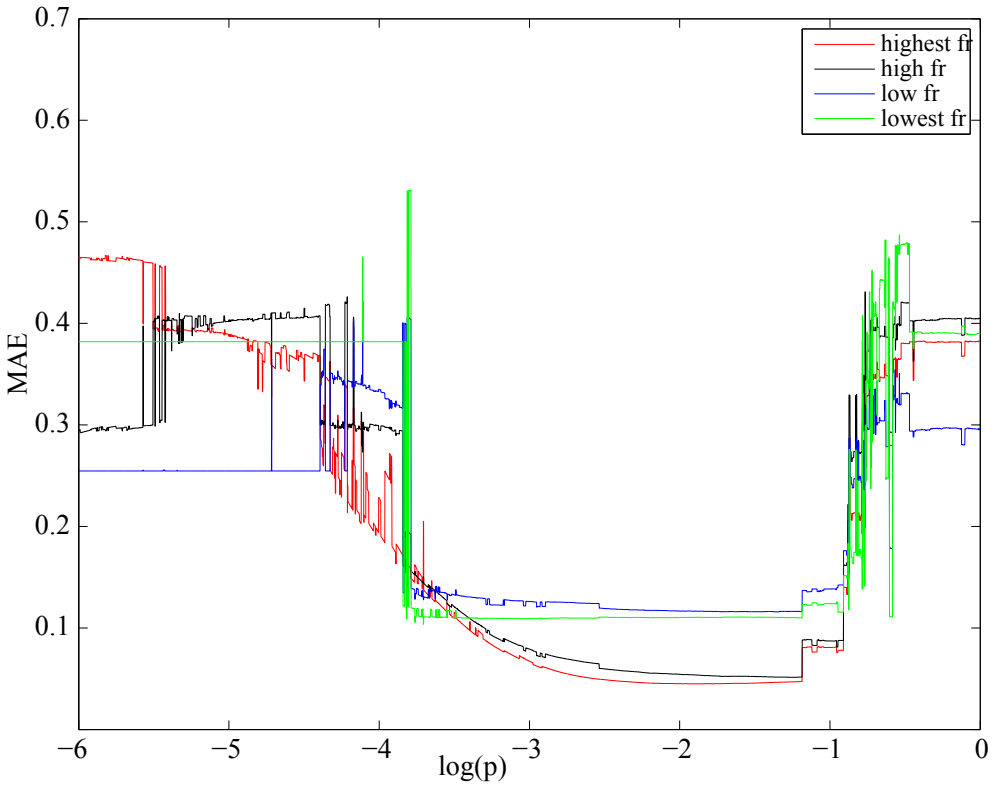

Figure 4.24: MAEs for the IMFs corresponding to the four sinusoidal components in Equation 4.16, as a function of the smoothing parameter p.

Chapter 5

Application of Empirical Mode Decomposition to Temperature Data

5.1 Introduction

Over the last century, global and local temperatures have risen significantly; many studies have been done in the last 50 years examining the significance of this warming. Coumou and Rahmstorf [2012] showed that the number of extremely high temperatures has risen significantly after 1980.

An increase in global temperatures also has follow-on effects on the environment, such as a rise in sea level [Rahmstorf, 2007], changes in precipitation [Trenberth, 2011a] and ocean circulation patterns [Levermann et al., 2005], and melting ice [Meehl et al., 2007]. Climate change has two types of causes, external forcing and internal variability in the climate system. External forcing can be a result of, for example, the shading of the Sun by aerosol pollution in the atmosphere from volcanoes [Neely et al., 2013] or from coal-fired power stations [Kaufmann et al., 2011]; the volcanic effect has only a short-term impact on the climate [Jones, 1994]. Another possible external forcing is a change in solar irradiation, but this has not changed sufficiently to explain the rise in the temperature in the last century [Crowley, 2000]. Other external forcings are the Earth's orbit around the Sun and, importantly, the addition of greenhouse gases to the atmosphere due to human activity (anthropogenic effect) [Trenberth and Fasullo, 2013, Mann and Park, 1994].

Trenberth [2011b] found small but significant and growing anthropogenic effects increasing global temperature. The other contributors to global temperature rise such as solar and volcanic effects are discussed by Ghil and Vautard [1991] and Foster and Rahmstorf [2011]. Ghil and Vautard [1991] said that a solar origin for the multi-decadal oscillation is unlikely for three reason. First, any insolation effects of the Sun's 22-year activity cycle should be dominated by the rectified 11-year period, which is absent in their analysis. Second, multi-channel signal spectrum

analysis of North American temperature data revealed no coherence with the 11-year insolation signal. Third, although the 0.1% amplitude of the solar irradiation cycle is sufficient to cause a variation of 0.1°C in the atmospheric temperature, it is definitely not sufficient to heat the upper ocean down to the penetration depth of a decadal signal. Foster and Rahmstorf [2011] showed that, because the effects of volcanic eruptions are very short-term and that of solar variability very small, neither of these factors can be expected to exert a significant influence on the continuation of global warming over the coming decades.

The strongest internal variability in the climate system is the change from El Niño to La Niña — a natural, stochastic 'seesaw' in the tropical Pacific called El Niño Southern Oscillation (ENSO). El Niño years are particularly warm, with 1998 the strongest El Niño year on record. During an El Niño, the ocean releases heat; during La Niña it stores heat. Randel et al. [2009] used a whole-atmosphere chemistry climate model to study the signature of ENSO in the global temperature. They found a strong ENSO signal in zonal mean temperature and ozone in the tropical lower stratosphere in both the model and observational datasets. They also noticed that there can be confusion of the ENSO and volcanic signals in short time series. Therefore it is more helpful to use longer time records to analyse decadal and multi-decadal temperature variability. Thompson et al. [2009] showed that land and ocean coupled temperature fluctuations have a complex spatial signature. These fluctuations penetrate the Atlantic Ocean, with larger sea-surface temperature amplitude than other locations. Hansen et al. [2006] argued that warming in the western equatorial Pacific was larger than in the eastern equatorial Pacific in the last century. This increase in the east-west temperature gradient may increase the likelihood of stronger El Niños.

Changes in atmospheric and sea-surface temperatures impact the ocean and the weather globally and regionally on both long and short timescales. To study the long-term variability in the temperature, it is necessary to use a long data record and subtract the underlying ENSO variability. Hulme et al. [2002] provided information on climate change over time intervals of 20, 50 and 80 years, and demonstrated that changes in climate state variables such as the warming of central England and surrounding coastal waters only showed directional change over a period of at least 50 to 100 years.

Foster and Rahmstorf [2011] analysed five global temperature time series (over land and ocean) from 1979 to 2010, three surface temperature records and two lower-troposphere temperature records based on satellite microwave sensors. All five results showed consistent global warming trends, ranging from 0.014°C to 0.018°C per

year. When the data were adjusted to remove the estimated impact of known factors on short-term temperature variations (El Niño/Southern Oscillation, volcanic aerosols and solar variability), the global warming signal became even more evident as the noise was reduced. The lower troposphere was affected by these factors more strongly than the sea surface temperature.

Gaining a better understanding of temperature variability requires a method that can more accurately extract information from the data. Some studies have already shown that data-adaptive methods such as the Empirical Orthogonal Function or Empirical Mode Decomposition/Hilbert-Huang Transform are suitable for this [Wu et al., 2011, Mann et al., 2000]. These methods decompose the data into components with different time scales, which relate to different temperature forcings.

As mentioned in Chapter 2, the Empirical Orthogonal Functions (EOF) Method requires stationary data. Recent global warming has changed the mean temperature and standard deviation, therefore adding more non-stationarity to the temperature data. Jones et al. [1999] showed that all monthly, seasonal and annual series are poorly approximated by linear trends due to two periods of warming from 1920 to 1940 and from the mid-1970s to 1999. Over both periods, the warming was slightly greater in the southern hemisphere than in the northern hemisphere. Dealing with such data is challenging, and requires a method that can deal with both non-linearity and non-stationarity. Molla et al. [2006] used EMD to analyse a 15-year air-temperature dataset with temperature measured daily. They used orthogonality and a normal-distribution check of the modes, and reconstructed the modes to test accuracy. However, 15 years of data are insufficient to investigate the climate-change influences on temperature, and orthogonality does not necessarily provide a good measure of the accuracy of the results [Bahri and Sharples, 2015].

The question is whether the EMD method and its extension, Ensemble Empirical Mode Decomposition (EEMD), are suited to dealing with the non-stationarity in the temperature data. If so, do the results change significantly using different interpolation methods? The EMD method was shown to be sensitive to the interpolation method in Chapter 4. Whether this sensitivity can affect the analysis of temperature data is investigated in this chapter.

In this chapter, unlike Chapter 4, we do not know the actual modes. Therefore we cannot directly compare the EEMD results with the actual modes. In order to make a judgment about which interpolation method gives the better results, we compare our results with other studies and obvious physical phenomena, and see if using different interpolation methods changes the detailed information that we extract from the EEMD method.

5.2 Temperature data

In this chapter, five temperature datasets are studied. The first dataset is the land-sea mean global monthly temperature index obtained from NASA's website (*data.giss.nasa.gov/gistemp/*). This dataset is the combined Land-Surface Air and Sea-Surface Water Temperature Anomalies (Land-Ocean Temperature Index, LOTI), from 1880 to 2015. The accuracy of this dataset has been improved several times. Quayle et al. [1999] combined long-term ship data, and buoy and satellite data to provide the greatest possible coverage of the sea surface temperature. The result is a global, century-scale surface-temperature index, updated monthly, that closely parallels other widely published global surface-temperature measurements. To combine these data into a single index, the land-air surface temperature was weighted with a coefficient of 0.3, the sea surface temperature 0.7. Smith and Reynolds [2005] investigated errors in the merged land-air/sea-surface temperature data and reconstructed the data in two separate parts: a low-frequency component and a high-frequency component: the low-frequency modes are anomaly modes, and the high-frequency modes seasonal and inter-annual modes. They found that anomaly persistence was stronger in SST than in land-surface temperature (LST); if a land station disappears, it has a effect on the LST analysis. The global ocean area is roughly twice as large as the global land area, so the impact of the SST overpowers the impact of LST. Land variations are often more complicated than sea-surface variations because of different elevations and land types.

Smith and Reynolds [2005] merged the two datasets by weighting the SST and LST components by the percentage of ocean and land area in each $5° \times 5°$ region. Errors arose from historical gaps in sampling and from bias uncertainties in the data. The errors in their study are due to the simple methods used for the low-frequency analysis; the low frequencies in their study were only resolved if there were sufficient local data for a large-scale ($10° \times 10°$ or $15° \times 15°$) average and for low frequency (15-yr) filtering.

Another study of the National Oceanic and Atmospheric Administration (NOAA) which merged land-ocean surface temperature from 1880 to 2006 was done by Smith et al. [2008]. They used the decadal-scale component to obtain the LF component, they averaged spatially over $15° \times 15°$ moving areas and then annually, to remove local anomalies. The smoothed annual averages were then median filtered using 15 annual averages to produce a 15-yr LF anomaly analysis. The high-frequency (HF) observations were obtained from a combination of the interpolation of SST measurements over the oceans and from the global historical climate network over the land. Smith et al. [2008] found that uncertainty occurred because of data errors

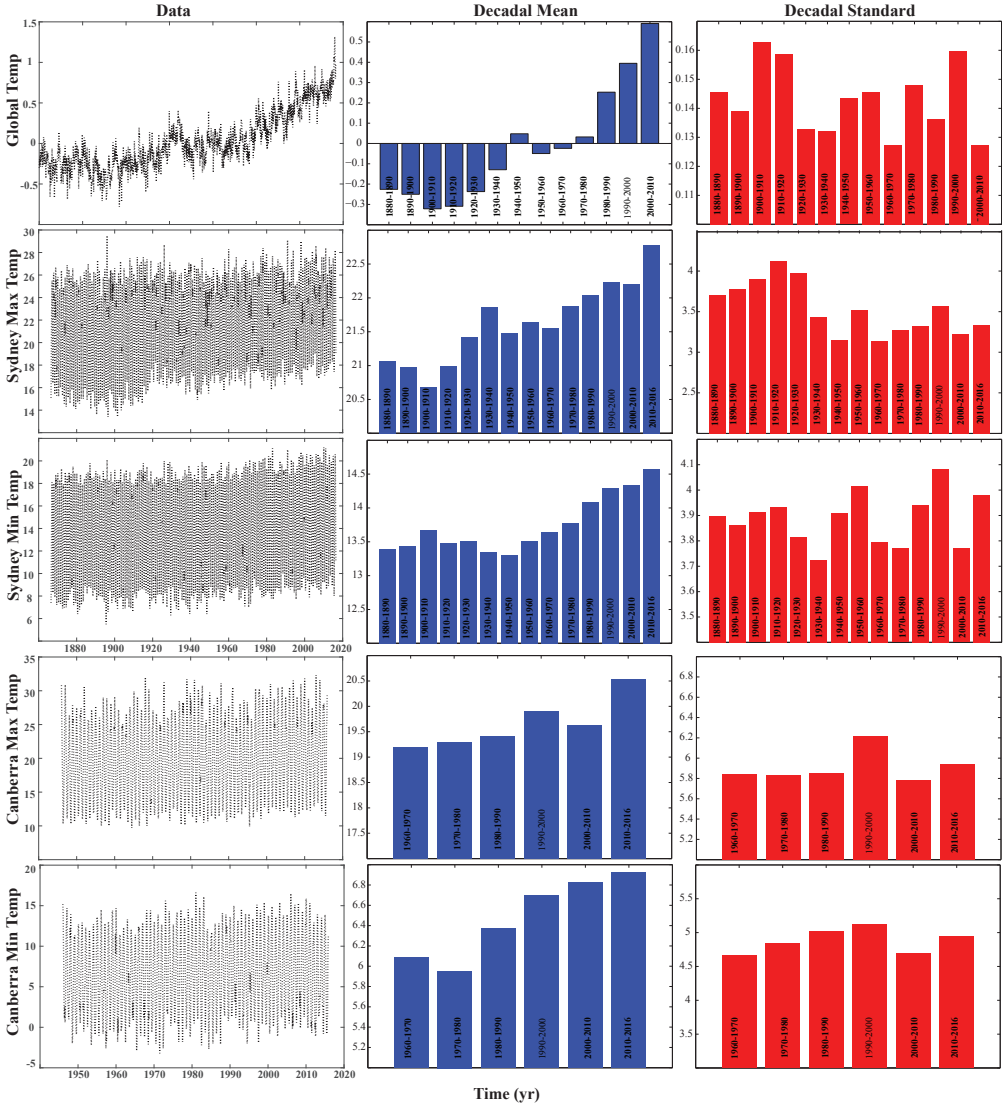

Figure 5.1: Monthly mean temperatures and the time variability of their mean and standard deviations. The left column of panels shows the data, the middle column the decadal mean and the right column the decadal standard deviations. The top row of panels shows the land-sea mean global monthly temperature index, the second row the Sydney, Australia monthly maximum temperature and the third row the Sydney monthly minimum temperature. The two bottom rows, with a different time scale, show the Canberra monthly maximum and minimum temperatures, respectively.

and incomplete sampling over the historical period. The improvements in this study, especially in late nineteenth century and since 1985, allowed a better understanding of the global temperature. The improvements in the late the nineteenth century were due to improvements in the analysis and data-collection methods.

The other datasets used here are the Sydney monthly maximum and Sydney monthly minimum temperatures, both from 1859 to 2015. These data were obtained from the Australia Bureau of Meteorology (*www.bom.gov.au/climate/data-services/*). The data were collected initially at a station on Sydney Observatory Hill, latitude 33.86°S

and longitude 151.21°E, at 39 m above mean sea level. Historically, if an observational site moved a relatively short distance (within about 1 to 2 km), it may still have used the same station number. Numerous changes have also occurred in instrumentation and/or observing practices over the period included in the dataset, which may have had an effect on the long-term record. In recent years, many stations have had observers replaced by automatic weather stations, either completely or at certain times of the day. The Sydney record started in January 1859. The first station on Observatory Hill was about 0.5 km from the harbour. In 1917, the weather station moved about 100 m downhill to the south to a 6 m lower elevation. From 1922 to 1946, the population of the area grew significantly. These changes might have contributed slight variations to the temperature records.

The final two datasets are the Canberra monthly maximum and minimum temperatures from 1940 to 2015, obtained from the Bureau of Meteorology website (*www.bom.gov.au/climate/data-services/*). The station is at Canberra Airport, latitude 35.30°S and longitude 149.20°E, at 578 m above mean sea level. The Canberra data came in two files, one called Canberra Airport comparison from 1939 to 2010, the other Canberra Airport from 2008 to 2016.

In this section, the temperature data are analysed to determined whether they are non-stationary, and so, what type of non-stationarity they have. The data are shown in Figure 5.1. To find if the data are non-stationary, a test is applied to find whether the decadal mean and/or decadal variance of the data vary with time. In order to do that we calculated the mean and variance of the data for each decade. The results are shown as bar charts in Figure 5.1.

Figure 5.1 shows that there was an overall increase in the global mean temperature (top row in Figure 5.1), with a sharp rise after 1980. The decadal variance in the global monthly temperature fluctuated more after 1940. The mean and the variance of the global temperature data are time-dependent, so these data are non-stationary; the mean changed non-monotonically from 1880 till 2016, with a slight bump around 1945. Although the changes in the variance over the whole time period were small, the fluctuations of the variance in the period 1880–1940 were different from those in the period 1940–2015. This shows not only a time dependence in the variance of global temperature but that the change in variance was irregular over time.

The Sydney monthly maximum temperature (second row in Figure 5.1) showed a roughly constant increase in decadal mean, but also a local peak around 1940. The decadal variance showed a sharp decrease in 1920s and remained low after that. The Sydney monthly minimum temperature (third row in Figure 5.1) dropped in 1940, then increased up to 2015. Its variance had a greater fluctuation after 1940.

The increase in both the minimum and maximum temperature shows that Sydney became warmer. The Sydney minimum temperature had a quadratic-like trend, with regular variations in the variance; the Sydney maximum temperature trend was more linear, with decreasing variance.

The Canberra, Australia monthly maximum and minimum temperature records both started after 1940, (the bar charts show the mean and variance variability after 1960) and showed an increase in the period after 1980. After 1985, the rate of decadal increase was gentle. However, due to the short time span of the data, the study of the longer time variations was harder. The variance of both the minimum and maximum temperatures peaks in 1990–2000 and then decreases after 2000.

5.3 EEMD analysis of temperature data

In this section, the results of using two different interpolation methods in EEMD on temperature data are presented; statistically significant modes are obtained. The EEMD method is used to avoid mode mixing effects in the results. It was shown in Chapter 4 that EMD/EEMD with linear interpolation had a poor performance. Therefore, in this chapter only the EEMD with cubic interpolation (CEEMD) and smoothing spline interpolation (SEEMD) is used. When using SEEMD, the smoothing parameter is set to $p = 0.001$, chosen according to the results in Section 4.4.

The IMFs from the EEMD analysis are shown in Figure 5.2 for the global temperature data. SEEMD mode 1 was equivalent to CEEMD modes 1,2 and 3 combined, SEEMD had an extra low-frequency mode with very small amplitude.

The method produces a number of modes (Figure 5.2), not all of which carry statistically significant information. Wu and Huang [2004] introduced the following method to test for statistically significant modes. First, decompose the targeted normalised noisy data into IMFs. Second, construct a long artificial white-noise record with the same standard deviation as the first mode as a reference. This reference record is then divided into sections of identical length (we used 100 sections, each with the same length as the time series). Next, apply the EEMD method to decompose each reference section into IMFs; the statistical characteristics of these IMFs should have energy $\times$ period (E_nT_n) constant for each section 'n'. Therefore, $\ln(E_n)+\ln(T_n) = 0$, so that the graph of $\ln(E_n)$ against $\ln(T_n)$ will be a straight line (for more information see Wu and Huang [2004]). A number of IMFs are produced for each reference section; the mean of each IMF over all sections is then obtained. By specifying the confidence interval (95% in this study), upper and lower bounds for the IMFs are determined. The spread lines are the 2.5 and 97.5 percentiles. The IMFs with energy above the upper spread line or below the lower spread line are

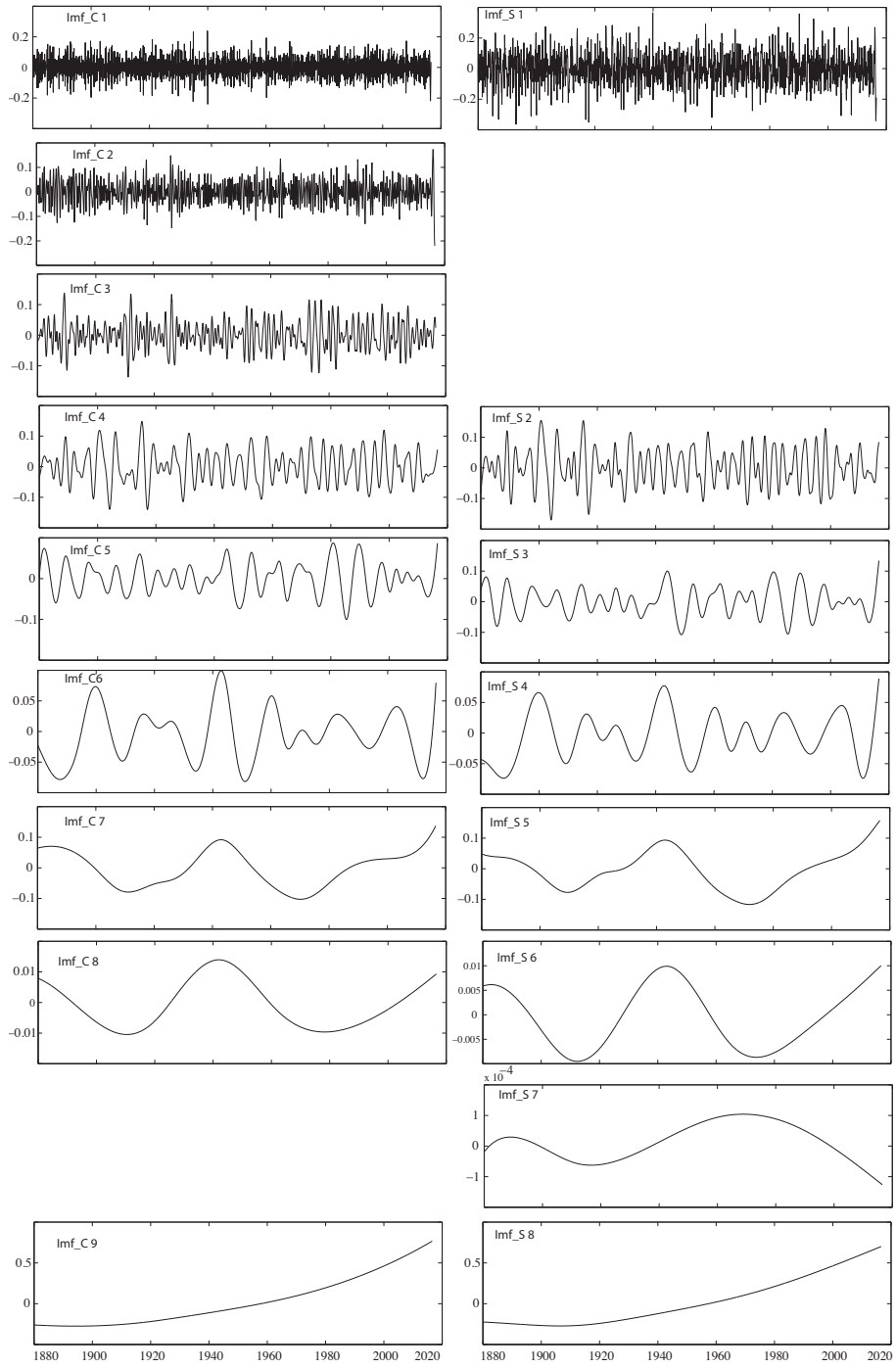

Figure 5.2: IMFs of the global land-sea mean monthly temperature index data: cubic-spline interpolation CEEMD (left column); smoothing-spline interpolation SEEMD (right column), high frequency at the top; the bottom IMFs is the residual or trend.

considered to contain information at that selected confidence level and are statistically significant (Monte Carlo tests). However the IMFs with energy lower than the lower spread line will have very low energy and do not play a significant role. Therefore, in this study only the IMFs above the upper spread lines are considered

as statistically significant IMFs. This method is used here on all temperature data. Figure 5.3 shows a posteriori and Monte Carlo tests for the data.

Figure 5.3 shows that, for global temperature, CEEMD gave seven statistically significant IMFs whereas SEEMD gave six statistically significant IMFs. For Sydney maximum temperature, SEEMD extracted four significant IMFs whereas CEEMD only extracted three. This shows that changing the interpolation in the EMD method, can change the total number of statistically significant IMFs.

		Period		
Dataset	**Interpolation**	**Intera-decadal**	**Decadal**	**Multi-decadal**
Global	Cubic	mode 5	mode 6	mode 7
	Smoothing	mode 3	mode 4	mode 6
Sydney Max	Cubic	mode 3	none	none
	Smoothing	mode 3	mode 4	none
Sydney Min	Cubic	none	none	none
	Smoothing	mode 3	mode 4	none
Canberra Max	Cubic	none	none	none
	Smoothing	none	none	none
Canberra Min	Cubic	none	none	none
	Smoothing	mode 4	none	mode 7

Table 5.1: Statistically significant IMFs (modes) for global, Sydney maximum and minimum, and Canberra maximum and minimum temperature data. In this table the blue shows the modes from CEEMD, and red the modes from SEEMD.

This situation applies for Sydney and Canberra minimum temperatures (Figure 5.3). In addition, the local temperature is influenced strongly by high-frequency drivers such as seasonal variability. Therefore the long-term low-frequency variables have smaller amplitudes, and mode mixing can result in IMFs containing relevant information being discarded as insignificant.

Ghil and Vautard [1991] used singular spectrum analysis (see Section 2.3.4) to analyse the time series of global annual surface air temperatures from 1854 to 1988. They found fundamental oscillations that were associated with peaks near 21 years, 6 years and 5 years. In this study we found fundamental oscillations similar to Ghil and Vautard [1991]. In Table 5.1, the statistically significant IMFs are listed according to their time scale. intra-decadal IMFs are IMFs with a period of 4–6 years, decadal IMFs a period of 10–20 years and multi-decadal IMFs a period of more than 30 years. As shown in Table 5.1, SEEMD extracted more statistically significant IMFs with lower frequencies; therefore SEEMD is a more suitable interpolation method if we are interested in studying the long-term variability. For Sydney minimum and maximum temperatures, smoothing splines gave a statistically significant decadal

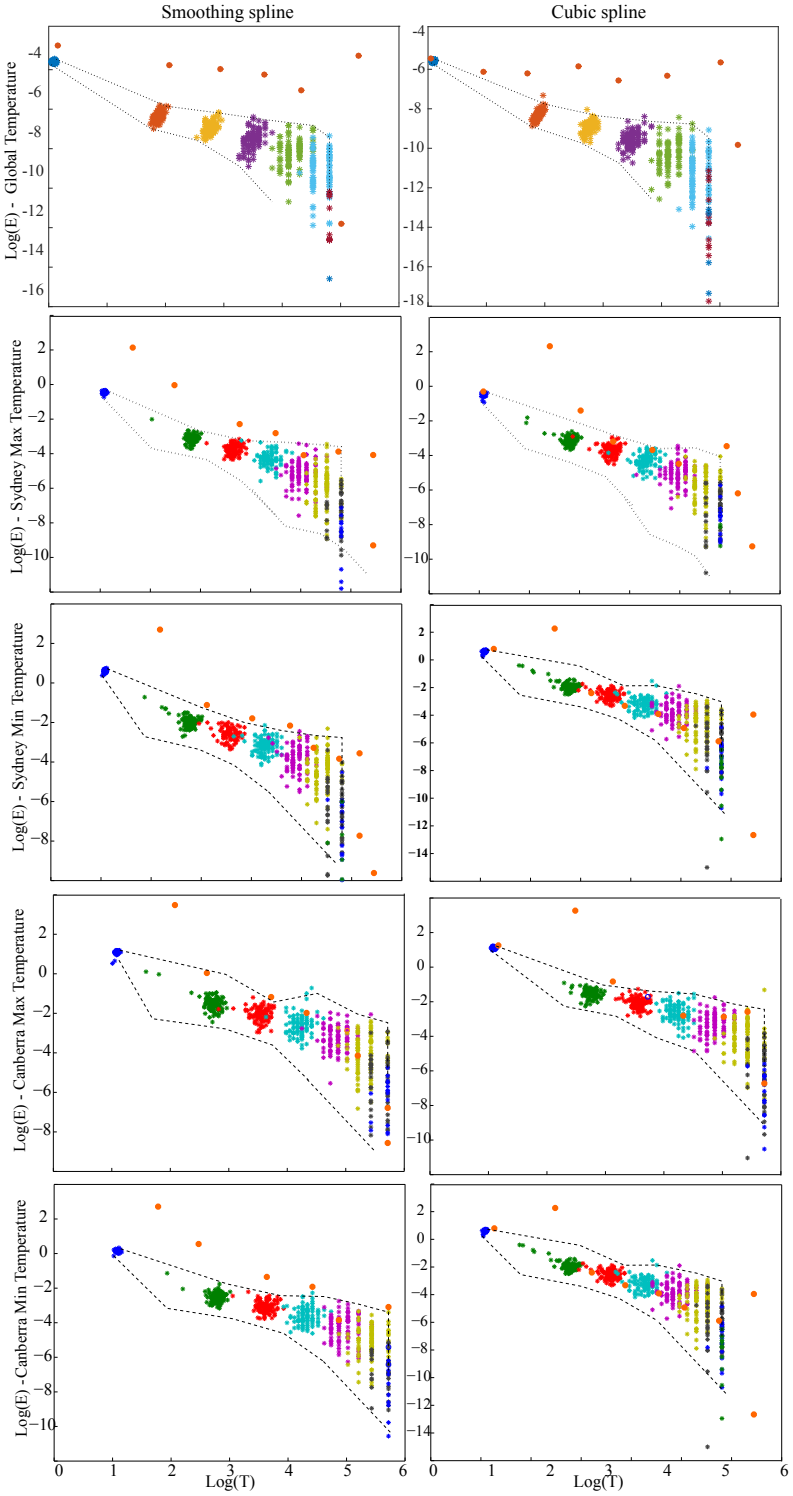

Figure 5.3: Results of the a *posteriori* and Monte Carlo tests for the five temperature datasets. The dotted lines show the upper and lower spread lines determined from the energy densities of the reference section IMFs (coloured dots). The orange circles show the energy density of the original IMFs. Those orange circles that lie above or below the spread lines correspond to statistically significant modes.

IMFs whereas cubic splines did not. Similarly for the Intra-decadal IMFs for the Sydney and Canberra minimum temperatures and for the multi-decadal IMFs for Canberra minimum temperature.

In Mann and Park [1994], the spatial patterns associated with significant peaks in the variance within a broad period of 2.8 to 5.7 years exhibited characteristic ENSO patterns. A decadal mode in the 15–18-year period appears to represent long-term ENSO variability. Mann and Park [1994] showed that a significant proportion of climatic variability on inter-annual to century timescales may be associated with quasi-periodic processes of either external or internal origin. Organised internal variability appears to be associated with ENSO or with extra tropical patterns that chiefly involve a North Atlantic Oscillation pattern. Decadal variability has a weak impact on global average temperature but gives rise to a strong redistribution of surface heat. The 15–18-year intra-decadal signal related to ENSO has a significant amplitude and projection onto global average temperature. This signal contributed to the anomalous warmth of the 1980s. Smith and Reynolds [2005] removed noise by fitting the data to a set of covariance modes, based on a densely sampled period. They found that for the period 1880–1997, the number of statistically significant LST modes increased gradually until about 1950. After 1950, the number of LST modes remained the same, with a slight decrease in the 1990s. The number of statistically significant SST modes remained constant for most of the twentieth century.

5.4 Interpolation sensitivity – application to temperature data

In Figures 5.4–5.7, the statistically significant IMFs in Table 5.1 at each of the three time scales, intra-decadal, decadal and multi-decadal, are shown.

Figure 5.4 shows the global and Sydney maximum temperature Intra-decadal temperature modes. The top panel shows the intra-decadal modes for global temperature using cubic and smoothing splines. The results are similar except for opposite trends in 1886, 1945 and 1990. The bottom panel in Figure 5.4 shows the corresponding results for Sydney maximum temperature. Here, there are more differences between the cubic-spline and smoothing-spline results. The amplitude of the CEEMD IMF curve is significantly less than that of SEEMD for the first half of the data, because of mode mixing in the cubic-spline results. In the second half of the data, there is more similarity between the SEEMD and CEEMD results. Neither of the interpolation methods found significant intra-decadal modes for Canberra maximum temperature, and no statistically significant modes were found for Sydney minimum temperature using CEEMD. The results for Sydney and Canberra minimum temperature using SEEMD are shown Figure 5.5. The IMFs show similar

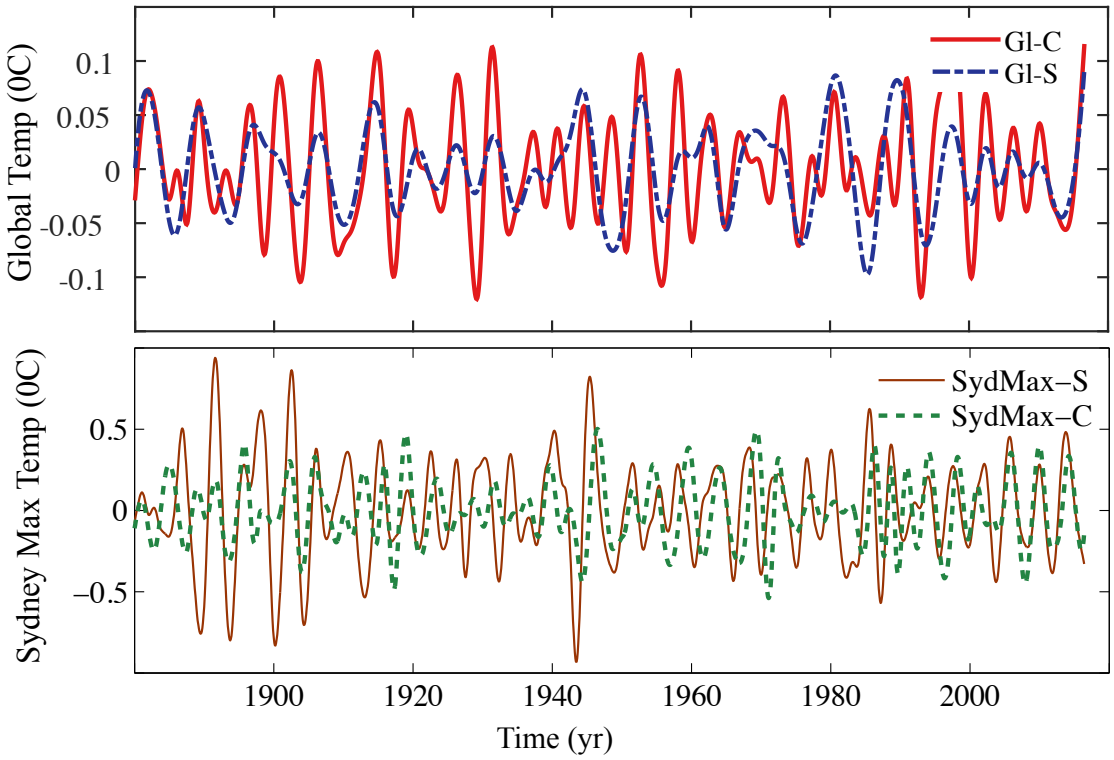

Figure 5.4: Statistically significant intra-decadal IMFs for global temperature and Sydney maximum temperature data. S: smoothing-spline interpolation; C: cubic-spline interpolation.

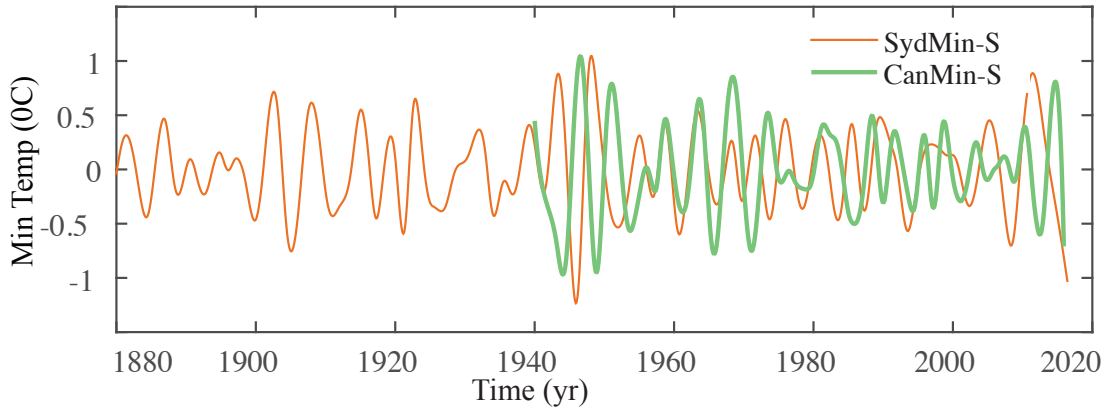

Figure 5.5: Statistically significant intra-decadall IMFs for global Canberra and Sydney minimum temperature data. S: smoothing-spline interpolation.

behaviour, with a some lag between the Canberra and Sydney intra-decadal results. This lag could be due to the different lengths of the Canberra and Sydney data.

Figure 5.6 shows the decadal IMF for global temperature with CEEMD and SEEMD. The two methods gave similar results for the years 1900 to 1940, but the SEEMD

results showed more peaks in the years 1940–2000. These peaks are consistent with the timing of El Niños in 1972, 1983, 1987 and 2003.

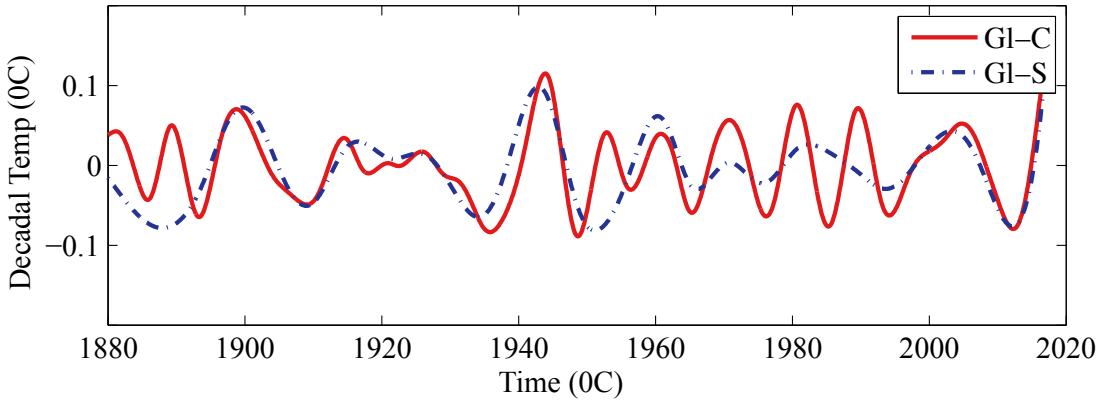

Figure 5.6: Statistically significant decadal IMFs for global and Sydney temperature data. S: smoothing-spline interpolation; C: cubic-spline interpolation

Figure 5.7 shows statistically significant IMFs with multi-decadal periods. Table 5.1 shows that only three IMFs with multi-decadal period across all the temperature datasets were statistically significant. Therefore, in Figure 5.7 just the multi-decadal IMFs from SEEMD and CEEMD global temperature and multi-decadal SEEMD Canberra minimum temperature are plotted. The two EEMD methods gave similar IMFs for the multi-decadal period for global temperature, with small differences after 1960. In this figure, there is a phase difference between Canberra minimum temperature and global temperature. This makes sense because the global temperature is the average of land and sea temperature, and the fluctuations in sea temperature are delayed relative to those in land temperature. Unfortunately, due to the short record for Canberra temperature, its multi-decadal behaviour from before 1940 is unknown. Therefore, it is difficult to distinguish global climate change effects from the local climate variability. Local climate variability can occur over sub-decadal to decadal scales, while climate change effects are only discernible in records longer than 50 years approximately. As such the Canberra time series may not be able to properly distinguish the local climate changes from those associated with global climate change effect, and is required in interpreting the results.

In the EEMD method, the residual is the IMF that shows the long-term variability in the data. Here, the long-term variability in the temperature data is related to climate change. Many studies have shown that climate change cannot be studied successfully with short records. Therefore, the trends in the Canberra temperature data, because of their short time records, probably will not give us realistic information about climate change.

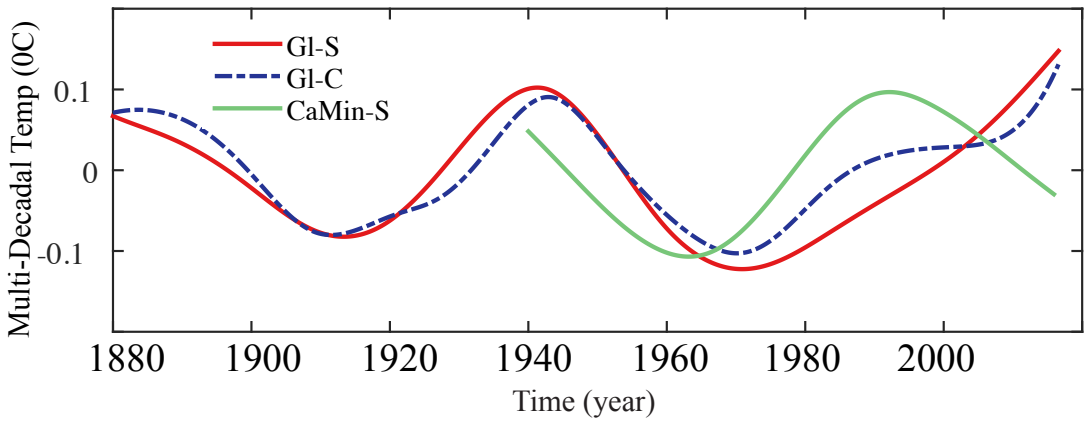

Figure 5.7: Statistically significant multi-decadal IMFs for global and Canberra minimum temperature. S: smoothing-spline interpolation; C: cubic-spline interpolation.

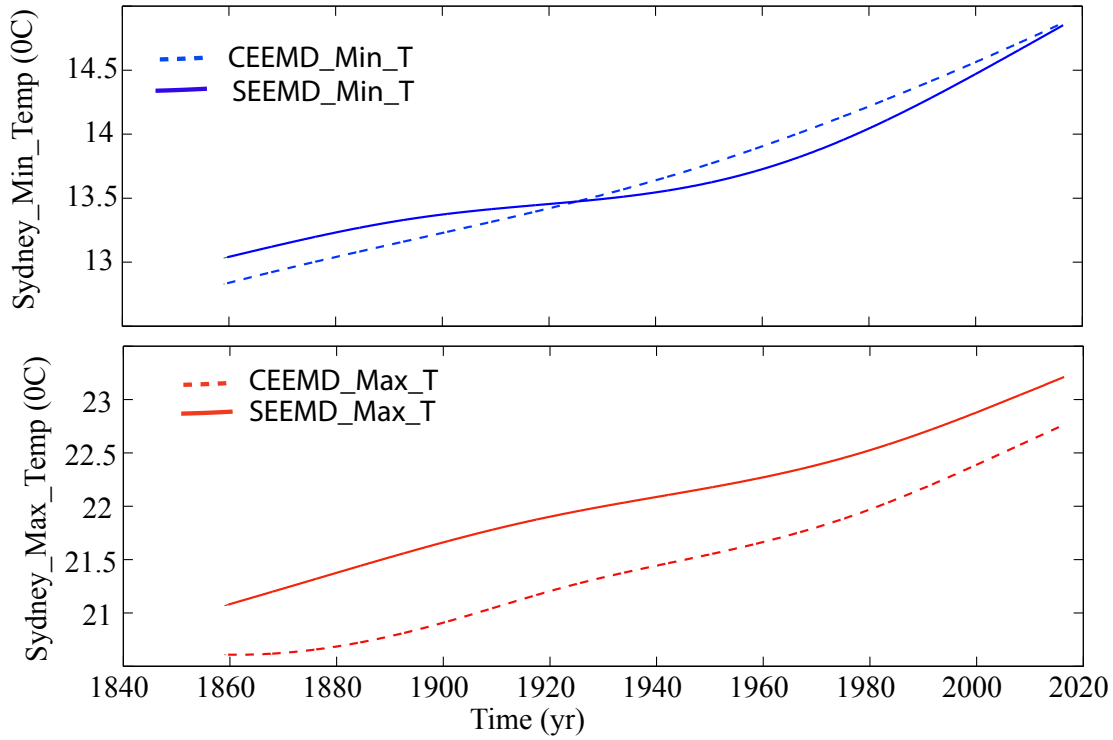

Figure 5.8: Sydney maximum and minimum temperature anomalies using CEEMD and SEEMD.

Figure 5.8 shows the trends (residuals) in the Sydney minimum (top) and maximum temperature (bottom) anomalies using SEEMD and CEEMD. This figure shows further evidence of the sensitivity of EEMD to the interpolation method. The residuals of the maximum and minimum temperature using SEEMD have different shapes, whereas those using CEEMD are more or less the same shape. SEEMD indicates that the minimum temperature had an acceleration after 1920 whereas

CEEMD indicates that its rate of increase was approximately constant during the whole time period.

According to the SEEMD results, over the whole period 1859–2015, Sydney maximum temperature increased from 21.1°C to 23.2°; CEEMD gave this increase as from 20.6°C to 22.8°C. For minimum temperature, SEEMD gave an increase from 13.05°C to 14.86°C, CEEMD from 12.8°C to 14.87°C.

Data	**Method**	**1859–1937**	**1938–2015**
Max T	CEEMD	0.45	1.25
	SEEMD	1.04	1.26
Min T	CEEMD	0.40	1.30
	SEEMD	0.02	1.38

Table 5.2: Increase in Sydney temperature for the given periods.

The SEEMD results for Sydney are similar to those of Jones [1994] and Ghil and Vautard [1991] for global temperature, and of Vose et al. [2005] for Sydney temperature. Jones [1994] reviewed the surface air temperature record of the past 150 years. He found that, over the two 20-year periods of greatest warming this century, 1925–1944 and 1978–1997, global temperatures rose by 0.37°C and 0.32°C, respectively. In recent decades, there have been much greater increases in minimum temperatures than in maximum temperatures; over the period 1950–1993, the difference between maximum and minimum temperatures decreased by 0.08°C per decade. Jones [1994] found that for the period 1951–1990, minimum temperatures warmed at three times the rate of maximum temperatures and that minimum temperature in Sydney rose faster than maximum temperature over the entire last century. Annual global surface temperatures warmed by 0.57°C over the period 1861–1997 and by 0.62°C over the period 1901–1997. However, he could not obtain accurate results for the southern hemisphere due to lack of data coverage. Table 5.2 shows the rate of change in temperature estimated using SEEMD and CEEMD. The increases in the Sydney temperature from SEEMD and CEEMD are similar for the period 1938–2015 but quite different in the period 1859–1937. According to CEEMD, the minimum temperature in the periods 1859–1937 and 1937–2015 increased by about 0.9°C per year, whereas according to SEEMD, it was about 1.36°C per year.

In addition, Figure 5.8 also shows the average difference between the CEEMD and SEEMD results. The average of the trend using CEEMD is 21.4590 whereas when the SEEMD is used the average is 22.0774. This 0.6184 average difference is another evidence of the sensitivity of the EMD method to interpolation. This is also shown in the Bahri and Sharples [2015] results for sea level data. They showed that two different interpolation have different start and end points. In their results, there

was a 5 cm difference for the trend between the results from linear interpolation and cubic interpolation.

5.5 Temperature and its physical drivers

As mentioned in Section 5.1, changes in temperature have two types of drivers, external and internal. In this section, SEEMD and CEEMD are used to investigate whether the resulting IMFs can be related to specific drivers and whether the different interpolation methods can change the interpretation of the results.

5.5.1 CO_2 and temperature

The strongest external temperature driver in the last century has been the anthropogenic. Mann et al. [2000] investigated the effects of solar, greenhouse gas and volcanic forcing. They found that, while natural (solar and volcanic) forcing was an important factor governing the natural variation in temperature in past centuries, only human greenhouse-gas forcing, as noted by Mann and Park [1994], can statistically explain the unusual warmth of the past few decades.

In this study, CEEMD was used to analyse the global CO_2 concentration data for 1960–2015 from the NOAA website (*www.esrl.noaa.gov/gmd/ccgg/trends/data.html*). Due to our interest in long-term variability, only the CO_2 residual is used here. Figure 5.9 shows the global temperature residual using CEEMD (blue dashed line) and SEEMD (red solid line), together with the CO_2 residual calculated using CEEMD.

SEEMD and CEEMD gave similar residuals for the global temperature data and show a sharp increase after 1920. The CO_2 trend also shows a sharp increase from 1960 till the present. The time period for the CO_2 data is not very long, so that it is not clear if the CO_2 trend had similar behaviour with temperature before 1960.

5.5.2 Southern Hemisphere variability

One of the significant internal drivers of temperature in the Southern Hemisphere is ENSO. In the last century, there were stronger El Niño effects than previously experienced. These strong El Niños are called *super El Niño*. Hong [2016] showed these events happened in 1972–1973, 1982–1983 and 1997–1998. In this study we use these three *super* El Niños as examples of strong physical phenomena, and see if CEEMD and SEEMD can distinguish these events. The reason to take El Niños into account is that the presence of an El Niño event affects global mean temperature anomalies. For example, the big El Niño over the winter of 1997–1998 directly warmed a large part of the Pacific, indirectly warmed an even larger region and caused a warm year in 1998. The opposite effect was seen in the La Niña event in

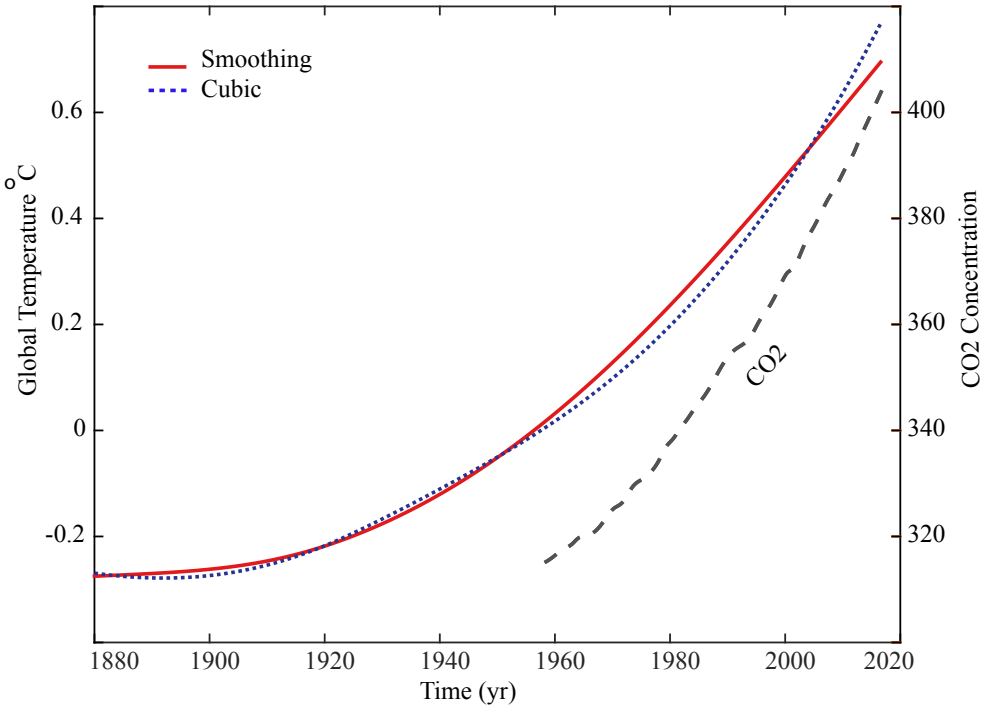

Figure 5.9: SEEMD and CEEMD trends for global temperature (left-hand scale). The black dashed line shows the CEEMD trend in the CO_2 concentration (right-hand scale)

the 2007 winter. Since the variability associated with these events is large compared to expected global warming trends over a short number of years, the underlying trends might be more clearly seen if the El Niño events were extracted from the climate model. There is no perfect way to do this.

Unfortunately ENSO influences the climate on different time scales. Trenberth and Fasullo [2013] emphasised the role of the long-term variation in ENSO, called the Pacific Decadal Oscillation (PDO), and found that PDO events may persist for up to two decades in the tropical Pacific, causing a somewhat slower warming at the surface, because more heat is stored deeper in the ocean. Rasmusson et al. [1990] noted that the multi-decadal oscillations in the temperature anomaly are probably related to the low-frequency component of ENSO.

The EEMD decadal IMFs (see Section 5.3) are shown in Figure 5.10. Smoothing-spline interpolation showed the decadal temperature fluctuations more clearly than cubic-spline interpolation. In Figure 5.10, the SEEMD result showed that the temperature reached maxima in 1952, 1970, 1983 and 1990, 1997 years reported as major El Niño years (Figure 5.10).

Hong [2016] showed that super El Niño happened in 1982–1983, and Mann and Park [1994] found 1983–1993, to be the warmest decade up to that time, as shown by the

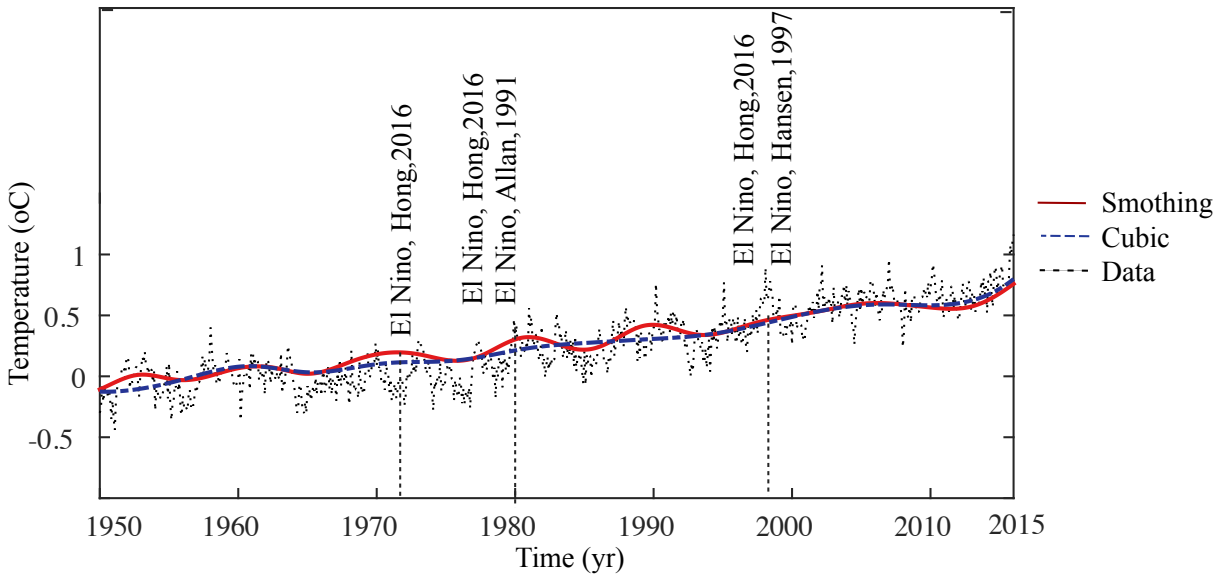

Figure 5.10: Decadal oscillation in global temperature and super El Niño events.

SEEMD results in Figure 5.10. The SEEMD results also show another significant peak in 1990, which is also mentioned as a super El Niño year by Hong [2016], Spencer [1990].

After 1990, neither interpolation method showed any significant peak in the global land and sea temperature; this time was called a cooling phase by Ghil and Vautard [1991]. About half of the slow down can be explained by a slower increase in radiative forcing. A problem is that the data on the net radiative forcing are too imprecise to better quantify its contribution. The often cited data on heat storage in the ocean are therefore just further evidence that El Niño played a crucial role in the pause. According to Herold and Santoso [2017], the ENSO effects in 86, 87, 91/92, 94, 02 and 09 were quite weak. In Figure 5.10, the SEEMD results show a brief cooling period after 1990 and no further peak in the temperature, just a gentle increase. Ghil and Vautard [1991] also found a large fluctuation just after 1950, which they related to the quasi-biennial component of ENSO; this is consistent with the SEEMD results in Figure 5.10.

The effect of the 1998 El Niño on temperature was not resolved in either EEMD method. The El Niño events in 1997–98 and 1982–83 temperature records appear to be somewhat anomalous in the context of the past several centuries. Hansen et al. [2006] called the El Niños in 1983 and 1998 super El Niños because of the sharp temperature gradient between the east and west Pacific.

Over the period 1998 to 2013, global temperature rose more slowly than previously but this does not mean that global warming has stopped. Due to natural variability,

trends based on short-term records are very sensitive to values at the beginning and end, and do not in general reflect long-term climate trends. The long-term warming and short-term pauses have nothing to do with each other, since they have very different causes.

5.6 Discussion and conclusions

The results in this chapter show that the EEMD method, regardless of the interpolation method used, can describe long-term behaviour in the temperature. However, changing the interpolation method changes the degree of variability in the trend, and even the interpretation of the data.

The SEEMD results had less mode mixing than CEEMD; less mode mixing can result in more statistically significant modes. For the decadal mode for the global temperature data, both CEEMD and SEEMD show the overall trend in the cooling and warming phases. However, SEEMD exhibited more detail in the decadal variability and contained features that matched known extreme events and El Niño occurrences reported in the literature.

EEMD showed a sensitivity to the interpolation method in finding the trend for the Sydney maximum and minimum temperature data, with greater sensitivity for the minimum temperature data. According to the SEEMD results, Sydney maximum temperature has increased about 2.3°C in the last 136 years, whereas CEEMD found an increase of about 2°C. For Sydney minimum temperature, SEEMD gave about a 1.4°C increase, CEEMD a 1.3°C increase. In addition, the slopes of the trend from SEEMD and CEEMD were not the same. The CEEMD result showed linear trends for both minimum and maximum temperature, with positive slopes, whereas SEEMD showed that the minimum temperature rise accelerated from 1920. The SEEMD result in this case agrees with the literature.

Although the differences between the smoothing-spline and cubic-spline interpolation results might seem small, they still cause serious issues in interpreting the results. For example, the maximum temperature difference between the SEEMD and CEEMD results for Sydney was about 0.3°C. Comparison of this value with the overall global temperature increase since 1880, reported in on Climate Change [2014] (0.85°C), shows the huge impact of the interpolation method in the EEMD method. Again its appears that smoothing-spline interpolation is the better interpolation method to use in EMD/EEMD.

Chapter 6

Application of Empirical Mode Decomposition to Sea-Level Data

6.1 Introduction

Global warming has resulted in significant environmental changes, among which its impact on the ocean, and more specifically on sea-level rise, is one of the most important leading to serious environmental hazards [Meehl et al., 2005]. According to the Intergovernmental Panel on Climate Change (IPCC) 6th Report [on Climate Change, 2014], global mean sea level has increased significantly in the 20th century, by about 1.7 mm per year between 1901 and 2010. In Mastrandrea et al. [2010], sea level was estimated to rise by up to 74 cm by the year 2100, corresponding to an average annual increase of 8.5 mm. These significantly different values could be the result of using different methods to estimate them. More accurate and more reliable prediction of sea-level change would help develop more efficient coastal management plans to mitigate possible damage from sea-level rise.

Church and White [2011] applied satellite altimeter data from the relatively short period 1993 to 2009 and tidal-gauge data from 1880 to 2009 to estimate the global mean sea level (GMSL) using the Empirical Orthogonal Function Expansion (EOF) method. For the period 1993 to 2009, both tidal and altimeter results from this method showed an increase in GMSL. The linear trend from the altimeter data from January 1993 to December 2009 was 3.2 ± 0.4 mm per year. However the value from the combined data was slightly smaller than the altimeter estimate after correction for glacial melting. For the 130 years from 1880 to 2009, the total rise was about 210 mm, commensurate with the 195 mm in their prior study which had eight years' less data [Church and White, 2011, Cai, 2006]. Church and White [2011] identified two different rates in the data: one from 1880 to 1935 was 1.1 ± 0.7 mm per year; the other from 1936 to 2009 was 1.8 ± 0.3 mm per year.

6.1.1 Historical sea-level variability

In the previous chapter, global warming in the last century was discussed. This warming has direct impacts on the environment, among which is impact on the ocean, and more specifically on sea-level rise [Meehl et al., 2005]. Meehl et al. [2005] showed that, even if the concentrations of greenhouse gases in the atmosphere had been stabilised in the year 2000, sea level would continue to rise due to future global warming of about another half degree, with an additional 32% sea-level rise caused by thermal expansion by the end of the 21st century.

A number of studies have been done on global sea-level rise, for example Church et al. [2006], Church and White [2011], Jevrejeva et al. [2008], Wenzel and Schröter [2014], Ray and Douglas [2011], Hay et al. [2014]; these studies, regardless of the method used, all demonstrated that the rate of rise in the last two decades (about 3 cm per decade) was the highest on record.

Global sea level has always been changing; in the last ice age, 20,000 years ago, sea level was about 120 m less than it is today [Solomon, 2007]. More recently, it was stable from 100 BC until 950 AD, then increased by 0.6 mm/yr for about 400 years, followed by a further stable period, then gradually decreased until late in the 19th century [Kemp et al., 2011, Vermeer and Rahmstorf, 2009]. Since then, sea level has risen at an average rate of about 2 mm/yr, representing the steepest century-scale increase of the past two millennia. This rate began to increase between 1865 and 1892. Using an extended semi-empirical modelling approach, Kemp et al. [2011] showed that these sea-level changes were consistent with the increase in global temperature for at least the past millennium.

From 1901 to 2010, total global sea level rose by 190 ± 20 mm, albeit at varying rates and spatial distributions [Church et al., 2013]. Kopp et al. [2016] showed that from late in the 20th century, sea level has risen faster than in any of the previous 27 centuries.

Lyu et al. [2014] defined the time of emergence (ToE) as the time when the climate-change signal from increasing greenhouse gases exceeded natural climate variability. They mentioned that the ToE is an important climate-change issue. In their study, they used three regional sea-level projection models to predict sea level in 2100, and estimated the ToE for sea-level changes relative to the reference period 1986 to 2005. They used a dynamic sea-level model, derived from ocean density and circulation changes, then added the global-ocean thermal expansion effect and additional contributions from land-ice mass loss, land-water storage change and glacial isostatic adjustment, which generally enhances the signal of regional sea-level rise (except in some regions with decreasing sea levels). Their results showed a ToE over more than

50% of the ocean area by 2020. This shows that ToE for sea level is substantially earlier than that for surface air temperature, and exhibits little dependence on the emission scenarios. It means that society will face detectable sea-level change and its potential impacts earlier than surface air warming.

Wenzel and Schröter [2010] used neural networks to estimate global sea-level variations based on long-term tidal-gauge records in the 20th century. The global mean sea level for the period January 1900 to December 2006 was estimated to have risen at a rate of 1.56 ± 0.25 mm/yr, reasonably consistent with earlier estimates, but they did not find significant acceleration.

6.1.2 What causes sea-level change?

There are several contributors acting on global and/or regional scales that affect sea level. The main global contributors are: thermal expansion, which is the increase in the volume of the ocean water mass due to atmospheric heat exchange; melting land ice from mountain glaciers, small ice caps and big ice sheets, which add water mass to the ocean; and changes in the depth of the ocean basin due to long-term movements of the Earth's crust [Rahmstorf, 2012b].

Church and White [2011] provided a sea-level budget analysis to estimate the contributions from each of these contributors. For the period 1972 to 2008, the sum of the contributions led to a total rise in sea level of about 70 mm. Over 50% of that was due to melting land ice, and a little less than 50% to thermal expansion. Church and White [2011] suggested that thermal expansion affects sea level over longer time scales, so that even if greenhouse-gas emissions were curbed, sea level would continue to rise. Therefore, in this chapter we pay special attention to the variability in sea level and its relationship to changes in global mean temperature.

Rahmstorf [2007] used a semi-empirical approach to connect global sea-level rise to global mean surface temperature. They proposed that the rate of sea-level rise $\frac{dH}{dt}$ on an anthropogenic time scale is roughly proportional to the magnitude of warming above the temperature in the pre-industrial age T_0:

$$\frac{dH}{dt} = a(T - T_0). \tag{6.1}$$

This holds to a good approximation for temperature and sea-level changes during the 20th century. Vermeer and Rahmstorf [2009] improved this method by adding a rapid-response term. They proposed a simple relationship linking global sea-level

variations on time scales of decades to centuries to global mean temperature:

$$\frac{dH}{dt} = a(T - T_0) + b\,\frac{dT}{dt}. \tag{6.2}$$

Vermeer and Rahmstorf [2009] applied this model to global sea-level data and temperature from 1880 to 2000, and found that it explained 98% of the variance.

Church and White [2011] quantitatively explained the rate of sea-level rise since 1970 by a gradual increase in ocean thermal expansion, with fluctuations at least partly related to volcanic eruptions and an increasing cryospheric contribution. The contributions from glaciers and ice caps [Cogley, 2009], and from the Greenland ice sheet [Rignot et al., 2011] both increased in the 1990s, and there was an increasing contribution from the West Antarctic ice sheet [Rignot et al., 2011]. The larger rate of rise in the 1940s may be related to larger glacier and ice-cap contributions [Oerlemans et al., 2007] and higher temperatures over Greenland.

Peltier et al. [2015] showed that land in most of the worldis subsiding at a fraction of a millimeter per year, compounding the problem of sea-level rise. However, at some locations in the northern latitudes, the reverse is happening; land is rising after being liberated from the mass of the ice sheets, again normally by less than 1 mm/yr, but in places of over 5 mm/yr.

6.1.3 From global to regional scale

Changes in sea level are spatially non-uniform [Church and White, 2011], so it is necessary to have a better understanding of the drivers of both global and regional sea-level changes to facilitate local adaptation planning [Rahmstorf, 2012a]. Regional sea levels are affected by global sea-level variability and by regional factors. With global contributors outlined above, this section discusses the impact of regional contributors. White et al. [2005] found that coastal sea level was rising faster than the global average in some parts of the world, for example around Australia, during the 1990s and around 1970, but that it rose slower than the global average during the late 1970s and late 1980s. They suggested that a more rapid coastal rise corresponds to La Niña-like conditions in the tropical Pacific Ocean, a slower rate to El Niño-like conditions.

Meyssignac and Cazenave [2012] studied the altimeter-based spatial trend patterns in sea level from 1993 to 2010. The variability in sea level was emphasised when they removed the global mean sea-level trend of 3.2 mm/yr. In some regions, such as the western and northern Pacific, the southern Indian Ocean and the south of Greenland, the local departures from the global mean trend were so large that they

actually dominated the sea-level rate signal over the relatively short period 1993 to 2010. In addition, all processes that influence global mean sea level vary in time and space. For example, local changes in ocean temperature and salinity lead to local sea-level changes through associated variations in the water-column density and volume (thermosteric and halosteric effects, respectively) [Wunsch et al., 2007, Lombard et al., 2009]. These variations are closely linked to atmosphere-ocean interactions (mostly through wind stress but also through exchanges of heat and fresh water) and associated changes in the ocean flow field [Köhl and Stammer, 2008, Timmermann et al., 2010].

Regional sea-level changes differ from the global sea-level change for a number of reasons. First, ocean waters circulates, driven by winds and other factors, so that even if the global water volume does not change, there will be regional sea-level changes. This can happen due to both natural oscillations in the climate system (such as El Niño/Southern Oscillation) and forced anthropogenic changes (e.g. a weakening of the Atlantic thermohaline circulation).

Second, the gravitational pull of land ice is reduced as the ice melts, which has a surprisingly large effect on the sea surface. For example, as ice on Greenland melts, this causes a global sea-level rise, but a regional drop in sea level in a circular area around Greenland extending all the way to Fennoscandia [Mitrovica et al., 2001].

Local sea-level changes can differ by some tens of centimeters or even more from the global mean sea-level change. This makes some locations, like low-lying delta cities on subsiding ground, particularly vulnerable.

On a regional scale, there are several factors than can affect sea level, for example ocean currents [Levermann et al., 2005], storm surges, modes of climate variability such as El Niño and long-term trends such as land movements [Church and White, 2011].

Church and White [2011] noted that sea level at any location contains the influence of local and regional meteorological effects (including storm surges), modes of climate variability (for example El Niño/Southern Oscillation) and long-term trends (from both the ocean surface and land movements), and including the impact of anthropogenic climate change. As the altimeter record has clearly demonstrated, global mean sea level has much less short-term variability (more than an order of magnitude) than sea level at individual locations because, as mentioned before, while the volume of the ocean is nearly constant, the distribution changes with time. While the variability between individual locations can be reduced by low-pass filtering of the data, there remains significant energy at yearly-to-decadal periods that may be either positively or negatively correlated between stations.

Levermann et al. [2005] investigated the relationship between thermohaline circulation, and the local and global variability of sea level. They found that local sea level can rise by up to 1 m due to local current circulation variability. Changes in sea level due to regional variations can be as large as the climatic signal on multi-decadal time scales [Rahmstorf, 2012a,b].

Factor	**Time Scale**	**Length (km)**	**Amplitude (m)**
Seiches and edge waves	1 (min)	10–80	0.03
Tropical surge	1 (hr)	20–100	5
Tsunami	10 (min)	100–600	1–10.7
Semi-diurnal tides	12 (hr)	100–10000	1–10
Diurnal tides	1 (dy)	110–11000	0.1–1
Extra tropical surges	12 (hr) – 7 (dy)	300–1100	1
Rossby waves	10 (dy) – 1.5 (yr)	100–1200	0.1
Climate change	5–20 (yr)	500–20000	0.2
Crustal movements	10 (yr) – 4.5 (by)	10–40000	1000

Table 6.1: Sea-level contributors of different times and lengths. Min: minute; hr: hour; dy: day; yr: year; by: billion years. Data from Pugh [2004].

Table 6.1 lists regional sea-level-change contributors [Pugh, 2004]. It should be mentioned that, although the time scale for climate-change contributors in Pugh [2004] is up to 20 years, there is evidence that climate-change contributions can have time scales of up to 60 years [Douglas et al., 2000].

Meyssignac and Cazenave [2012] summarised observations of sea-level variations, globally and regionally, during the 20th century. Over this period, the global mean sea level rose at rates of 1.7 mm/yr and 3.2 mm/yr , as a result of an increase in ocean thermal expansion and land-ice loss, respectively. Regional sea-level variations, however, have been dominated by thermal expansion over the last few decades, even though other factors such as ocean salinity and the subsidence response to the last deglaciation could have played a role [Nicholls and Cazenave, 2010, Antonov et al., 2002].

Lombard et al. [2009] showed that non-uniform ocean warming and therefore non-uniform thermal expansion, is most responsible for the observed spatial patterns in the sea-level trend. Wunsch et al. [2007] reproduced local sea-level data trend patterns observed using satellite altimetry from 1993 to 2004. They too showed that the change in the thermal expansion in the upper ocean is the dominant contribution to observed spatial trend patterns, but also that approximately 25% of the temperature contribution is locally compensated by salinity.

Several studies have indicated that the trends in sea-level rise around Australia are affected by climate change [Hughes, 2003, Mitchell et al., 2000]. Sea-level rises over the last several decades have varied from about $-0.95\,\mathrm{mm/yr}$ at Geraldton in the west to $+2.08\,\mathrm{mm/yr}$ at Port Adelaide in the south [Hughes, 2003, Mitchell et al., 2000]. The average of these changes is about $+0.3\,\mathrm{mm/yr}$, less than the IPCC global estimate [Hughes, 2003, Mitchell et al., 2000].

6.1.4 Limitations in sea-level studies

In principle there are two main challenges in studying sea-level rise. First is the data complexity and the second the choice of appropriate methods of analysis.

Tidal gauges are one of the most common, easy and widely accessible ways to obtain sea-level data. However, obtaining accurate sea-level data from tidal gauges is quite challenging, with problems such as: 1) tidal-gauge measurements are not available over a very long period and are not distributed evenly over the oceans; 2) many of the time series have data gaps; 3) tidal gauges measure sea-level relative to the land, so these data are corrupted by land movement.

Satellites are now a readily available source of sea-level data. Unlike tidal-gauge data, satellite data are continuous, are not corrupted by land movement and are of a high quality. However, the time span of satellite data is too short to provide enough information about sea-level variation over a multi-decadal time scale [Church and White, 2011]. Multi-decadal and decadal trend analyses are necessary to investigate sea level and its acceleration due to climate variability. Therefore, tidal-gauge data are the only option for long-term studies of sea level, and identifying a method that appropriately accounts for the errors in those data is necessary.

Watson [2011] mentioned that there are numerous meteorological and oceanographic processes that affect ocean water levels over differing time and spatial scales. Several influences are on time scales less than a year, such as seasonal factors, weather systems and variability in ocean water properties. Other longer-term influences such as ENSO (up to 10 years) and Pacific Decadal Oscillation (PDO) (20–30 years) will influence ocean water levels over much longer time frames. Choosing the 'best' method to isolate these dynamical signals from sea level data can be quite challenging.

A suitable method is that able to cope with the challenges in the data. The two studies by Church and White [2011] and Pokhrel et al. [2012] are good examples to show the importance of the chosen method in the results. Church and White [2011] showed that land-water storage makes a small negative contribution to sea level variability, whereas Pokhrel et al. [2012] claimed that 42% of sea-level rise over the period 1961 to 2003 was due to reduced land-water storage. In comparison to the

Church and White [2011] study, reservoir storage was assumed to be smaller, but the flow of groundwater to the sea was calculated to be several times larger. The problem with the study of Pokhrel et al. [2012] is that it did not chose realistic assumptions for their model. One of the problems in their study is they calculated the rate of the sea level by just using the values at the beginning and the end of the dataset whereas Church and White [2011] calculated the rate using linear interpolation through all the data points. Therefore, Church and White [2011] seems to be closer to reality.

Estimating sea-level acceleration is extremely important for coastal adaptation planning, and unfortunately estimates are extremely sensitive to the method used to calculate them. Cheng et al. [2016] found the sea-level acceleration in certain areas using linear-least-squares fitting and sea-level rate-difference approaches. They found significant sea-level acceleration in the Bohai Sea, Yellow Sea and Yangtze River Estuary, with relatively lower sea-level acceleration in the open ocean adjacent to the Bohai and East China Seas.

The results presented in this dissertation demonstrate an alternative way to calculate sea-level acceleration (SLA). However the accuracy of using the EMD method to calculate the SLA is not yet known.

The EEMD method has been used recently to estimate sea-level rise (SLR). Veltcheva and Soares [2004] used sea-level data collected at Figueira da Foz, Portugal from March to June 1994. Their results showed that the EMD method could reliably extract all oceanographic oscillations. Duffy [2005] applied HHT to sea level observed at Chesapeake Bay, to solar radiation at Des Moines, Iowa, and to barotropic observations in central Illinois. They found that EMD/HHT could distinguish periodic features such as the tides, and episodic events such as snow melt and heavy precipitation events. EMD/HHT could also capture a wide variety of phenomena such as diurnal cycles, frontal passages, baroclinic instabilities and seasonal cycles. They suggested this technique be used for investigating important weather events, from floods to ENSO events. They also found the EMD/HHT method quite useful in extracting the location and nature of significant events from their data.

Ezer et al. [2013] used EMD/HHT to study sea-level rise on the mid-Atlantic coast of southern USA and its correlation with a regional contributor (the Gulf Stream). They found a strong correlation between sea-level rise and Gulf Stream variations with a high level of confidence (99.99%). Saramul and Ezer [2014] obtained rates of relative-sea-level rise (RSLR) using two different methods, linear regression and non-linear EMD/HHT. Their results showed extremely large spatial variations in RSLR, varying from about 1 mm/yr to about 20 mm/yr. Furthermore, there were

indications that RSLR increased significantly in all locations after the 2004 Sumatra-Andaman earthquake and the Indian Ocean tsunami that followed, so that recent RSLR seemed to have fewer spatial differences than in the past, but high rates of about 20 to 30 mm/yr almost everywhere. A conclusion from these results is that RSLR in this region is dominated by regional land movements over the smaller global sea-level rise. Faster rates, about 10-20 mm/yr, or 5–10 times greater than the mean rates from global tidal gauges [Church and White, 2011, Church et al., 2006], were found where groundwater extraction caused significant land subsidence.

Kenigson and Han [2014] used EMD and EEMD to isolate oscillations and provide robust acceleration estimates for sea-level trends. They mentioned that the reliability of these methods in calculating accelerated sea-level rise had not yet been fully tested. They applied the EMD and EEMD methods to both tidal-gauge observations and synthetic sea-level time series, constructed as a sum of oscillations extracted from tidal-gauge records and trends with prescribed acceleration rates.

Bahri and Sharples [2015] applied linear and cubic EEMD methods to Brisbane sea-level data to determine the long-term trend. Their results indicated significant differences in the long-term trend estimated using the two interpolation methods. In particular, the linear EEMD method exhibited a non-monotonic piecewise-linear trend. Differences in the two long-term trend estimates were most significant at the ends of the domain, where they differed by approximately 50 mm. While this may not seem large in absolute terms, it is important to note that when dealing with the ramifications of global and regional sea-level change, trends of the order of millimetres per decade can be important, and so the differences encountered here could be quite significant in applications. Moreover, without knowing what the actual trend in the sea-level dataset is, it is difficult to assess which trend estimate is the more accurate. However, both the cubic and linear EEMD trends reproduced the general behaviour of the time series reasonably well. This shows the importance of understanding the sensitivity of the method used in estimating sea-level long-term behaviour. The rest of this chapter is dedicated to extending our understanding of the sensitivity of EMD on the interpolation method and its application to sea-level data.

In this chapter, unlike Chapter 4, we do not know the actual modes. Therefore we can not directly compare the EMD results with the actual modes. In order to make a judgment on which interpolation method gives the better results, we compare our results with other studies and obvious physical phenomena and see if using different interpolation method changes the detailed information that we can extract using EMD.

6.2 EEMD analysis of sea-level data

In this section, the non-stationary characteristics of sea-level data for eastern and southern Australia will be discussed. Most of the tidal gauges around Australia are located along the coast. In addition to the tidal-gauge data challenges discussed above, there are particular challenges for tidal gauges along the coast. Sea level on the coast can be influenced by local effects such as the wind piling up water against the shore. Although variability in the wind cannot affect the global sea level, it can affect local sea-level estimates. This wind variability can impact sea-level-change records for up to a decade [England et al., 2014]. In this chapter, the EMD method is used to remove these false decadal impressions and to give a better understanding of long-term sea-level variability.

Figure 6.1 shows sea-level time series at seven locations in eastern and southern Australia. The regression lines show that the mean sea level rose at most of the locations, and that the rates of rise were non-uniform. A non-uniform rate is evidence of non-stationarity in the sea-level data, both temporal and spatial. In this section we discuss sea-level variability at different time scales, its acceleration and the type of non-stationarity in more detail.

A number of different methods have been used to estimate sea-level acceleration. One method is the quadratic approach; Church and White [2011] estimated acceleration in global mean sea level by fitting a quadratic function to the time series, taking account of the time-variable uncertainty estimate. They also removed all direct (non-climate) anthropogenic changes in terrestrial water storage from their observations to focus on climate-related variability. They found a considerable variability in the rate of rise during the twentieth century but there was a statistically significant acceleration since 1880 of $0.009 \pm 0.004\,\mathrm{mm\,yr^{-2}}$. Cheng et al. [2016] estimated SLA using sea-level rate differences used linear regression. They split the sea-level time series into two halves, and calculated the difference between the sea-level rates of the two half-series. For a series over the period 1950 to 2009, the two half series from 1950 to 1979 and from 1980 to 2009 were used to calculate the linear trends.

Based on these three commonly used approaches for estimating sea-level acceleration, consistent sea-level accelerations of $0.085 \pm 0.020\,\mathrm{mm\,yr^{-2}}$ (1950–2013) and $0.074 \pm 0.032\,\mathrm{mm\,yr^{-2}}$ (1959–2013) were observed in the Bohai Sea and the East China Sea, respectively. The acceleration in the Yellow Sea was not statistically different from zero, but the record length (20 years) was relatively short.

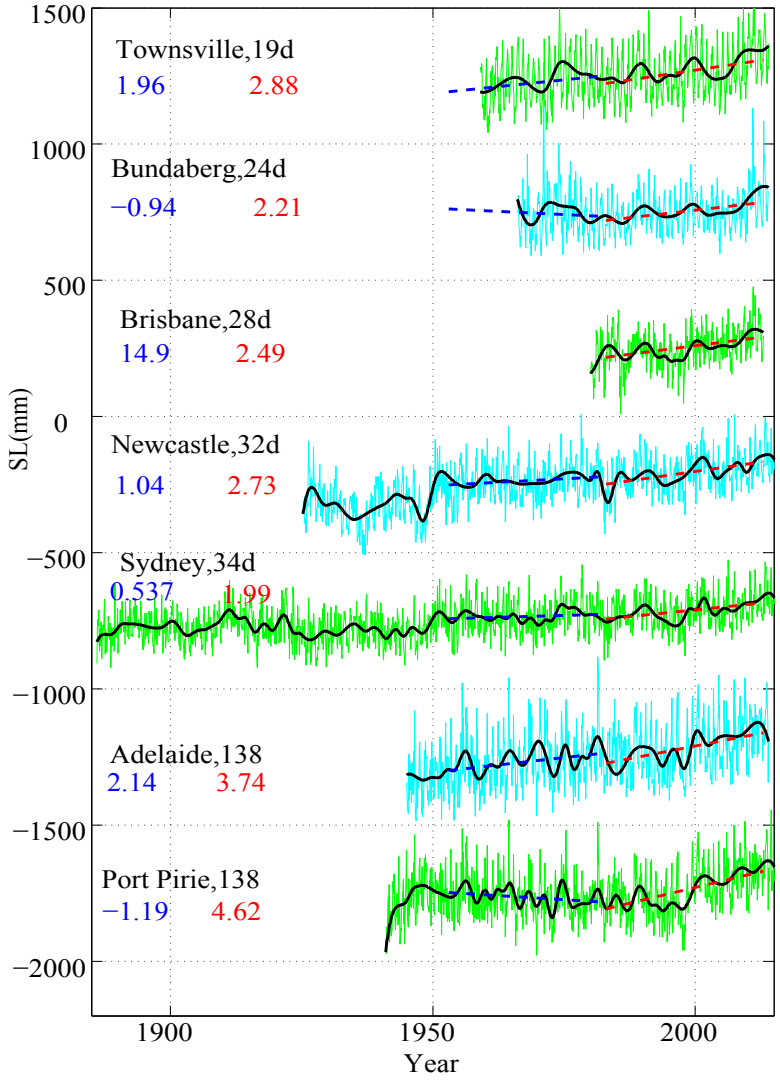

Figure 6.1: Sea-level variability along the east and south coast of Australia. Solid green and light blue lines show the observed sea-level data; solid black curves the sub-decadal variation in sea level; dashed blue line the linear regressions over the 30-year time period from 1953 to 1983; dashed red line a linear regression for the period 1983 to 2013.

For the locations in Figure 6.1, the minimum acceleration, $0.015\,\mathrm{mm\,yr^{-2}}$, occurred at Townsville. The Bundaberg and Brisbane sea-level accelerations were similar to those at the other stations in the same region for the period 1983 to 2013, but these records are not long enough to show if the rates have changed in the last 60 years. The sea-level accelerations for Newcastle and Adelaide were similar, $0.028\,\mathrm{mm\,yr^{-2}}$ and $0.026\,\mathrm{mm\,yr^{-2}}$, respectively. Port Pirie had the maximum rate of sea-level acceleration of $0.097\,\mathrm{mm\,yr^{-2}}$. According to Harvey et al. [1999], this high rate might be due to land-level changes. Nevertheless, the data for Port Pirie have significant changes in the mean over time and consequently a strong non-stationarity.

Douglas et al. [2000] advised that record lengths of at least 60 years are required to distinguish between trends and long-period relative sea-level fluctuations in individual records. Therefore, in this study, the three longest tidal-gauge datasets (Newcastle, Sydney and Port Pirie) from the east and south of Australia were chosen.

The Newcastle sea-level dataset is one of the longest sea-level records in Australia, from 1925 to 2015, and was obtained from the Permanent Service for Mean Sea Level website (*www.psmsl.org/*). The Newcastle sea level data came in several files; in this study, we used two files, one called Newcastle III, latitude $-32.91°$ and longitude $151.80°$ from 1925 to 1988. The other file called Newcastle V, latitude $-32.92°$ and longitude $151.78°$ from 1957 to 2015. These two gauges were run in parallel for a number of years until station Newcastle V was synched with the monthly records of station Newcastle III; Newcastle V then became the primary gauge. After obtaining the data, the two files were combined and the missing data eliminated or interpolated. In cases where more than 10 values were missing in the record, the values from the missing point onward were deleted from the analysis. In cases where just a couple of points in the middle of the dataset were missing, the missing values were replaced by the mean value of the surrounding two points.

The second dataset is Sydney monthly sea-level data from 1886 to 2015, obtained from the Permanent Service for Mean Sea Level website (*www.psmsl.org*). The Sydney data came in two files, one called Sydney, Fort Denison latitude $-33.85°$ and longitude $151.23°$ from 1886 to 1993. For 1914 and earlier, the sea-level values are taken from a re-computation of the data by Hamon [1987]. Later data from Fort Denison can be found in other PSMSL records and are mean sea level (MSL), based on data from a new tidal gauge. The other file is called Sydney, Fort Denison2 latitude $-33.85°$ and longitude $151.22°$ from 1914 to 2015. The two datasets were combined and the missing data eliminated or interpolated.

The third dataset is from the Port Pirie tidal-gauge station in South Australia for 1917 to 2015. This dataset was also obtained from The Permanent Service for Mean Sea Level website (*www.psmsl.org*) for $-33.177°$ latitude and $138.01°$ longitude. Although the data were available from 1917, the source of data from this year until 1941 was not reliable due to lack of detailed information from the station. Also, the measurement method appears to be different from 1941 onward. Therefore, we used the data from 1941 to 2015. Harvey et al. [1999] found the tidal-gauge data from this station to have been strongly impacted by land-level changes. They estimated a sea-level rise of 0.31 mm per year. They warned that if we did not consider land movement in this area, we would overestimate the rate of sea-level rise.

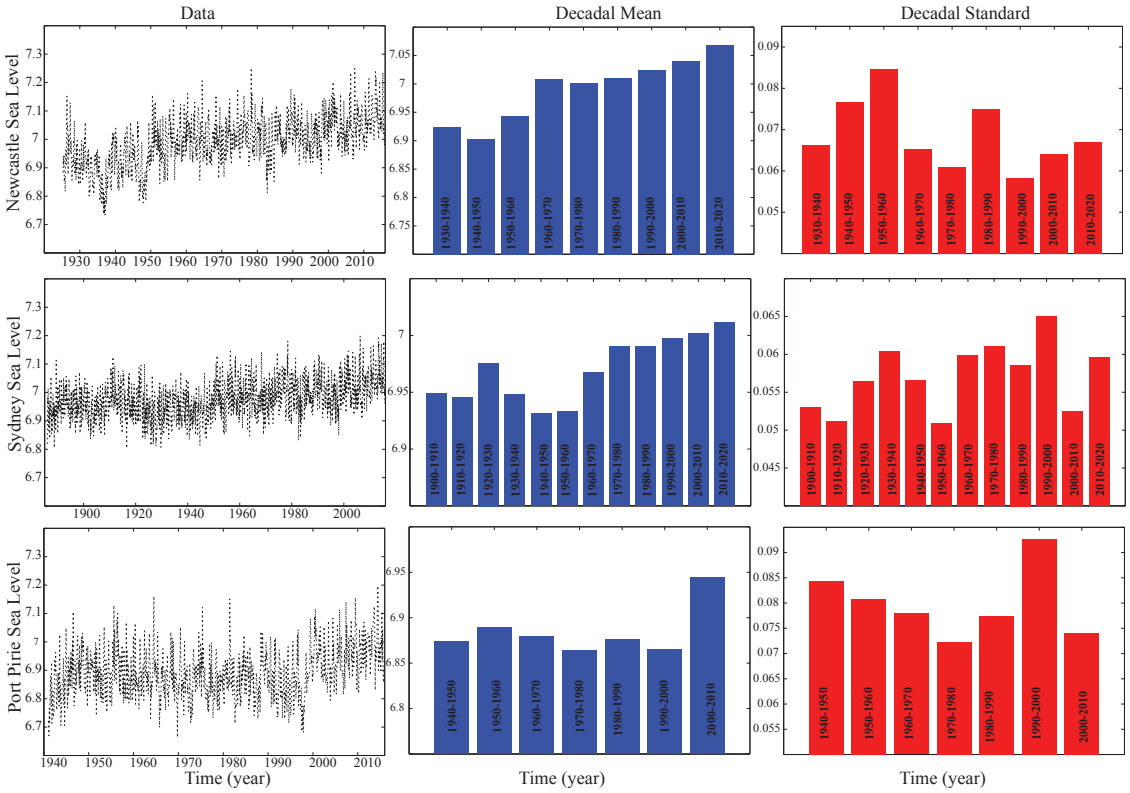

Figure 6.2: Monthly mean sea level and the time variability of their decadal mean and standard deviations. The left column of panels shows the data, the middle column the decadal mean variable and the right column the decadal standard deviations. The top row of panels shows the land-sea mean global monthly sea-level data at Newcastle, the second row, Sydney and the third row, Port Pirie.

The decadal variability of the mean and standard deviation of the three sea-level datasets is presented Figure 6.2 as a evidence of non-stationarity of the data. In order to do that, we calculated the mean and variance of the data for each decade. The results are shown as bar charts in Figure 6.2. At all stations, the decadal mean increased over time. The decadal mean of the Newcastle data increased to a local peak in the period 1960 to 1970. The decadal standard deviation for this station fluctuated over time. In the Sydney data, the decadal mean sea level rose slightly, with a small peak around 1920 and a minimum in 1940 to 1950. Church et al. [2006] found maxima in the rates of sea-level rise of over 2 mm per year in the 1940s and 1970s, and nearly 3 mm per year in the 1990s in Australia. As in earlier studies, using 10- and 20-year windows [Church et al., 2006, 2008], the most recent rate of rise over these short 16-year windows was at the upper end of the histogram of trends but was not statistically higher than the peaks during the 1940s and 1970s.

In the Port Pirie data, after 2000, the decadal mean rose sharply and the decadal variance decreased dramatically. Although Harvey et al. [1999] mentioned changes in this area might be due to land movement, their study was before 2000. Therefore,

it is not obvious whether the changes after this year are still just a result of land movement. Nevertheless, this sharp change in the mean and standard deviation is expected to impact the EMD results.

6.3 EEMD analysis of sea-level data

In this section, the Newcastle, Sydney and Port Pirie monthly tidal-gauge sea-level data are analysed using cubic-EEMD (CEEMD) and smoothing-EEMD (SEEMD). In particular, we obtain the statistically significant IMFs.

As an example, an output of the CEEMD and SEEMD with 500 ensemble members is shown in Figure 6.3 for the Sydney sea-level data. Both interpolation methods extracted similar IMFs but CEEMD extracted more than SEEMD at the higher frequencies (Figure 6.3). The reason is that, when the smoothing parameter is set low, $p \simeq 0$ (as in Figure 6.4), the smoothing splines do not resolve the high-frequency variations in the data. One of the challenges in the original EEMD results with cubic splines is that often the information on one physical driver is spread over several modes, and this makes choosing the physically meaningful modes harder.

Figure 6.4 shows the statistically significant IMFs (using the Monte Carlo method explained in Chapter 5) for sea-level data using SEEMD (left column) and CEEMD (right column). In Figure 6.4 for the Newcastle sea-level data, SEEMD extracted three IMFs whereas CEEMD found eight IMFs. IMFs 7 and 8 in the cubic EMD results both have a similar period but different energies, a result of mode mixing.

For the Sydney sea-level data, CEEMD extracted seven statistically significant modes whereas SEEMD only found five. The first, third and fourth CEEMD IMFs are very close to the error bound, and it is likely that they carry little energy. These IMFs make interpreting the CEEMD results more complicated because it leaves the final results to user judgment.

For the Port Pirie data, SEEMD again found fewer IMFs than cubic splines (three vs seven; Figure 6.4). CEEMD did not extract a mode with a decadal period, but SEEMD did. IMFs 7 and 8 in the CEEMD results and IMFs 5 and 6 in the SEEMD results again have the same period with different energies. This is due to mode mixing, in which a single IMF is divided into two.

In this study, IMFs within the confidence intervals and those close to the confidence intervals were not considered; only the IMFs that possessed considerably more energy than might be expected from noise were considered.

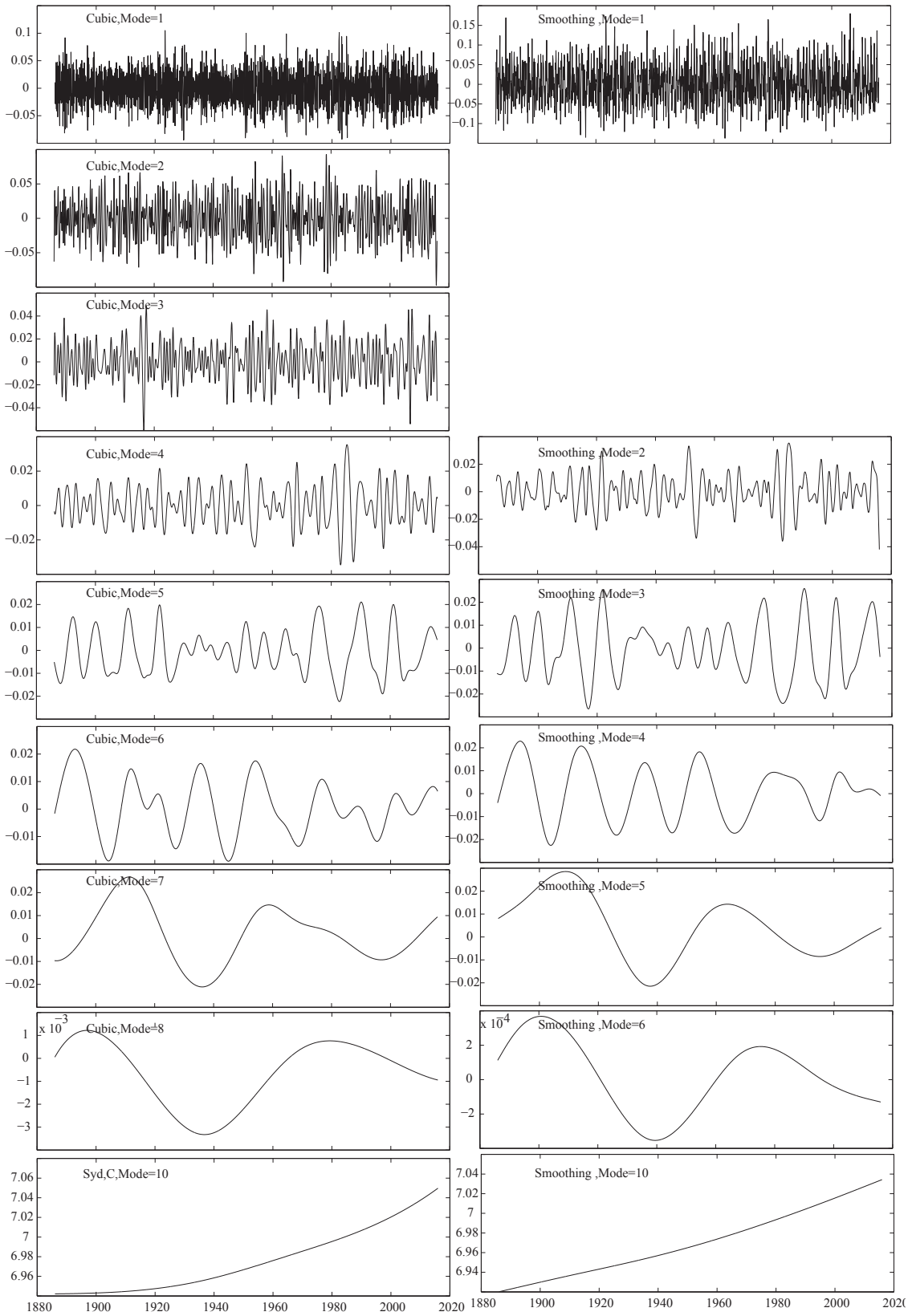

Figure 6.3: IMFs (modes) from EEMD for the Sydney sea-level data. Left-hand panels cubic interpolation; right-hand panels smoothing spline interpolation with smoothing parameter $p = 0.0015$. the last two modes are the residuals (trends).

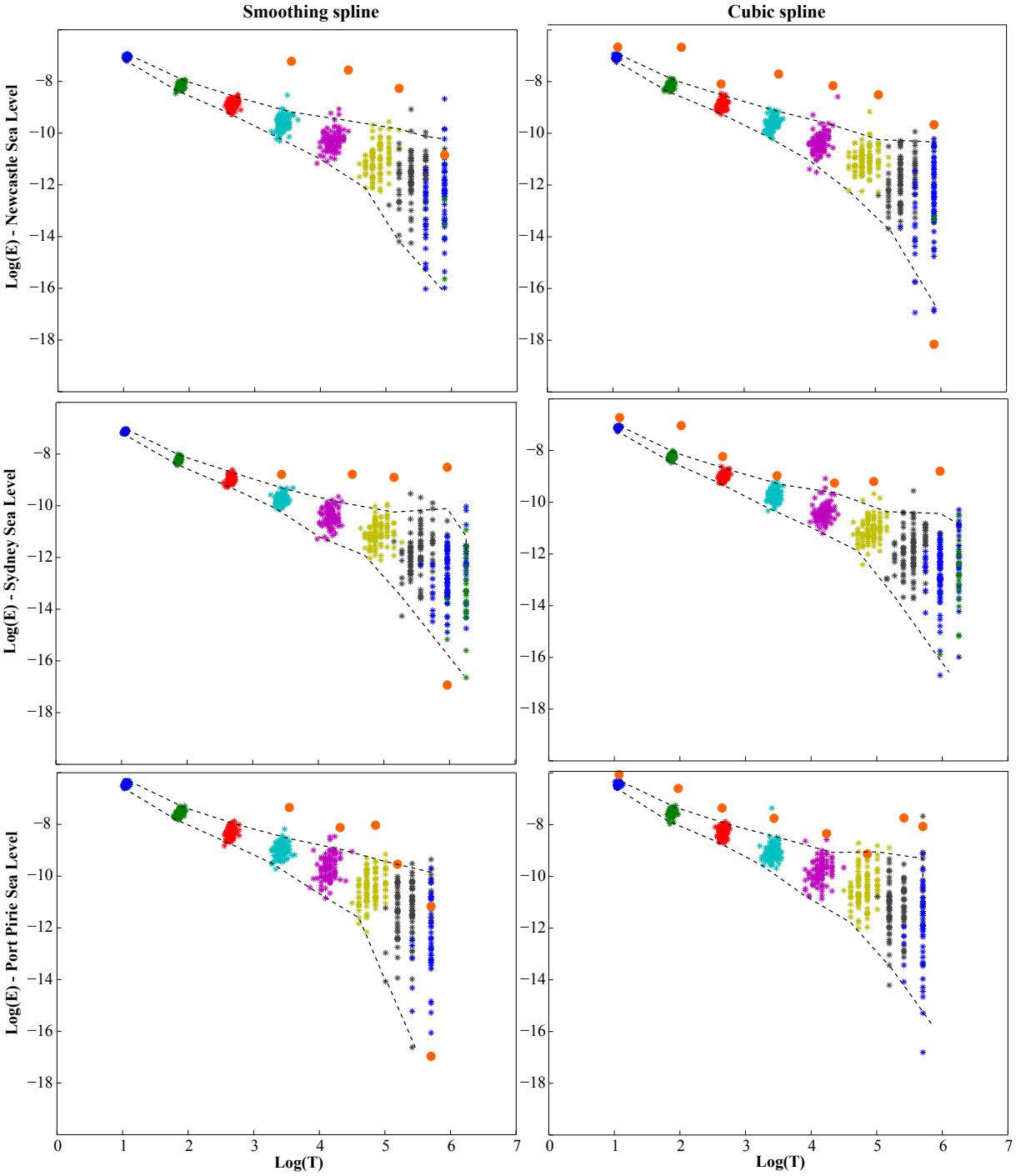

Figure 6.4: Statistically significant IMFs for the three sea-level datasets. The dotted lines show the upper and lower confident intervals determined from the energy densities of the reference section IMFs (coloured dots). The orange circles show the energy densities of the IMFs. Those that lie above upper confidence limit correspond to statistically significant modes.

Mode mixing impacts the results of the statistically significant tests and hence the number of statistically significant IMFs obtained using each method. In the presence of mode mixing, when the information in one IMF is distributed to more than one IMF, the mixed modes may still have enough energy to remain outside the confidence intervals or they may not.

To study the long-term variability in sea level, it is common to eliminate the short-term variability. Some studies eliminate all inter-annual and even decadal oscillations [Burgette et al., 2013]. In this study, in order to extract the IMFs related to climate change that are statistically significant, we have summarised the statistically significant modes according to their time periods in Table 6.2.

	Period				
Data	**Interpolation**	**Intra-Dec**	**Decadal**	**Multi-Dec 1**	**Multi-Dec 2**
Newcastle	Cubic Smoothing	Mode 4 Mode 2	Mode 5 Mode 3	Mode 6 –	– –
Sydney	Cubic Smoothing	– Mode 2	Mode 5 Mode 3	Mode 6 Mode 4	Mode 7 Mode 5
Port Pirie	Cubic Smoothing	Mode 4 Mode 2	Mode 5 Mode 3	– –	– Mode 5

Table 6.2: Statistically significant IMFs at the different time scales: intra-decadal (period ∼3 years); decadal (period ∼10 years); multi-decadal 1 (period ∼20 years); multi-decadal 2 (period ∼ 30 years). In this table, the blue shows the CEEMD modes, the red the SEEMD modes.

The statistically significant mode at each of the time scales is shown in Figures 6.5 to 6.9.

6.4 Interpolation sensitivity – application to sea-level data

Intra-decadal variability:

Figure 6.5 shows the statistically significant intra-decadal modes (period ∼3 years) at the three stations. The Newcastle and Port Pirie results using CEEMD and SEEMD are similar. However for Sydney, there are some minor differences at the endpoints between the smoothing-spline and cubic-spline results.

Decadal variability:

In the decadal results in Figure 6.6, the Newcastle cubic and smoothing results again show good agreement. However, they are both influenced by mode mixing. This mode mixing occurs in the period 1960 to 1970, when the peak in the mean of the input data occurs (Figure 6.2). It is shown in Chapter 4 that the results in the EMD method, regardless of interpolation used, are strongly sensitive to abrupt changes.

Similarly for the Sydney decadal IMF, some minor disagreement between the two interpolation methods happens in the period 1920–1940, when the fluctuations in the mean of the input data occur. The mode mixing in period 1960 to 1970 Newcastle and 1920 to 1940 at Sydney is further evidence of this sensitivity. For the Port Pirie decadal IMFs in the period 1960 to 1970, SEEMD and CEEMD estimated the amplitude quite differently. After 2000, the CEEMD and SEEMD IMFs show opposite variability in sea level.

Multi-decadal variability

Only Sydney had statistically significant SEEMD and CEEMD IMFs with period around 20 years (multi-decadal). There were no statistically significant IMFs resulting from CEEMD for the Newcastle data, and in Port Pirie none of the multi-decadal IMFs were statistically significant. For this reason, only the Sydney multi-decadal IMFs are shown in Figure 6.7. In Figure 6.7, SEEMD and CEEMD show a good agreement, apart from when the CEEMD results show some mode mixing in years 1910–1920 and 1980; the SEEMD results have less mode mixing than the CEEMD results.

For multi-decadal IMFs with longer time periods ($\sim$30 years; multi-decadal 2), the Sydney data again produced the only statistically significant IMF. The SEEMD and CEEMD multi-decadal IMFs are shown in Figure 6.8. Overall, both methods showed the long multi-decadal IMFs, with a small difference in the endpoints (especially between 1885 and 1910), and a small amount of mode mixing in the CEEMD result around 1970.

Rate of sea-level rise

Figure 6.9 shows the trend from SEEMD and CEEMD using the three datasets. Both methods found similar trends for the Sydney and Newcastle data, with some differences between the interpolation results at the endpoints, more obvious in the Sydney result. From the Sydney data, SEEMD and CEEMD both estimated the average rate of sea-level rise over the whole period 1880–2015 as about 0.8 mm per year. In Newcastle, both methods indicate a fast rise in sea level from 1925 to 1960

and less rise in sea level after 1960. Both methods estimated the rate of the sea-level rise about 2 mm per year.

Apart from the total sea-level rise, it is important to estimate the shape of the trend accurately. The shape of the trend gives us information about whether sea-level rise is accelerating or decelerating. As mentioned in Chapter 4, the EEMD method sometimes fails to find the proper trend shape; the trend results for Port Pirie show a clear example of this. As seen in Figure 6.9 for Port Pirie, CEEMD shows a linear trend, whereas SEEMD shows a long period from 1940 to 1990 with little change in sea level followed by a significant rise.

For Port Pirie, SEEMD estimated the rate of sea-level rise at about 1.9 mm per year but the CEEMD estimate was 1 mm per year. The SEEMD results agree with those of Hughes [2003] and Mitchell et al. [2000]. CEEMD and SEEMD estimated the overall rate of sea-level rise up to 1990 as 0.04 and 0.2 mm per year, respectively. However, after 1990, SEEMD found a sea-level increase of 0.19 mm per year whereas CEEMD reported a sea-level rise of 0.23 mm per year. Using two different interpolation methods in EEMD definitely changed the rate and the shape of the resultant trend. This clearly shows the sensitivity of the EEMD method to interpolation.

Sea-level acceleration

As mentioned above, sea-level acceleration (SLA) is an important parameter in planning for climate change adaptation and mitigation, and is has been considered in a number of studies.

Watson [2011] investigated sea-level acceleration in Australia. They used a 20-year moving-average water-level time series, and fitted second-order polynomial functions to consider trends of acceleration in mean sea level over time using long sets of data from Fremantle (from 1897), Auckland (from 1903), Sydney (from 1914) and Newcastle (from 1925). They estimated the average rate of acceleration in the mean sea level over the period 1940 to 2000 at the Sydney and Newcastle stations as $-0.04\,\mathrm{mm\,yr^{-2}}$ and $-0.096\,\mathrm{mm\,yr^{-2}}$, respectively.

In this study, SLA is estimated by finding the derivative of the residual resulting from the SEEMD and CEEMD methods. The results are shown in Figure 6.10. For Newcastle, both methods showed a time-dependent SLA, with a decrease in the acceleration from 1935 to 1960, then a pause from 1960 to 1970, followed by an increase in the acceleration until 1990 from SEEMD and until 2000 from CEEMD. Although both methods showed similar behaviour in the SLA, the amount of decrease up to 1960 was different; SEEMD found that the SLA dropped to $-0.13\,\mathrm{mm\ yr^{-2}}$ whereas CEEMD suggested that it only fell to $-0.06\,\mathrm{mm\ yr^{-2}}$. There was also some disagreement in the SEEMD and CEEMD results at the endpoints. For Newcastle, CEEMD estimated the average sea level acceleration as $-0.0052\,\mathrm{mm\,yr^{-2}}$ which is smaller compared with the one from by SEEMD, $-0.017\,\mathrm{mm\,yr^{-2}}$. Watson [2011] also estimated negative acceleration for sea level but the estimated amount is quite different with what is obtained by EMD. The SLA estimated by the linear-regression method in Section 6.2 is completely different from the result from EMD and Watson [2011].

For Sydney, CEEMD found that the SLA rose up to $0.021\,\mathrm{mm\ yr^{-2}}$ in 1930, then dropped to less than zero in 1970. The SLA then rose sharply up to $0.05\,\mathrm{mm\ yr^{-2}}$ in 2016. The SEEMD results however, showed a gentle drop to $-0.005\,\mathrm{mm\ yr^{-2}}$ in 1910, then a rise up to $0.008\,\mathrm{mm\ yr^{-2}}$ in 1930, after which it remained relatively constant till 2000. After 2000, the SLA decreased. Clearly the CEEMD and SEEMD sea-level accelerations are quite different for Sydney. CEEMD suggest that SLA increased but SEEMD results showed that SLA stayed the same for the last couple of decades. The average sea-level acceleration in Sydney is estimated at $0.0038\,\mathrm{mm\,yr^{-2}}$ by SEEMD and as $0.0167\,\mathrm{mm\,yr^{-2}}$ by CEEMD. The sea-level acceleration results from both methods disagree with the results of Watson [2011] and those in Section 6.2 from linear regression.

For Port Pirie, the SEEMD SLAs were completely different from the CEEMD SLAs. CEEMD found a gradual linear increase from about 0.001 to $0.01\,\mathrm{mm\ yr^{-2}}$ in the SLA whereas SEEMD found strong fluctuations from 1940 until 2016, with the SLA varying from $-0.025\,\mathrm{mm\ yr^{-2}}$ in 1955 to about $0.1\,\mathrm{mm\ yr^{-2}}$ in 1990. For Port Pirie, CEEMD and SEEMD estimated the average sea-level acceleration as $0.005\,\mathrm{mm\,yr^{-2}}$ and $0.037\,\mathrm{mm\,yr^{-2}}$, respectively. Overall the SLA results show a strong sensitivity to the interpolation method, with the differences significant in the context of sea-level rise.

6.5 Sea level and its drivers

6.5.1 Wind stress

Global warming can change the regional wind stress, which can impact regional and local sea level and circulation patterns. Timmermann et al. [2010] summarised the relationship between global warming and wind stress, and its impact on sea-level variability; their results are shown schematically in Figure 6.11. They found that the sea surface temperature warming pattern in the tropical and subtropical Pacific generates a robust wind response in the tropical South Pacific. The related change in wind-stress curl induces local Ekman suction and hence a decrease in sea level relative to the global mean sea level in many areas in the South Pacific and southern Indian Oceans.

Holbrook et al. [2011] used satellite altimeter data to estimate large-scale sea-level changes in Australia. They showed that inter-annual and decadal sea-level changes can be affected by local along-shore winds, coastal trapped waves and the offshore sea level.

Merrifield [2011] showed that over the past two decades, sea level has increased in the western tropical Pacific Ocean at rates that are approximately three times the global average. A general circulation model was used to show that these high rates are caused by a gradual intensification of the Pacific trade winds since the early 1990s. These results show that sea-level changes in the western tropical Pacific are indicative of a multi-decadal increase in the Pacific trade winds.

Dangendorf et al. [2013] used sea-level data at Cuxhaven (in Germany) from 1871 to 2008, and compared the effects of wind stress, sea-level pressure and precipitation on sea-level variability. They found that there was a robust estimation of the regression coefficient if more than 60 years of data were used. They looked at the effect of these drivers on sea level at seasonal and long-term time scales. At seasonal time scales, most of the variability in sea level was caused by zonal wind-stress changes; sea-level pressure and meridional wind stress had smaller but still significant effects. At decadal time scales, local meteorological forcing was significantly weaker, and remote forcing over the North Atlantic played the main role.

6.5.2 East Australian Current

Levermann et al. [2005] showed that regional ocean circulation can affect sea level at a regional scale. In eastern Australia, the East Australian Current (EAC) is the major southward current, and plays a significant role in the climate of this area. The EAC is the link between the Pacific and Indian Ocean gyres, and is dominated

by a series of meso-scale eddies of variable strength and direction. Changes in the wind-stress curl can change the strength of the gyre and therefore the EAC; changes in the EAC can then change the regional sea level. EAC is a western boundary current with a volume transport of 20–30 Sv [Ridgway and Dunn, 2003]. Suthers et al. [2011] showed that the poleward-flowing EAC is characterised by its separation from the coast, to form the eastward flowing Tasman Front and a southward flowing eddy field. The separation zone greatly influences coastal ecosystems on the relatively narrow continental shelf (only 15–50 km wide), particularly between 32–34°S. Godfrey et al. [1980] found that eddies are dominant in the region of the separation point. The separation point is determined by wind stress, coastal geometry (coastal boundary) and bathymetry. Due to climate change, wind stress has changed in this area over recent decades [Cai, 2006].

Ridgway [2007] found that penetration of the EAC has increased southward in the last 60 years. As a consequence, this region has become warmer and saltier, with mean trends of 2.28°C/century and 0.34 psu/century over the years 1994 to 2002. Cai [2006] used the Godfrey's Island Rule model to show that changes in Antarctic ozone will change the surface wind stress and thereby affect the large-scale circulation in the southern hemisphere. Variations in large-scale circulation will strengthen the EAC flow by up to 20%, and perhaps even more in the south. Over the last decade, the EAC flow has decreased at $20-30$°S, whereas at 30°S it has increased by about 9 Sv or 20%. According to Cai [2006], the EAC is becoming warmer, saltier and stronger in the south; this could affect local marine ecosystems. The CSIRO Mk3.5 coupled climate model and the ocean downscaling simulations from OFAM also show similar results [Sun et al., 2012].

Suthers et al. [2011] also examined the effects of climatic wind-stress-forced ocean dynamics on EAC transport variability and coastal sea level, from ENSO to multi-decadal time scales. Inter-annual variations in the EAC transport produce a detectable sea-level signal in Sydney Harbour, which could provide a useful fisheries index, as does the Fremantle sea level and Leeuwin Current relationship.

Holbrook et al. [2011] showed that the EAC and sea level are affected by Rossby waves. Ridgway et al. [2008] showed that the EAC has increased in south Australia and decreased in north-east Australia. Seasonal, inter-annual and strong decadal changes in the EAC mask the underlying long-term trends related to greenhouse-gas forcing. However, the mean EAC flow is weak [Ridgway and Godfrey, 1994].

Sydney is in the area that the EAC separates into two parts. In addition, the EAC is dominated by cyclonic and anti-cyclonic eddies which make calculating the EAC transport, not its eddies, a bit hard. Cetina-Heredia et al. [2014] identified the core

of the geostrophic EAC current by using the maximum southern geostrophic velocity at 28°S (between the coast and 160∘E) where the EAC is well-concentrated [Ridgway and Dunn, 2003]. This maximum southern velocity then followed and is considered the core of EAC transport. In this study, data provided by Cetina-Heredia et al. [2014], were used for the monthly core EAC transport at 33°S. This dataset is shown in Figure 6.12.

Figure 6.13 shows the decadal variability in the EAC transport and sea level in Sydney using SEEMD and CEEMD. The decadal EAC IMF estimated by SEEMD just shows a gentle peak around 2003 and no changes from 1980 to 1992. The decadal EAC from CEEMD shows a full cycle which is out of phase with sea level from 1982 to 1993 and in phase with sea level from 1995 to 2005. So the EAC decadal results derived from CEEMD and SEEMD show completely different behaviours. For decadal sea-level data CEEMD and SEEMD agree, the only difference is with the estimated amplitude, the SEEMD estimate the decadal amplitude is significantly higher than CEEMD.

Figure 6.14 shows the EAC transport trends from CEEMD and SEEMD. Both interpolations show the EAC transport first increasing then decreasing over the time period (1981–2010). However, SEEMD found that the EAC increased up to 2000, is agreement with the results of Hill et al. [2008]. Hill et al. [2008] used a low-pass-filter method on the wind stress and EAC transport data, and found that the EAC increased due to an increase in wind stress over the period 1950–20003. Ezer et al. [2013] showed that there is an inverse relationship between the Gulf Stream western boundary current and sea level in the mid-Atlantic region. He showed that weakening in the Gulf stream is the result of extreme sea level rise in that region. It is expected that weakening in the EAC west boundary current will also cause a similar increase in SLR. Therefore, according to the SEEMD results in Figure 6.14, from 1981 to 2000, because of the increase in the EAC the sea level in Sydney did not rise significantly, but after year 2000, because of the weakening in the EAC, the sea level rose faster. This explains the fast rise in the SEEMD results for Sydney data after 2000 in Figure 6.9, and is further evidence that SEEMD produced more physically meaningful results.

6.6 Conclusions

In this chapter, the use of EEMD with smoothing splines (SEEMD) and cubic splines (CEEMD) on sea-level data was investigated. The results were as follows.

1. SEEMD obtained fewer IMFs than CEEMD because the smoothing parameter p was set to a small value to favour the accurate extraction of the low frequency IMFs, including the trend.
2. SEEMD also obtained fewer statistically significant IMFs, again because of favouring the low-frequency modes by choice of the smoothing parameter p. Smoothing the data removed the high-frequency IMFs. CEEMD did find statistically significant high-frequency IMFs but they were mostly very close to the noise bound, and therefore less reliable. Although SEEMD obtained fewer statistically significant IMFs, they were all clearer and more reliable. At low-frequency time scales, SEEMD tended to find more statistically significant IMFs than CEEMD.
3. At an intra-decadal time scale, the results of SEEMD and CEEMD agreed quite well.
4. At a decadal time scale, there was some mode mixing in both the SEEMD and CEEMD results but, in general, the results of both methods were similar.
5. For multi-decadal modes with periods of about 20 years, the CEEMD results contained more mode mixing than the SEEMD results. However they both described the multi-decadal variability in a similar way.
6. For multi-decadal modes with periods of more than 30 years, the results of both interpolation methods agreed overall except at the end points.
7. CEEMD and SEEMD showed similar trends for the Newcastle sea-level data. However, the rate of sea-level rise from SEEMD agreed more closely with previous studies. For the Sydney sea-level data, the trends from the two methods differed at the endpoints, causing different estimates of the rate of sea-level rise. The SEEMD rate agreed with other values in the literature. For Port Pirie, the SEEMD and CEEMD results disagreed in both the shape and magnitude of the estimated sea-level rise. SEEMD found the sea-level rise to be nonlinear, with the rate of rise increasing more rapidly after 1990 whereas CEEMD found a relatively slow linear rise in the sea level. Again the results of SEEMD agreed better with existing literature.
8. Sensitivity of EEMD to the interpolation method is again obvious when estimating the acceleration in sea-level rise. At all three locations, the sea-level acceleration found using the two interpolation methods was different in either or both the magnitude of the acceleration and the shape of the acceleration curve. It is not clear in estimating sea-level acceleration which method gives the more reliable results.

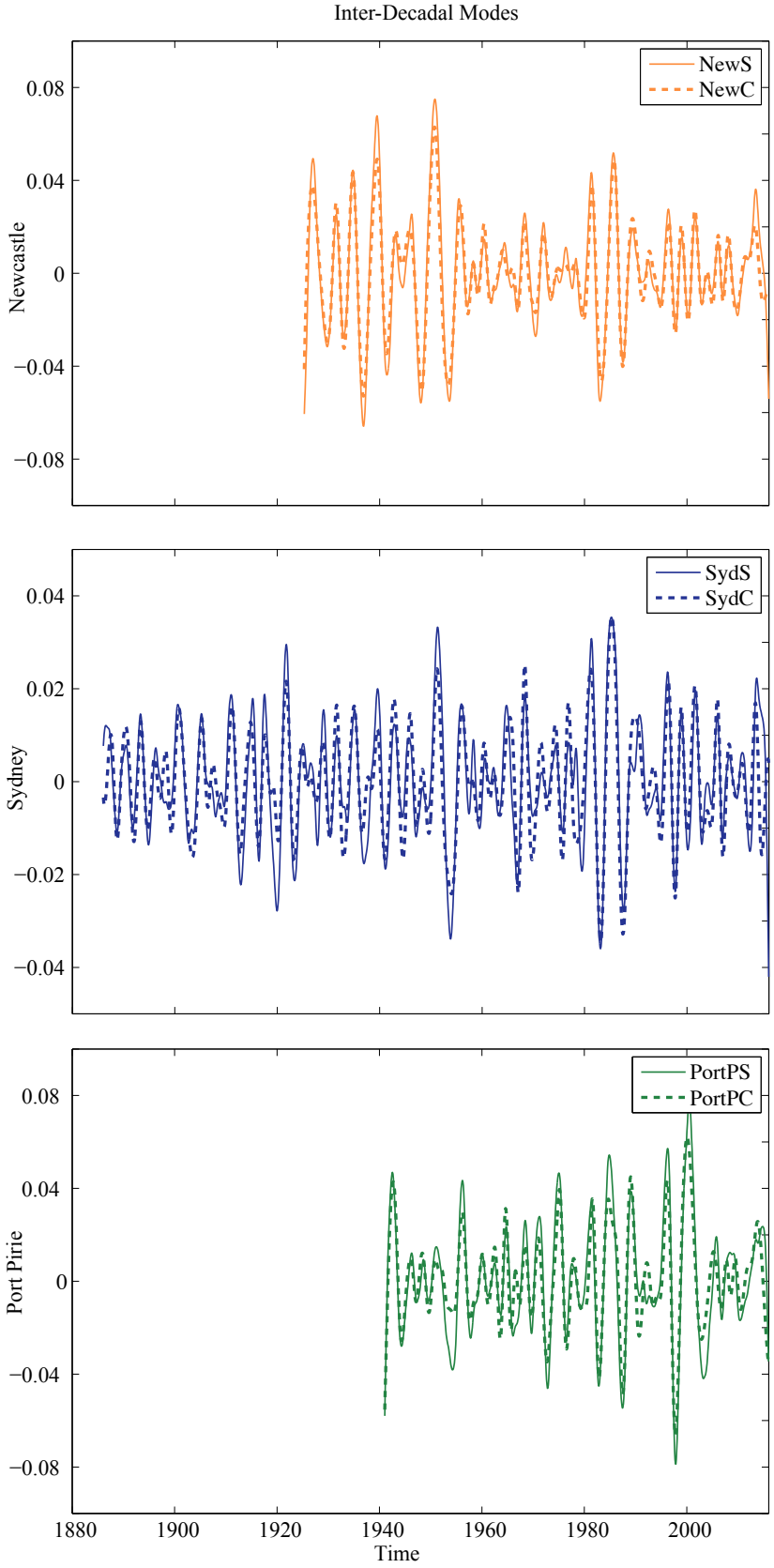

Figure 6.5: Statistically significant IMFs with intra-decadal time period ($\sim$3 years). Vertical scale is in meters. S: smoothing-spline interpolation; C: cubic-spline interpolation.

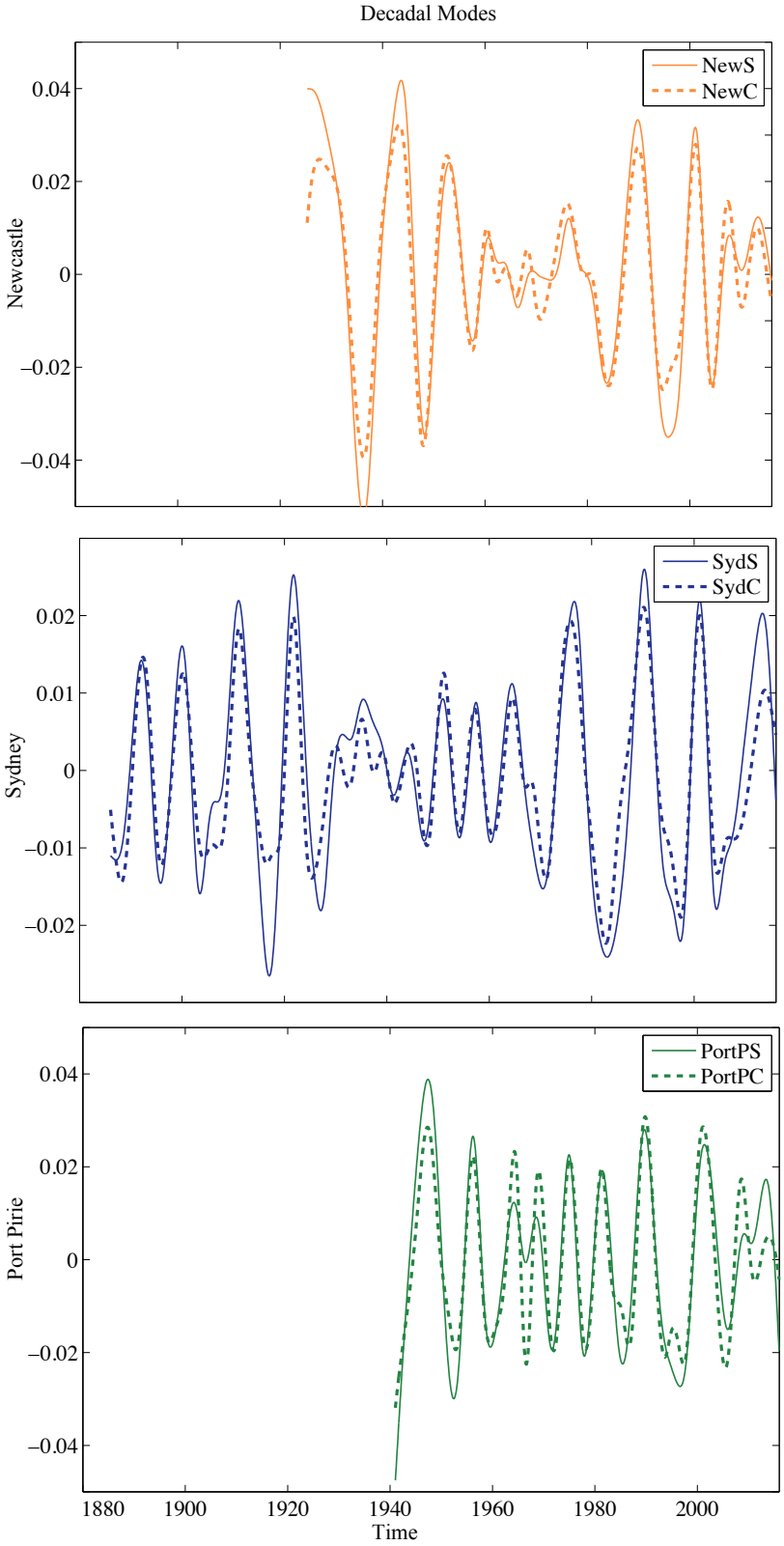

Figure 6.6: Statistically significant IMFs with decadal time period ($\sim$ 10 years). Vertical scale is in meters. S: smoothing-spline interpolation; C: cubic-spline interpolation.

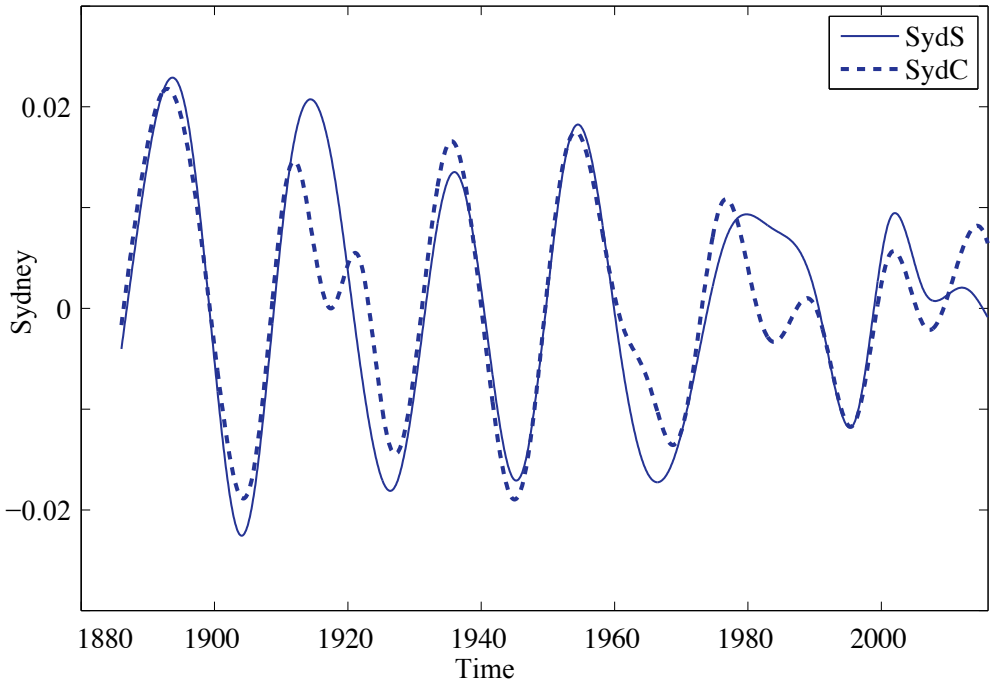

Figure 6.7: Statistically significant IMFs with multi-decadal 1 period ($\sim$20 years) for the Sydney data. Vertical scale is in meters. S: smoothing-spline interpolation; C: cubic-spline interpolation.

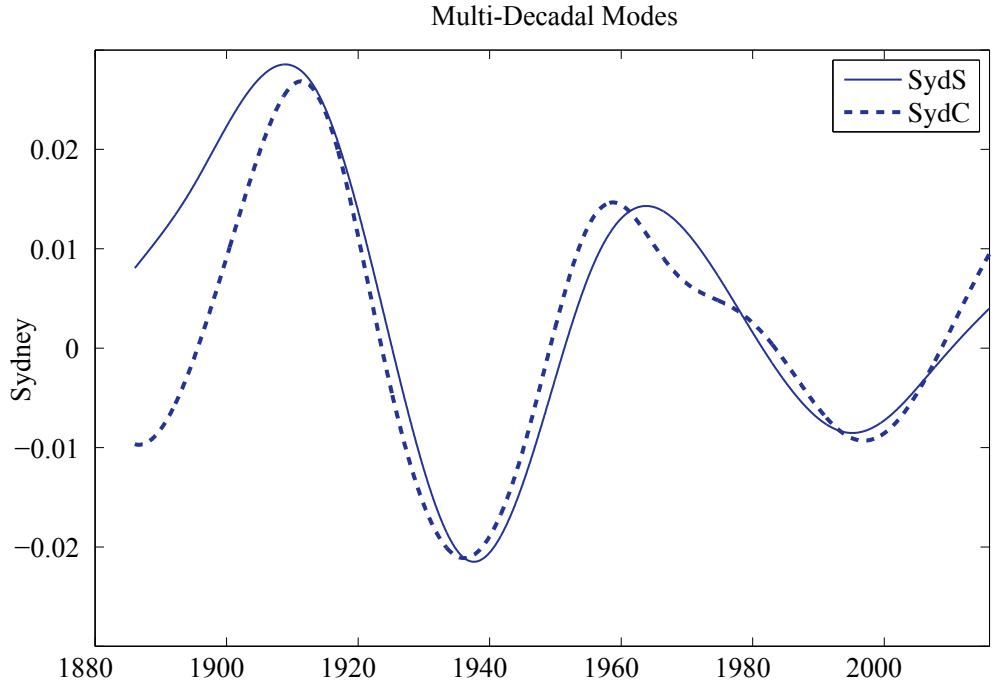

Figure 6.8: Statistically significant IMFs with multi-decadal 2 period ($\sim$30 years). Vertical scale is in meters. S: smoothing-spline interpolation; C: cubic-spline interpolation.

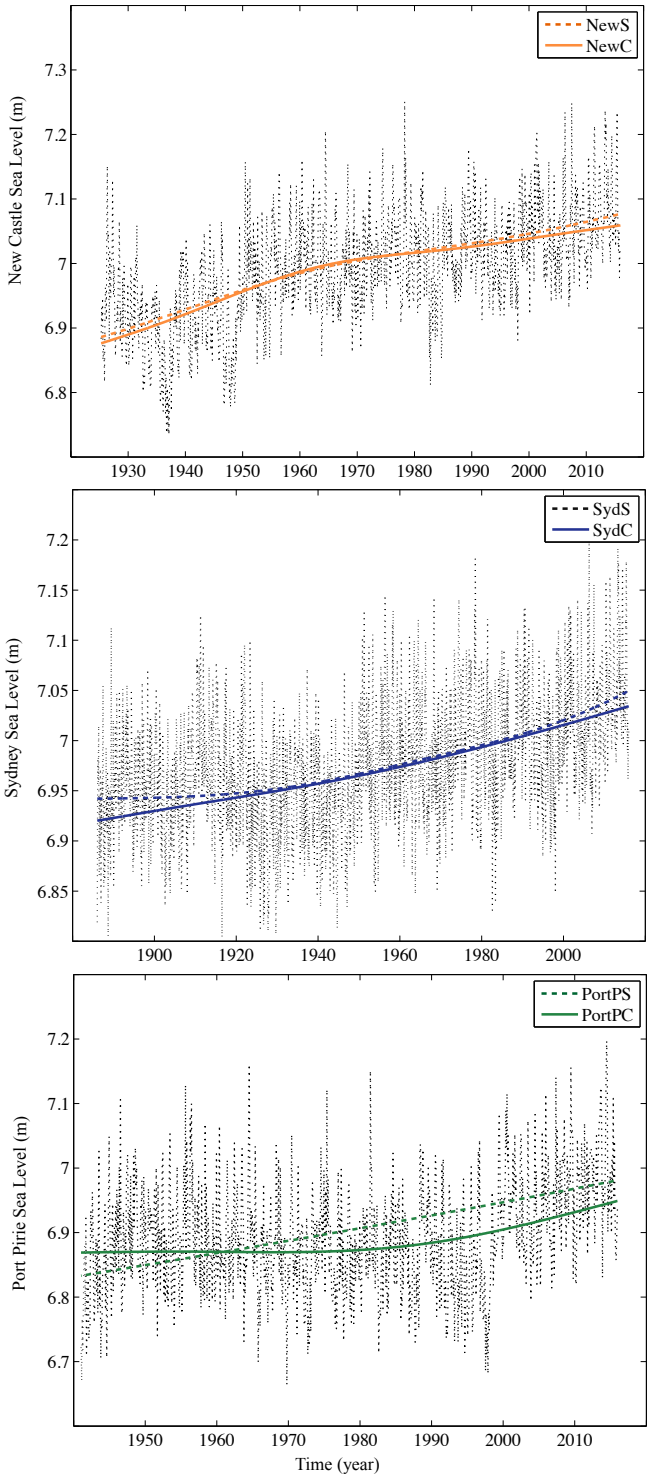

Figure 6.9: Sea-level trends (residual modes) for the Newcastle, Sydney and Port Pirie datasets using the EEMD method with smoothing-spline (S) and cubic-spline (C) interpolations. The dots show the data. Vertical scale is in meters.

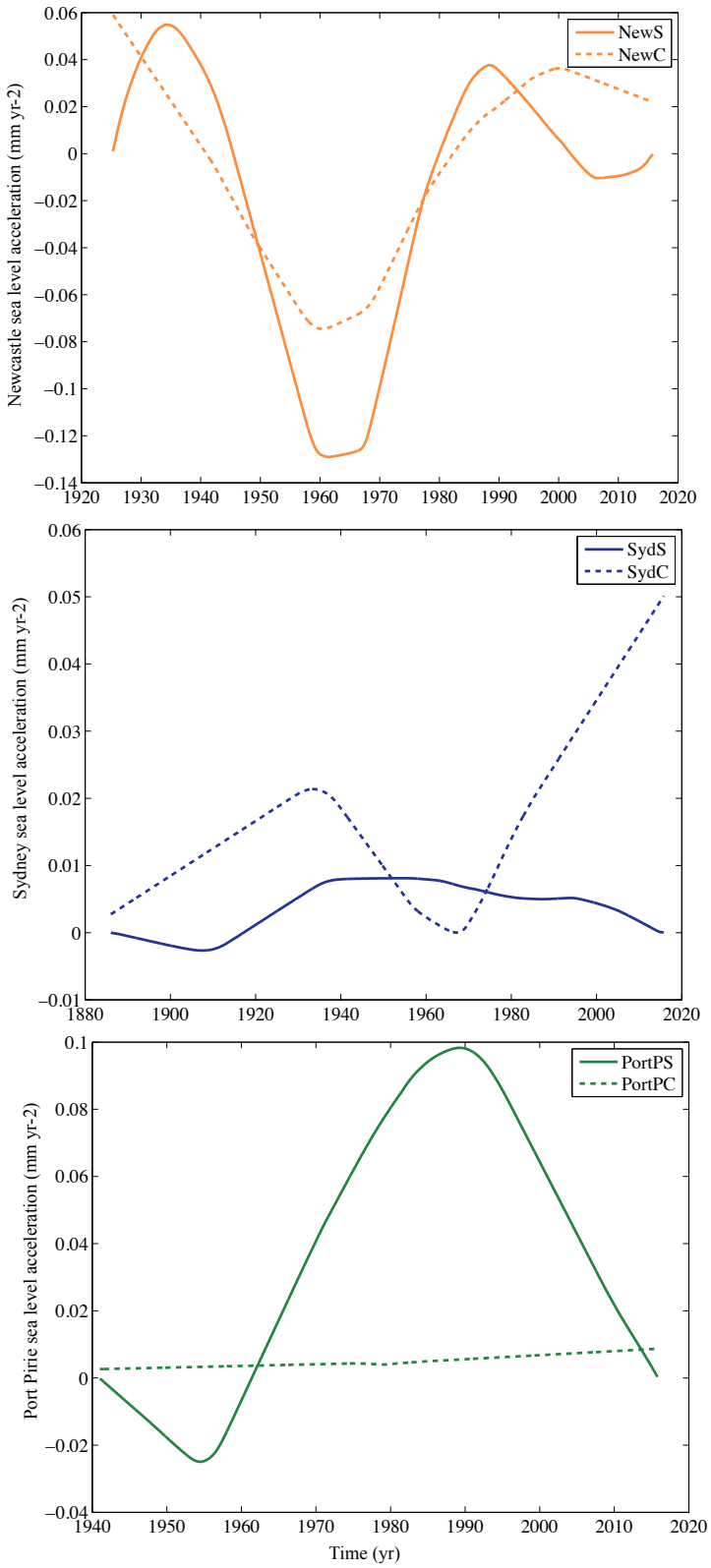

Figure 6.10: Sea-level acceleration ($\mathrm{mm\,yr^{-2}}$) for the Newcastle, Sydney and Port Pirie datasets using the EEMD method with smoothing-spline (S) and cubic-spline (C) interpolations.

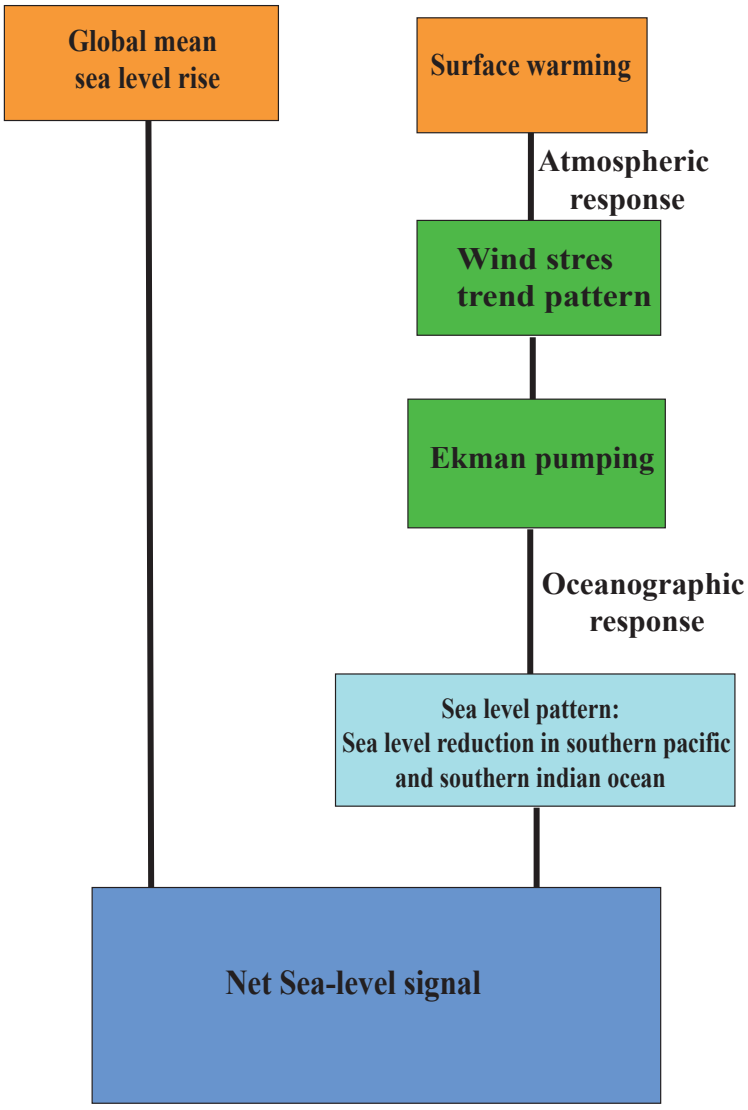

Figure 6.11: Schematic outline of the relationships between global warming and wind stress, and its impact on sea-level variability. Modified from Timmermann et al. [2010].

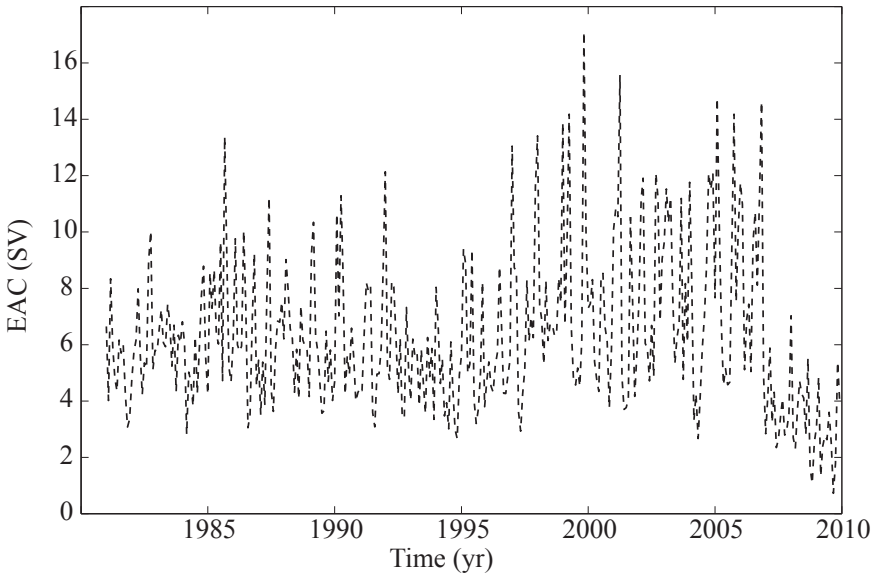

Figure 6.12: EAC core transport for 33°S from 1980 to 2010.

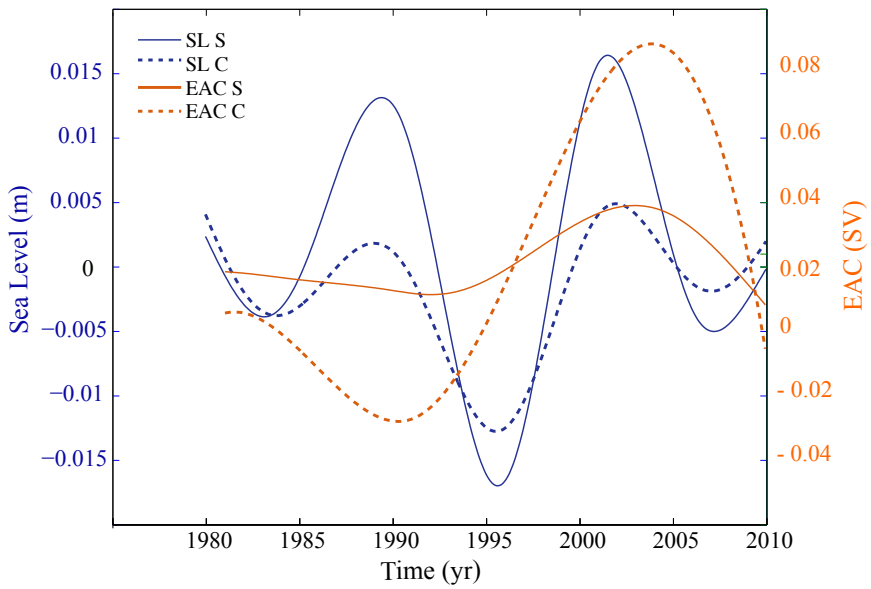

Figure 6.13: Sydney sea-level variability and Sydney EAC transport (from 1980).

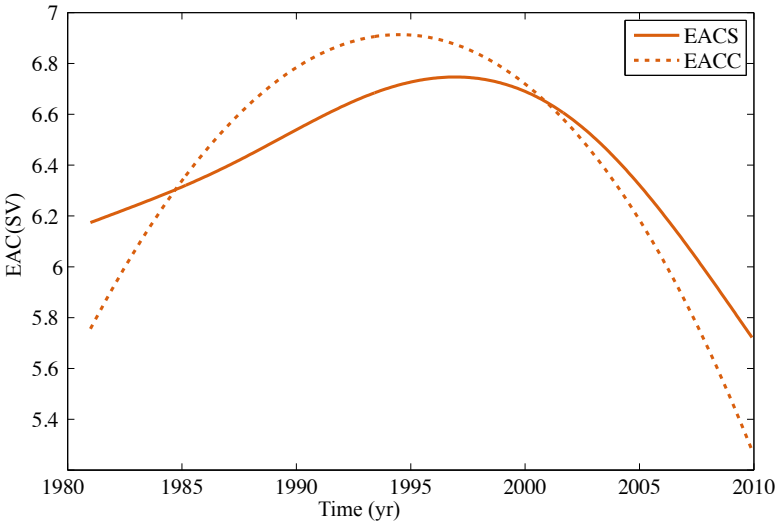

Figure 6.14: The EAC transport trends using SEEMD and CEEMD.

Chapter 7

Application of Empirical Mode Decomposition to Forest Fire Danger Index Data

7.1 Introduction

Australia has unique ecosystems, vegetation and species. One of the major hazards in Australia is bushfires which threaten these ecosystems [Lucas et al., 2007].

Bushfires are of three different types. The first type is crown fire, mainly affecting the tops of the woody vegetation. This fire is strongly influenced by wind. The second type is surface fire, which burns grass, bushes and other vegetation close to the ground. The third fire type is ground fire, a flameless fire that burns through a thick layer of the ground containing organic matter and roots. Surface fires initiate the other two types of fire [Scott and Reinhardt, 2001]. To assist in controlling bushfires, it is common to use a fire danger rating system which takes into account the vegetation and climate of the region.

Bushfires have complex interactions with vegetation, soils, weather and human activities. Each of these elements can change due to climate variability, thereby changing the bushfire regimes. Resolving the relative roles of each of these key factors that influence fire dynamics demands comparative studies that illuminate how fire regimes vary across climatic gradients [Fox-Hughes et al., 2014]. Many studies on the long-term variability in weather variables such as temperature, rain and wind have been conducted. However, the variability in the drivers of bushfire behaviour over different time scales has not been paid much attention in the literature.

In this chapter, the variability in the Forest Fire Danger Index (FFDI) is investigated over various time scales using the EMD method. The parameters taken into account in calculating it are also investigated in a similar way, to examine which are the dominant factors driving Fire Danger over different time scales.

7.1.1 Fire-danger rating system as a control tool

To minimise the damage caused by fire, the risk of fire needs to be assessed, motivating the introduction of fire-danger rating systems; these are a combination of the fire-danger factors, ranked into classes for public warning and preparation of the support resources in cases of high fire danger [Tolhurst et al., 2010].

Fire occurrence depends on meteorological conditions, vegetation status and topography [Chandler et al., 1983]. Some of the relevant meteorological conditions are the temperature, relative humidity, and wind speed and direction in the area. Vegetation plays an important role as fuel for bushfires.

Meteorological conditions can change during a bushfires, and intensify the fire or modify its behaviour [Sharples, 2009]. Wind speed and direction depend on topography, and vegetational local heating and cooling. The topography may also cause dramatic changes in fire behaviour as a fire progress over the terrain [Sharples et al., 2012]. In addition, the fire itself may influence the environment and thus the fire behaviour; heating from the fire can modify or produce local winds, contributing to atmospheric instability and causing cloud development [Sharples, 2009].

There is a variety of fire-danger rating systems used around the world, such as: the Forest Fire Danger Index (FFDI: McArthur [1967]); Forest Fire Behaviour Tables (FFBT: Sneeuwjagt and Peet [1985]); the Fire Weather Index (FWI: Van Wagner et al. [1987]); and the US National Fire Danger Rating System (NFDRS: Deeming [1972]). The two most common rating systems are the FWI and FFDI.

The Canadian Forest Fire Danger Rating System is based on experimentation and observation of fire behaviour. This rating system works best on areas with pine vegetation, which is widespread in Canada [Van Wagner et al., 1987].

The Australian McArthur Forest Fire Danger Index or FFDI is based on an empirical model of fire behaviour in open Eucalyptus forests; this rating is applied broadly to all forest types, shrubs, mallee and eucalyptus. The FFDI requires meteorological information such as temperature, relative humidity, rainfall and wind speed. The FFDI lies between 1 and 100^{+}, where 100^{+} corresponds to the catastrophic fire conditions on Black Friday in 1939 [Tolhurst et al., 2010].

A study by Viegas [1999] that compared five different indices in six different regions of France, Italy and Portugal (the FFDI and NFDRS were not included) showed that the FWI System was the best performer among the five systems. Matthews [2009] compared the Fire Forest Danger Index, the Forest Fire Behaviour Tables and the Fire Weather Index for Australian vegetation for use in eucalypt forests. They calculated the three fire-danger indices over a range of values for fuel availability,

the moisture content in the fuel (fuel moisture) and wind. They found that all three indices were correlated, with some variability. The main variability was a result of how or whether fuel availability was incorporated into each index. The FFDI and FWI, both of which take fuel availability into account, were significantly correlated. The FFDI is more suitable for eucalypt forest, the FWI for pine forest. As the present study seeks to be relevant to Australia, the FFDI was chosen as the fire-danger rating system to analyse.

The equation for the McArthur FFDI is [Noble et al., 1980]:

$$F = 2\exp\big(-0.450 + 0.987\ln(D) - 0.0345H + 0.0338T + 0.0234V\big), \qquad (7.1)$$

where F is the fire danger rating (in the range 0–100^{+}) and D is the drought factor, which represents the proportion of the fuel available to burn, expressed as a number between 0 to 10. It is calculated from soil moisture and antecedent rainfall [Griffiths, 1998]). As such, can be considered as a measure of fuel availability that incorporates aspects of meteorological and horological drought. H is percent relative humidity, T is air temperature (°C) and V is average wind speed ($\mathrm{km\,h^{-1}}$).

Fire risk	**FFDI**	**Grassland**
Moderate	0–11	0–11
High	12–24	12–24
Very High	25–49	25–49
Sever	50–74	50–99
Extreme	75–99	100–149
Catastrophic	100^{+}	150^{+}

Table 7.1: Bushfire danger levels and the corresponding FFDI ranges.

The moisture content of a fuel sample is defined as the relative mass of moisture in the fuel sample compared to the oven-dried mass of the sample, and is expressed as a percentage. Fuel moisture can change in response to latent heat effects, vapour exchange and rainfall [Nelson Jr, 2000, Viney, 1991]. The FFDI range is divided into six categories, each of which indicates a level of fire risk to report to the public (Table 7.1). Sharples et al. [2009] introduced a simple fuel-moisture index as a function of temperature and relative humidity:

$$\mathrm{FMI} = 10 - 0.25(T - H). \qquad (7.2)$$

FMI is a good proxy for fuel moisture, assuming that no free water is present [Sharples et al., 2009].

7.1.2 FFDI and climate change

Millions of years ago (early Pleistocene) in south-east Australia, which was covered by rain forest (now occupied by dry Eucalyptus forest), bushfires were not as common as nowadays [Sniderman and Haberle, 2012]. Sniderman and Haberle [2012] showed that fire frequency increased by an order of magnitude during the climate-driven transitions to sclerophyllous Eucalyptus-dominated vegetation. Mooney et al. [2011] used sedimentary charcoal records from Australasia from the last 70,000 years to show that there was considerable spatial and temporal variability in fire occurance during deglaciation.

Bradstock et al. [2014] examined the changes in fire occurrence from 1975 to 2009 by examining trends in the areas burned. They concluded that an increase in fire, due to climatic warming and drying, was more likely to occur in the moist, temperate forests near the coast than in the arid and semi-arid woodlands of the interior. Significant warming and drying occurred across most of the 32 bioregions examined. The results were mostly consistent with predictions, with an increase in area burned in seven of eight forest bioregions, whereas the area burned either declined (two) or did not change significantly (nine) in the drier woodland bioregions. In 12 woodland bioregions, data were insufficient for analysis of temporal trends in fire. Increases in fire, attributable mostly to warming or drying, were confined to three bioregions.

There are a number of studies that use climate-change scenarios to predict fire dangers and its contributing factors into the future. Beer and Williams [1995] investigated the changes in fire danger in Sale, Victoria, Australia under conditions of doubled carbon dioxide concentration in the atmosphere. They found more days with high fire danger, with the annual mean FFDI and annual mean of daily minimum relative humidity strongly negatively correlated. They expanded their study to the whole country to investigate the spatial variation in FFDI. Coastal areas, especially in the north and northeast, had decreased values of the FFDI. They suggested that changes in relative humidity play a strong role in changing the FFDI. They used a One-Box Model [Livingstone, 1993] to investigate the changes in relative humidity to determine whether the relative humidity change depends on the relationship between actual and potential evaporation, the expected changes in wind, the present value of humidity and/or the present value of the evaporation rate. Their estimate predicts that relative humidity always decreases as a result of warming. This means widespread increases in the FFDI.

Pitman et al. [2007] explored the impact of future climate change on the risk of forest and grassland fires over Australia in the month of the January using a high-resolution regional climate model. They used two future emission scenarios (rela-

tively high and relatively low) for 2050 and 2100. Their results showed a consistent increase in regional-scale fire risk over Australia, driven principally by warming and reductions in relative humidity in all simulations, under all emission scenarios and at all time periods. The probability of severe fire risk increased by around 25% in 2050 compared to the present day under both relatively low and relatively high emissions; this increased by a further 20% by 2100 under the relatively low emission scenario.

Grose et al. [2014] found that a warmer climate is likely to increase the frequency and impact of the very worst fire weather days for south-east Tasmania. Flannigan et al. [2009] suggested that fire seasons will lengthen in temperate and boreal regions. There will be a general increase in area burned and fire occurrence but a lot of spatial variability, with some areas of no change or even decreases in area burned and occurrence.

King et al. [2011] investigated fire and carbon dynamics for projected warmer and drier climates in the south-eastern Australian high country. A carbon-accounting model FullCAM [Richards, 2001] and FIRESCAPE [King et al., 2011], a landscape fire-regime simulator, were combined and used to simulate several fire management options under three climate scenarios: the recent climate (1975–2005); a moderate climate projected for 2070; and a more extreme climate projected for 2070. For warmer and drier climates, their model simulations predicted: an increase in fire incidence; larger areas burned; higher mean fire intensities; shorter fire cycle lengths; a greater proportion of fires burning earlier in the fire season; a reduction in carbon stores; a reduction in carbon sequestration rates; and an increase in the proportion of stored carbon emitted to the atmosphere. Prescribed burning at historical or twice historical levels had no effect on the fire or carbon dynamics. In contrast, increasing the initial attack success partially offset the adverse effects of warmer and drier climates on fire activity, but not on carbon dynamics. For the south-eastern Australian high country, simulations indicated that fire and carbon dynamics are sensitive to climate change, with simulated fire management only being able to partially offset the adverse effects of a warmer and drier climate.

Liu et al. [2010] investigated the trend in global wildfire potential under climate change. They used the Keetch-Byram Drought Index (KBDI) to measure the fire potential under projected changes in temperature and precipitation at the end of this century (2070–2100) from general circulation models (GCMs) for present and future climate conditions. They showed that future wildfire potential increases significantly in the United States, South America, central Asia, southern Europe, southern Africa and Australia; fire potential moves up by one level in these regions, from currently low to future moderate potential or from moderate to high potential. Relative

changes are the largest and smallest in southern Europe and Australia, respectively. The annual period with the KBDI greater than 400 becomes a few months longer. The increased fire potential is mainly caused by warming in the US, South America, and Australia, and by the combination of warming and drying in the other regions. Sensitivity analysis shows that future fire potential depends on many factors such as the climate model and emission scenario used for climate-change projections.

Barbero et al. [2014] investigated the climate-fire relationship on an inter-annual timescale. They provided a spatio-temporal analysis using data with a 60-km spatial and weekly temporal resolution. They found an increase in the occurrence of very large fires in southwestern and southeastern US from 1984 to 2010. They also found that the probability of very large fires reached its peak during spring in the eastern and southern half of the US, and in mid to late summer for most of the northwest US.

Mariani and Fletcher [2016] found that heatwaves did not have an extreme influence on fire activity in the short term. They suggested that heatwaves might contribute to the landscape availability of fuel by drying out any remaining wet soaks or creeklines in the landscape and affecting the moisture content of live fuels. Williams et al. [2001], in a doubled-CO_2 climate simulation, found that Alice Springs, Miles, Sale and Mildura all had significant changes in the probability distribution of the daily FFDI, and the seasonal FFDI increased throughout both the northern and southern fire zones. The seasonal FFDI was most sensitive to large changes in temperature. The seasonality of fire danger was also affected, but the degree of change varied widely. Of the eight detailed sites, only at Katanning and Hobart did the length of the fire season not change. Hasson et al. [2009] found that using different models for predicting fire probabilities gave large differences in results. Williams et al. [2001] mentioned that the limitations of the scenarios, because of the assumptions built into the GCM, impact the interpretation of fire-danger simulation.

As mentioned above, there have been some studies on the seasonal variability of fire rating systems and future prediction based on climate-change scenarios but there have not been many studies done on the decadal and multi-decadal time scale behaviour of fire rating systems. This long-term behaviour would help us understand the broad picture of how fire danger changes over time, and which of its drivers are the most important across different time scales.

7.2 EEMD analysis of Forest Fire Danger data

Data on the Forest Fire Danger Index (FFDI) and its contributing factors from three stations, Sydney, Canberra and Melbourne, are used in this study. These

data were obtained from the Centre for Australian Weather and Climate Research; details are given by Lucas [2010]. Lucas [2010] used historical FFDI datasets and, to have a uniform quality in the data, only used the data from June 1972 onward. He mentioned that the main problem with the data is the inhomogeneity in the wind measurements, a consequence of using different instruments and methodologies over time.

There were some missing data in the FFDI record, which we modified by taking means or interpolating. Also in this study, only the fire season (from 1 October until 31 March) is considered. The data from each station contain FFDI, maximum air temperature, relative humidity (RH) and wind speed at 3pm, and the drought factor from 1972 to 2014 (Figure 7.1).

In Figure 7.1, the top panel shows the FFDI for the three stations. The Canberra FFDI has larger fluctuations than the other two stations. The drought factor and maximum temperature are similar at all stations before 2000; after 2000, the drought factor at Sydney behaves slightly differently from the other two stations. As Lucas [2010] mentioned, there are some obvious inconsistencies in the mean of the Sydney wind data before and after the mid-1990s, due to using different methods and instruments to measure the wind speed. Relative humidity was higher in Sydney than in both Melbourne and Canberra, probably due to the fact that the Sydney weather station is on the coast, and so is affected by maritime air.

The EEMD method with smoothing-spline interpolation (SEEMD) was applied to all the data in Figure 7.1; smoothing splines gave the most reliable results for the synthetic data (Chapter 4), temperature data (Chapter 5) and sea-level data (Chapter 6). The significance tests were then used on the resulting IMFs to obtain the statistically significant IMFs (Chapter 5). Figure 7.2 shows the statistically significant IMFs; the first five IMFs were statistically significant for all datasets.

The first to fifth IMFs represent, respectively, seasonal, 2, 6, 8, 14 year variability. Due to our interest in the long-term variability of FFDI in this study, we focus on IMF 4 (decadal IMF) and IMF 5 (multi-decadal IMF). In Figure 7.2, Sydney wind-speed IMF 5 and 6, Canberra relative humidity IMF 5 and 6, and temperature IMFs 4 and 5, have the same period but different energies. This is evidence of mode mixing in the results of the EEMD method. In this study, the IMFs attributed to mixing issues have been added together. However, some of the IMFs had very small amplitudes, so that adding them to another IMF did not change the results significantly. Using more ensemble members might improve the results, but this was not attempted.

7.3 Decadal and multi-decadal variability in FFDI data and its contributors

The statistically significant IMFs from SEEMD with 300 ensemble members and smoothing parameter $p = 0.01$ for the individual contributors for Sydney, Canberra and Melbourne are presented in Figure 7.2. The results for temperature, wind speed, relative humidity and drought factor are compared with the corresponding FFDI IMFs.

7.3.1 Temperature

As shown in Equation 7.1, fire risk increases exponentially with temperature. Cary et al. [2012] found that average temperatures will increase by about 0.6°C to 4°C, depending on the assumed emission scenario, time frame and locality.

Hasson et al. [2009] investigated the impact of temperature on fire events. They used datasets for the period 1960–1999 to analyse the climate of the 20th century, then simulations for the periods 2046–2065 and 2081–2100 to predict changes in the middle and late 21st century. Their results showed that the frequencies of fire events will increase from around one event every two years during the late 20th century to around one event per year in the middle of the 21st century and one to two events per year by the end of the 21st century.

Temperature decadal variability

Figure 7.3 shows the normalised decadal variability found here for temperature and FFDI from 1972 to 2014 using SEEMD. In Canberra and Melbourne, the temperature and FFDI decadal variabilities are similar and in phase. In Sydney, however, the temperature and FFDI are not in phase (Figure 7.3 top panel); from 1972 to 1985, the temperature initially leads by three years but from 1985 to 1995, temperature and FFDI are more or less in phase. From 1995 until 2010, both FFDI and temperature show mode mixing, causing a small correlation, $R = 0.55$ (Table 7.2).

The peaks in both FFDI and temperature in the years 1983, 1987–88 and 1998 seem to be due to a strong El Niño effect in these years. Verdon et al. [2004] showed that the meteorological data can be strongly affected by El Niño.

Temperature multi-decadal variability

Figure 7.4 shows the multi-decadal variability of temperature and FFDI for Sydney, Canberra and Melbourne. In Canberra and Melbourne, FFDI and temperature behave similarly and are highly correlated, with R values of 0.75 and 0.8, respectively. The Sydney multi-decadal temperature is out of phase with FFDI before 1990 with

$R = -0.73$. After 1990, the temperature is in phase with FFDI with $R = 0.91$. Canberra temperature from 1985 to 2003 exhibits almost no change, which reduced R to 0.68. In Melbourne, the temperature and FFDI multi-decadal modes are in phase and highly correlated ($R = 0.84$).

Ahmed et al. [2013] found that at multi-decadal to centennial scales, temperature showed a distinct spatial variability. However, they found similarity in temperature behaviour at this time scale within each hemisphere. Given this, its expected that all three stations would have similar temperature variability. The results at the decadal time scale for all stations and at the multi-decadal time scale for Canberra and Melbourne match with those in Ahmed et al. [2013]. However, the anomalous behaviours for Sydney around 1980 and Canberra from 1985 to 2000, are more likely to be artefacts of the method.

Verdon et al. [2004] showed there is a strong relationship between temperature on a range of time scales and Forest Fire Danger Index (FFDI). This is also proved shown by the results from SEEMD and the high correlation between temperature and FFDI.

7.3.2 Wind speed

The decadal and multi-decadal variabilities in the wind speed are shown in Figures 7.5 and 7.6, respectively.

Wind-speed decadal variability

From 1972 to 1985, wind speed is out of phase with FFDI at all three stations (Figure 7.5). After 1985, Sydney decadal wind speed becomes in phase with decadal FFDI variability ($R = 0.32$). However, Canberra and Melbourne wind speed both remain out of phase with decadal FFDI until about 1990. The decadal wind-speed IMF for both these stations has a low correlation with the decadal FFDI (Canberra: $R = 0.06$; Melbourne: $R = 0.17$).

Wind-speed multi-decadal variability

The multi-decadal wind-speed variability in Canberra and Melbourne is in phase with the FFDI over the whole period, with a high correlation between multi-decadal wind speed and FFDI, ($R = 0.81$ and $R = 0.78$, respectively). In Sydney the wind speed lags the FFDI by about five years (Figure 7.6) and the correlation between wind speed and FFDI is low ($R = 0.24$). This could be due to the inconsistency in the wind-speed data for Sydney which was mentioned in Section 7.2, and will be discussed further in Section 7.5.

Lucas [2010] suggested that long-term behaviour found from wind data is likely to be unrealistic because of the many factors that can affect the wind data. These factors are also likely to change over years, such as the growth/removal of trees and building construction. Considering the Lucas [2010] study and the obvious inconsistency in the Sydney wind-speed data, the wind-speed patterns at all time scales should be considered with a suitable degree of skepticism.

7.3.3 Relative humidity

In Equation 7.1, the relative humidity and FFDI have an inverse relationship. Warming due to climate change will lower relative humidity [Murphy and Timbal, 2008]. Lower relative humidity means a relatively drier environment, and a drier environment is more likely to succumb to bushfires. Matthews et al. [2012] used a suite of models to examine the links between climate, fuels and fire behaviour in dry eucalypt forests in south-eastern Australia. They found that a warm and dry climate produced lower amounts of fine fuel, but a greater propensity of this fuel to burn due to lower moisture content. Changing the fuel load had only a small effect on fuel moisture. A warmer, drier climate also increased the rate of spread, an important measure of fire behaviour. Matthews et al. [2012] observed a strong climate signal in fuel moisture, with a shorter winter period, lower fuel-moisture levels and a greater number of fire days in warm-dry years, in all months. However, the effect of climate change on the amount of fuel is not yet fully understood.

Relative humidity decadal variability

Figure 7.7 shows the decadal variability of relative humidity and FFDI. At all three stations, the relative humidity and FFDI are out of phase and have a strong negative correlation, Sydney $R = -0.79$, Canberra $R = -0.88$ and Melbourne $R = -0.87$.

Relative humidity multi-decadal variability

Figure 7.8 shows that, at all three stations, the relative humidity and FFDI are out of phase except in Sydney from 1972 to 1985. However, as mentioned in Chapter 4, EMD is prone to endpoint effects. Therefore it is likely that the relative humidity from 1972 to 1985 is a method artefact.

These results suggest that relative humidity has a strong effect on FFDI variability at both decadal and multi-decadal time scales.

7.3.4 Drought factor

The drought factor depends strongly on environmental conditions such as precipitation [Keetch and Byram, 1968]. Trenberth et al. [2014] found that the occurrence and length of droughts are changing with climate change. Increased heating from global warming may not cause more frequent droughts but it is expected that, when droughts occur, they are likely to set in quicker and be more intense.

Drought-factor decadal variability

Figure 7.9 shows the decadal variability of drought factor and FFDI. In Canberra and Sydney, FFDI is in phase with drought factor but with a two-year lag. In Melbourne, the FFDI and drought factor are in phase with no lag, except from 1990 when the drought-factor decadal IMF shows mode mixing. Table 7.2 shows that the Sydney drought factor has smaller correlation with FFDI than the other two stations at this time-scale (Sydney: $R = 0.19$, Canberra: $R = 0.55$, Melbourne: $R = 0.78$).

Drought-factor multi-decadal variability

The multi-decadal drought factor mode in Sydney is out of phase with FFDI (negatively correlated, $R = -0.61$) whereas in Melbourne it is in phase and highly correlated with FFDI ($R = 0.85$). For Canberra, the drought factor is in phase with FFDI until 1987; after 1987, the drought factor becomes out of phase with FFDI. These differences make the drought factor not highly correlated to FFDI (Table 7.2).

Nicholls [2004] showed that, despite the similarity in rainfall especially in 1982, 1994 and 2002, the nature of the drought was not the same along the entire east coast of Australia. This might be the reason for the different behaviour in the multi-decadal drought factor in Sydney, as a coastal station, compared with the other two stations.

This results indicate that drought factor is a more dominant driver of FFDI at decadal time scales, but is less important at multi-decadal time scales, except at Melbourne.

7.3.5 Long-term variability

Figures 7.11 to 7.14 show the trends for all the FFDI contributors at all three stations. Figure 7.11 shows the drought factor and FFDI trends. In Melbourne and Canberra, the drought factors both increase after 1990 but relative humidity decreases (Figure 7.14). This behaviour is expected from Equation 7.1. However, in Sydney, the drought factor seems to increase linearly, its behaviour is similar neither to FFDI behaviour nor to the drought factors at the other two stations, both of which have a minimum around 1990. The reason could be again the position of Sydney in a coastal area which was less affected by drought than the other two stations. Differences in the weather and climate at Sydney compared to the other stations may have contributed to the differences observed. Sydney has a coastal climate, which is more humid and therefore less prone to changes in the drought factor. On the other hand Melbourne and Canberra data are gathered from inland areas,

where the impact of drought factor is stronger. We suspect that coastal climate has a significant impact on the decadal and multi-decadal relationship of FFDI and its drivers especially drought factor, but we are not certain. Further study is necessary on this topic.

The long-term behaviour of temperature for all three stations is consistent with the FFDI long-term behaviour (Figure 7.12). All stations, both the temperature and the FFDI increase after 1990. This corresponding behaviour is expected from Equation 7.1.

Only the long-term variability in wind speed in Melbourne is completely consistent with the FFDI. In Canberra, wind speed is almost constant before 1990; after 1990 both FFDI and wind speed increase (Figure 7.13). For Sydney, the wind speed and FFDI show some consistency but the accuracy of the long-term wind behaviour is questionable [Lucas, 2010]. For Sydney, wind speed after 1990 increases; this does not agree with previous studies such as England et al. [2014], and also differs from the results at the other two stations. This could again be due to inconsistency in the wind data at the Sydney station. The fact that Sydney wind data had an abrupt change around 1990 might cause this (see Section 7.5). Relative humidity before 1985 decreases in Sydney but does not change much for Canberra and Melbourne; during this time, the FFDI stays the same or decreases. From 1985 to 1995, the relative humidity and FFDI at all stations do not change much; after 1995 at all stations relative humidity drop to its minimum and FFDI increase significantly (Figure 7.14). This is the reason for negative correlation coefficients in Table 7.2.

7.4 Relationship between FFDI and its contributors

Table 7.2 shows the correlation coefficients between FFDI and its contributors at the different time scales. In the decadal variability (IMF 4) for all stations, temperature has the highest positive correlation with FFDI, and relative humidity is strongly negatively correlated. In Melbourne and Canberra for the decadal time scale, drought factor is also highly correlated to FFDI. This shows than for these three stations at the decadal time scale, the main drivers of the FFDI for the areas away from the coast are the temperature, relative humidity and drought factor, whereas for the area near the coast, drought factor is less important.

At the multi-decadal time scale (IMF 5), again the results for Sydney are different from Canberra and Melbourne. In Sydney, drought factor and relative humidity are both negatively correlated with FFDI; wind speed and temperature are not significantly correlated. The negative correlation of drought factor with FFDI seems

IMF	Station	DF	RH	V	T
Decadal	Sydney	0.19	−0.79	0.32	0.55
	Canberra	0.55	−0.88	0.06	0.75
	Melbourne	0.78	−0.87	0.17	0.80
Multi-decadal	Sydney	−0.63	−0.46	0.24	0.23
	Canberra	0.43	−0.93	0.81	0.68
	Melbourne	0.85	−0.93	0.78	0.84
Trend	Sydney	0.94	−0.94	−0.46	0.98
	Canberra	0.96	−0.66	0.67	0.88
	Melbourne	0.89	−0.99	0.99	0.97

Table 7.2: Correlation coefficients between FFDI and drought factor (DF), relative humidity (RH), wind speed (V) and temperature (T) at the three stations.

odd, but, the evidence is in Figure 7.10. More studies need to be done on the relationship between drought factor and FFDI for coastal region, at this time scale.

For Melbourne at the multi-decadal time scale, all variables are highly correlated, with relative humidity correlated negatively. It seems that in Melbourne, the multi-decadal variability in all the contributors significantly impacts the fire danger. This shows fire danger in Melbourne is quite sensitive to climate change. For Canberra, wind speed and temperature have the highest correlation with FFDI at the multi-decadal time scale.

According to the results in Table 7.2, at the multi-decadal time scale the main driver of FFDI is different from one station to another. It seems that the potential impact of climate change on fire danger is strongly related to the geographical situation of the area. More studies need to be done to fully understand the impact of multi-decadal variability on fire danger.

For the long-term trend at all three stations, the FFDI is highly positively correlated with drought factor ($R > 0.89$) and temperature ($R > 0.88$) and highly negatively correlated with relative humidity (Table 7.2). In Melbourne, long-term variability in all contributors has a strong relationship with fire danger. This again emphasises the sensitivity of Melbourne fire danger to climate change. In Sydney, wind-speed long-term variability shows a negative relationship with FFDI. However, this might be due to inconsistency in the wind data in Sydney. This will be investigated in the future.

Table 7.2 shows that for all three stations, relative humidity has strong relationship with FFDI. According to Equation 7.2, when both temperature and relative humidity are strongly correlated with FFDI, changes in FFDI are related to changes in the fuel moisture. This happens at a decadal time scale for Canberra and Melbourne.

In multi-decadal time scale, for Melbourne and for long term variability in Sydney and Melbourne.

In Melbourne and Sydney at decadal and multi-decadal time scales, the relative humidity has a stronger impact on the FFDI than temperature, this means more fuel moisture at decadal and multi-decadal time scales was available for these two stations. Long-term variability in Canberra shows that temperature has a stronger impact on the FFDI than relative humidity, which means the in long-term, fuel moisture will be reduced in Canberra.

7.5 Non-stationarity of the Sydney data and its impact on EEMD results

In the previous section, the inconsistency of the Sydney wind-speed data, due to instrumental and method changes in measuring wind speed over decades was mentioned; this inconsistency caused a corresponding inconsistency in the FFDI data. In an attempt to improve these data, the mean bias method was applied to the Sydney wind-speed data, and the resulting wind speeds used to recalculate the FFDI. The original and modified data are shown in Figures 7.15 and 7.16, together with the corresponding IMFs. The original wind-speed data show a step-function change after 1990 which causes a very sharp non-stationarity in the input data; in Chapter 4 it is shown that the EEMD method is highly sensitive to abrupt changes in the data.

Both sets of wind-speed data (original and modified) were then used in the SEEMD method with 300 ensemble members and smoothing parameter $p = 0.01$. The decadal and multi-decadal modes, and the trend are shown in Figure 7.15. The decadal modes of the original and modified data are out of phase in the middle of the dataset where the inconsistency in the data occurs. The differences between the two multi-decadal modes are more obvious in the middle of the dataset. These differences start from 1985 and continue until 2005.

The other obvious differences between the original and modified wind data appear in the trend (bottom panel in Figure 7.15). The original wind speed shows an overall long-term increase whereas the modified data show an overall long-term decrease. Troccoli et al. [2012] showed that in Sydney the long-term surface (2 m) and upper-layer (10 m) wind-speed behaviours are different; the surface wind speed has decreased over the long term whereas the upper-layer wind speed has increased. In our study, modifying the wind data using mean bias produced inconsistent results in the long-term wind behaviour. As shown in Figure 7.15, the modified wind-speed

results (solid line) at 10 m show a decrease in wind speed. The original wind-speed trend however, agrees with the results of Troccoli et al. [2012].

The wind results in Figure 7.15 show that modifying the input data can change the results dramatically in the sense that they might not show the true variability of the data at different time scales. This is more evidence for the sensitivity of the EMD method to abrupt changes in the data and its significant impact on the results.

Figure 7.16 shows the FFDI calculated from the original wind-speed data and the modified wind-speed data. In this case, the FFDI has changed only slightly. Lucas [2010] showed that the change in FFDI due to a change in the wind is:

$$\Delta\text{FFDI} = 0.0234 \times \text{FFDI} \times \Delta V, \tag{7.3}$$

where V is wind speed. This means for a change in wind speed from 27 km per hour to 20 km per hour with an FFDI of about 7, the change in the FFDI is only about 1.2, which is not a large change. However, even this small change in the data had significant impact on the multi-decadal results of EMD (Figure 7.16), particularly on the lower-frequency IMFs. The dependence of each step in the method on the previous step magnifies even small errors, and these magnified errors become more obvious in the later, lower-frequency IMFs (Figure 7.16).

The decadal FFDI original and modified IMFs are quite similar except around 1997. The original FFDI multi-decadal IMF has a smaller amplitude than the modified FFDI IMF from 1972 to 1995, but a more pronounced peak in the period 2000 to 2005. The trend in the original FFDI data shows the FFDI accelerated from 1987, whereas the modified FFDI trend accelerated from 1995, and at a greater rate.

7.6 Discussion and conclusions

In this chapter, SEEMD was used to analyse the long-term variability of FFDI and its contributors. This was quite challenging. Apart from the normal nonlinearity and non-stationarity in the meteorological data, there was a considerable inconsistency in wind data. For example, the Sydney wind-speed data showed a large and sudden inconsistency, making the data highly non-stationary. The high sensitivity of the EMD method to abrupt changes in the data and how it strongly impacts the results (especially for low-frequency modes).

However, there were some valuable results that achieved in this chapter. It was shown that the main driver of the FFDI is different at different time scales. On a decadal time scale, the FFDI is mainly driven by temperature, probably due to the decadal variability in ENSO. On a multi-decadal time scale, however, wind

influenced the FFDI stronger than other contributors. In the long-term time scale, FFDI seems to be strongly driven by temperature, relative humidity and the drought factor. These results can change depending on the geographical situation of the weather station. In all time scales, relative humidity showed a strong relationship with FFDI of all three stations.

The long-term FFDI trend showed that the FFDI started to accelerate after 1990 due to accelerations in the drought factor, temperature and wind speed, and a deceleration in relative humidity. This acceleration in the FFDI predicts an increase the in frequency and intensity of bushfire occurrence.

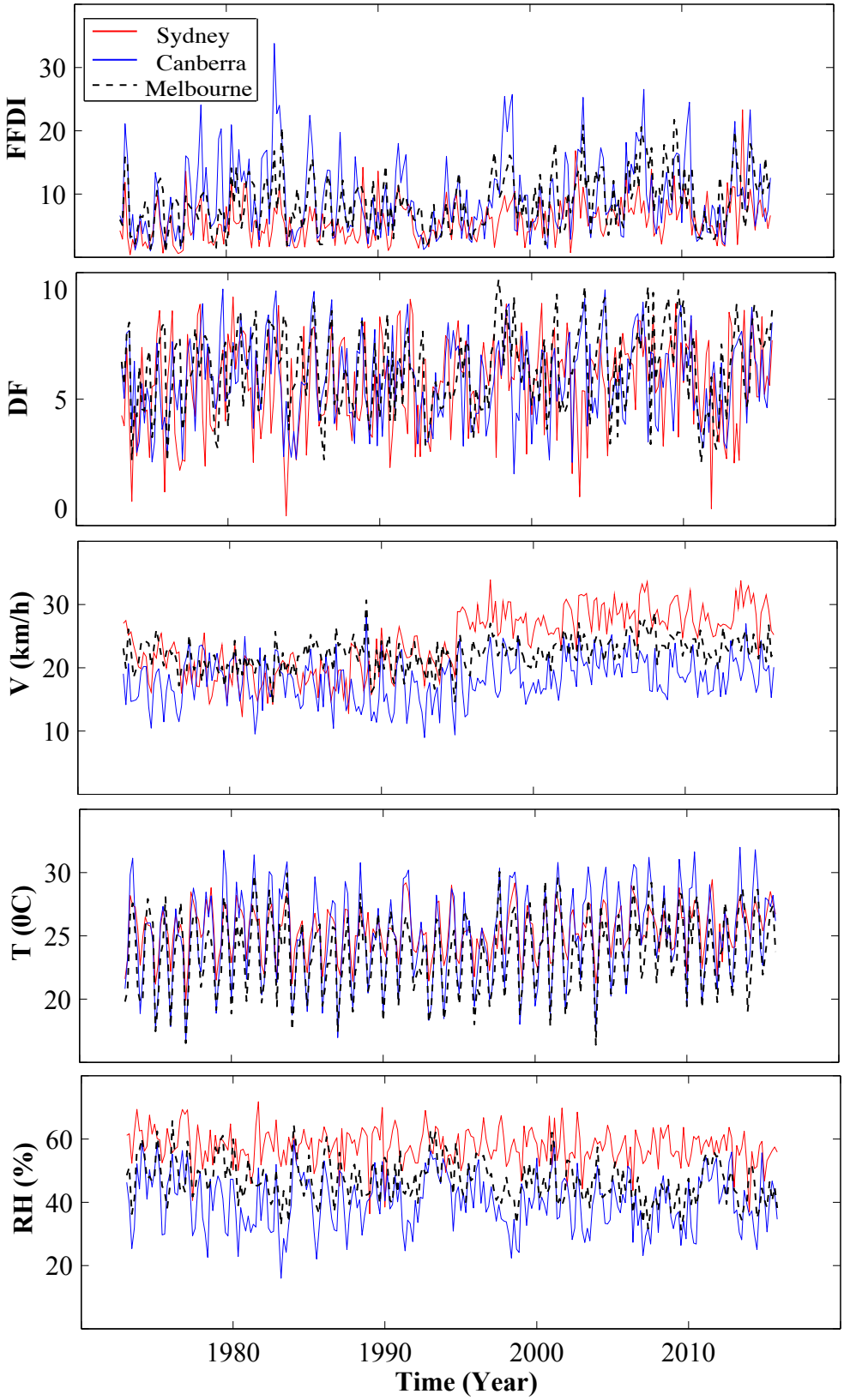

Figure 7.1: Forest Fire Danger Index (FFDI), drought factor (DF), wind speed (V), maximum temperature (T) and relative humidity (RH) data, all during the fire season, for Sydney, Canberra and Melbourne.

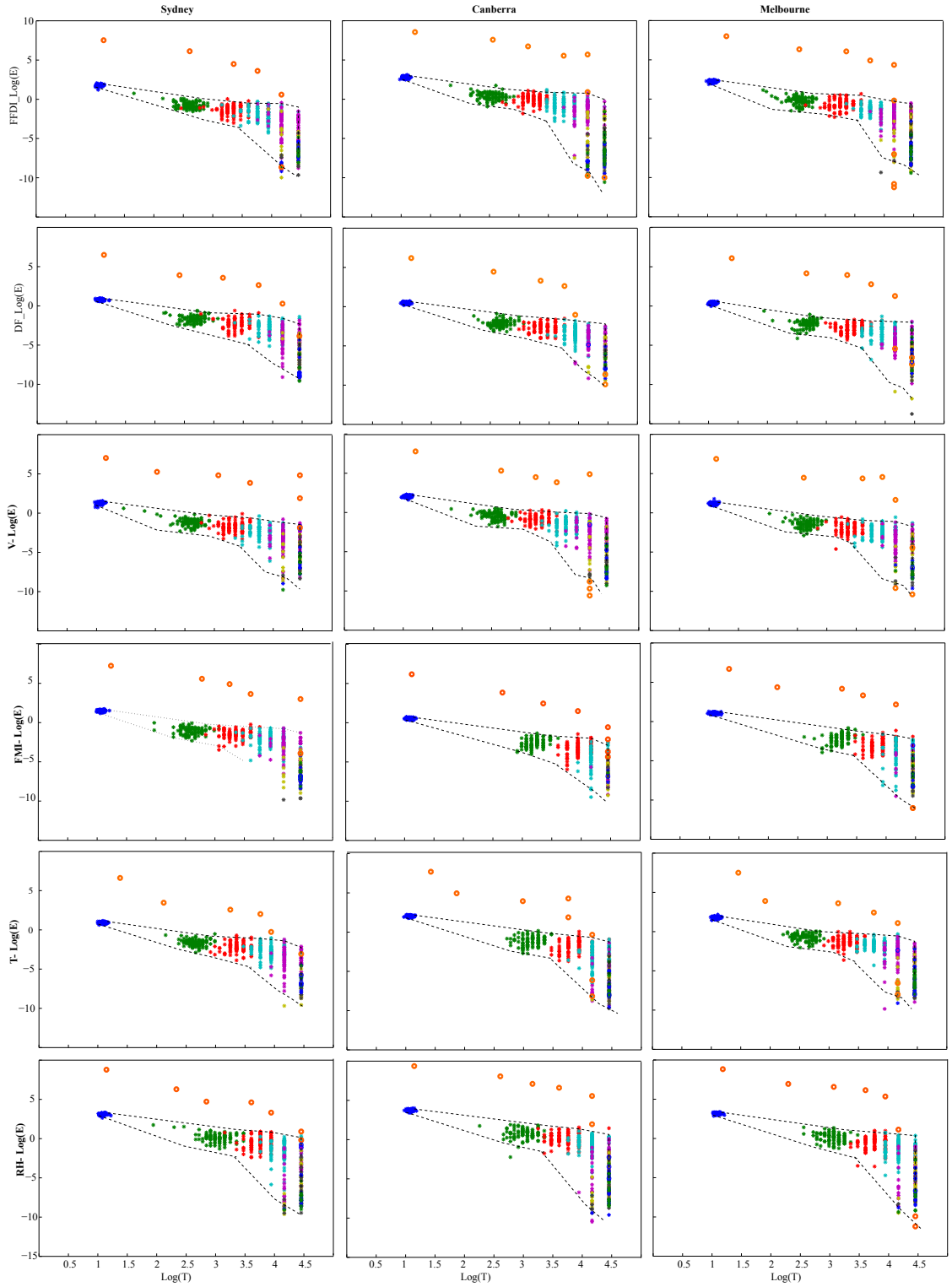

Figure 7.2: Log energy versus log period from the Monte Carlo tests (Chapter 5) for the datasets in Figure 7.1. The statistically significant IMFs are the open circles at the top of each plot.

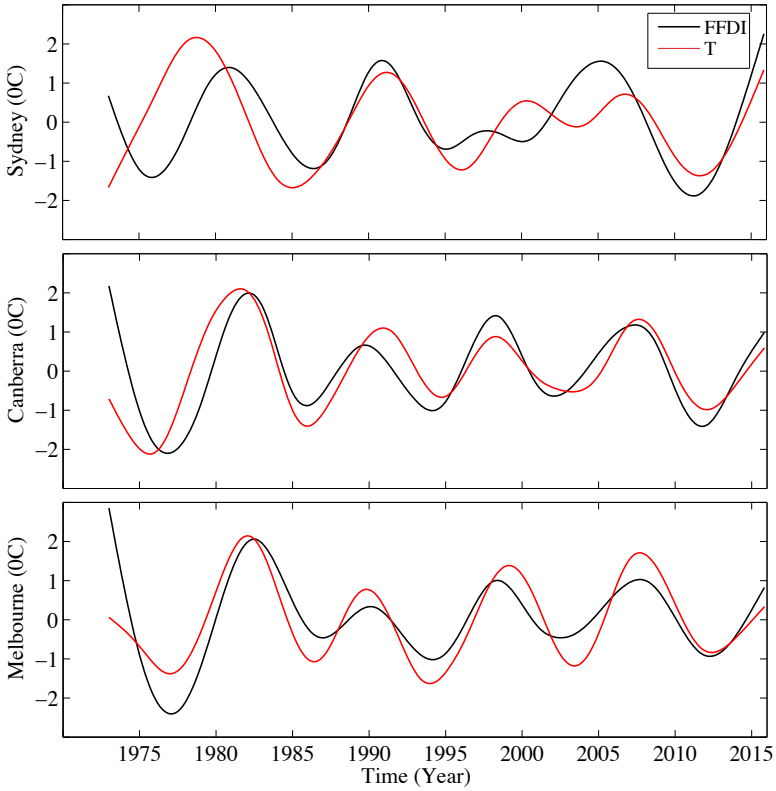

Figure 7.3: Temperature and FFDI decadal variability, IMF 4.

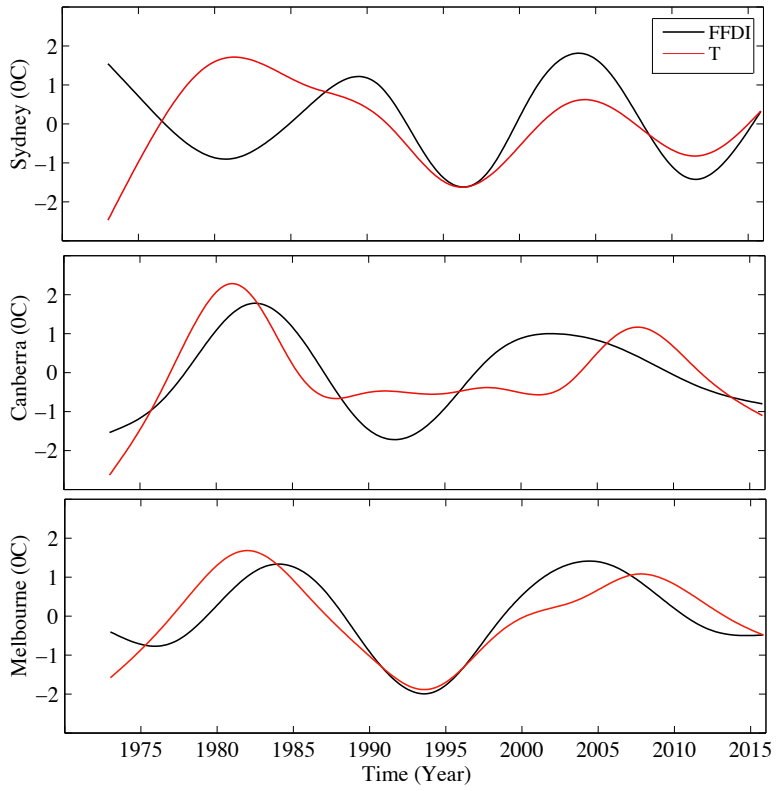

Figure 7.4: Temperature and FFDI multi-decadal variability, IMF 5.

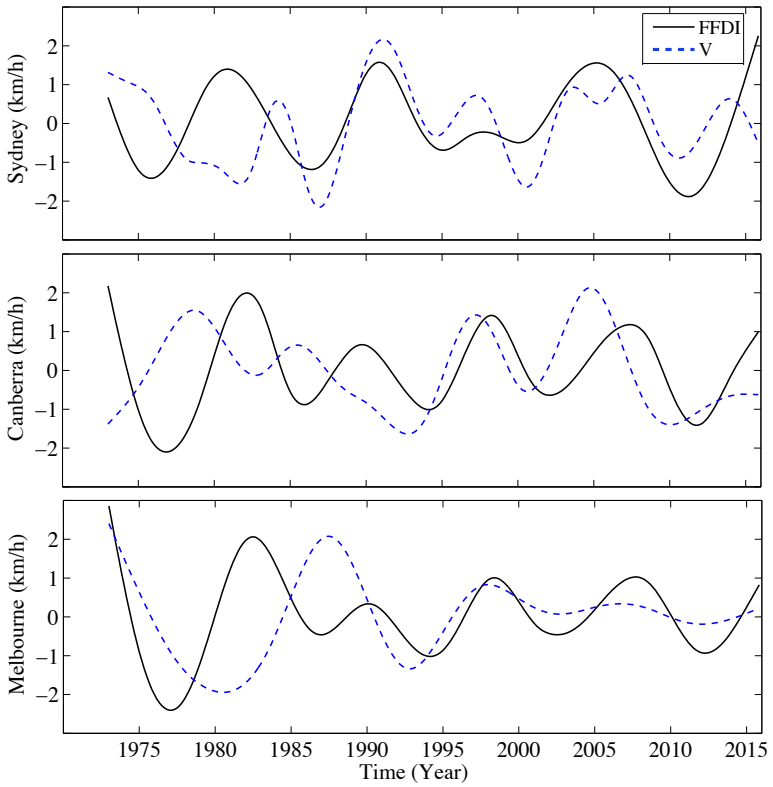

Figure 7.5: Wind and FFDI decadal variability, IMF 4.

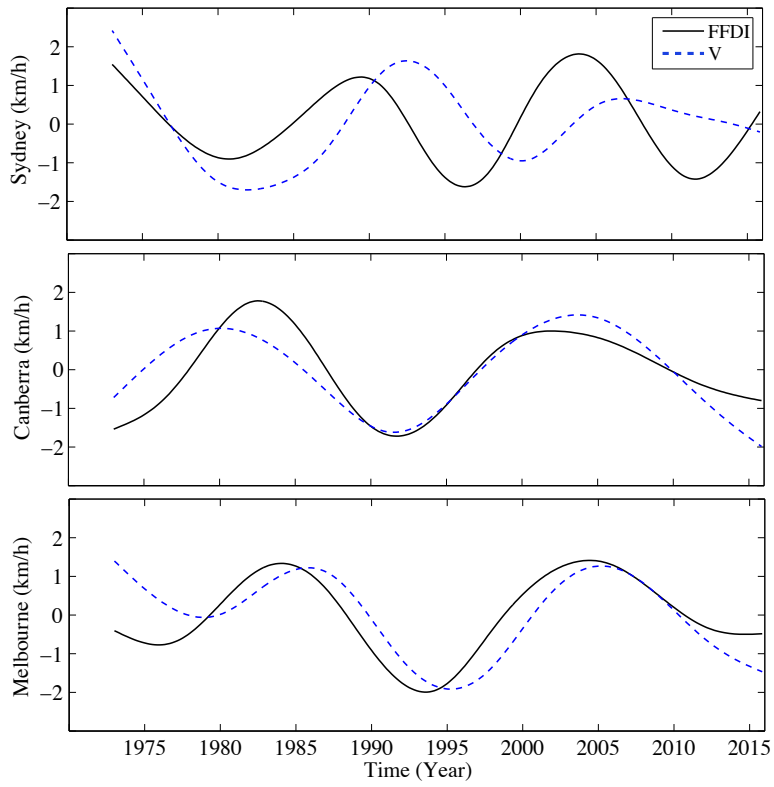

Figure 7.6: Wind and FFDI multi-decadal variability, IMF 5.

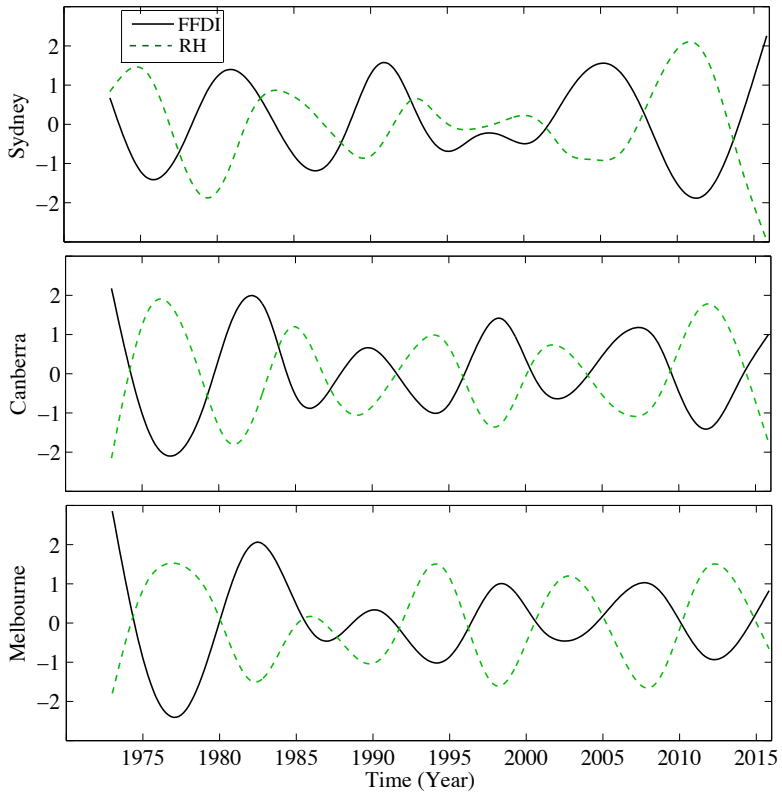

Figure 7.7: Relative humidity and FFDI decadal variability, IMF 4.

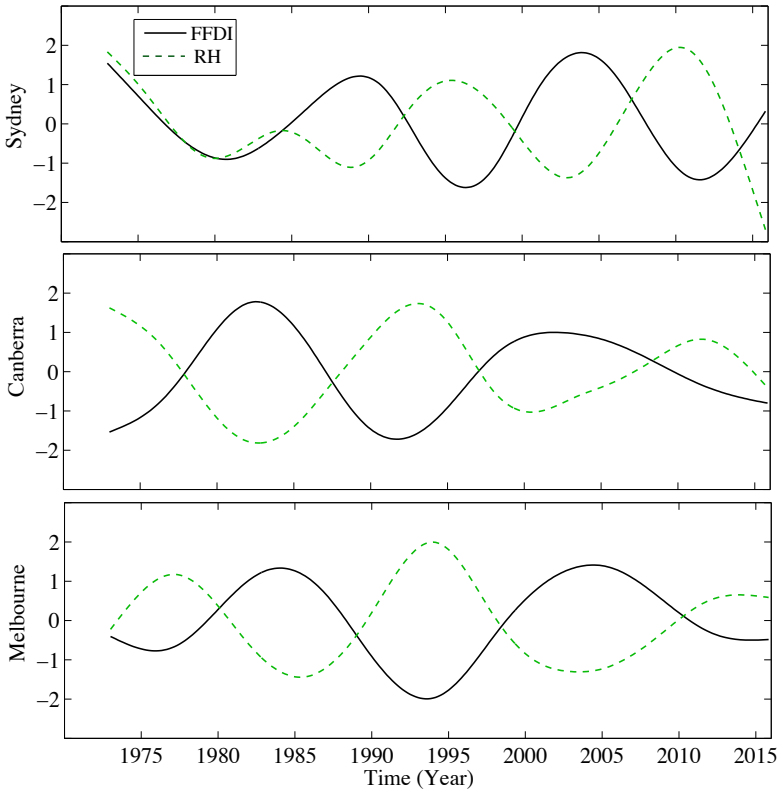

Figure 7.8: Relative humidity and FFDI multi-decadal variability IMF 5.

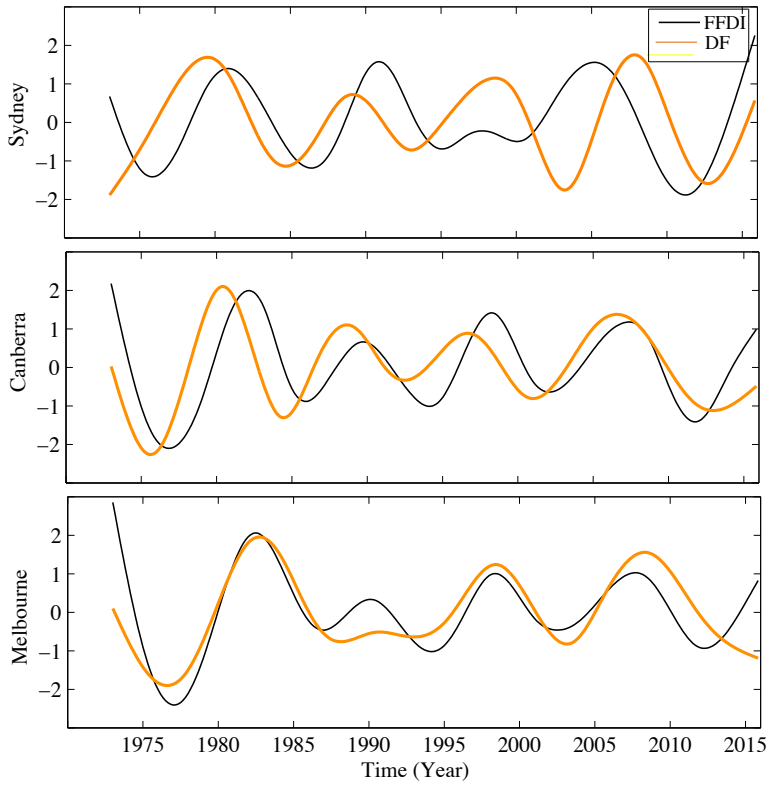

Figure 7.9: Drought factor and FFDI decadal variability, IMF 4.

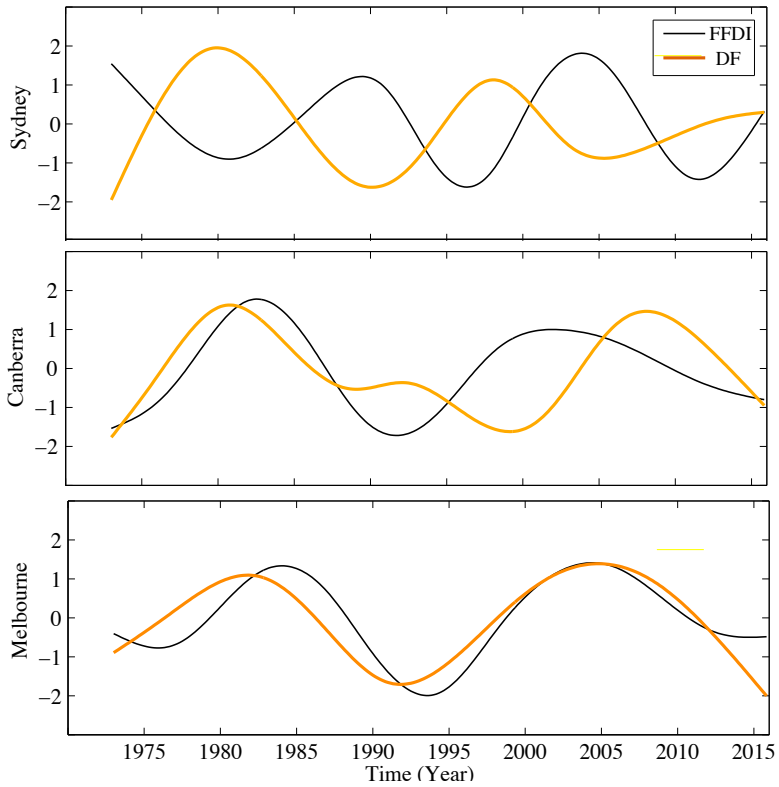

Figure 7.10: Drought factor and FFDI multi-decadal variability, IMF 5.

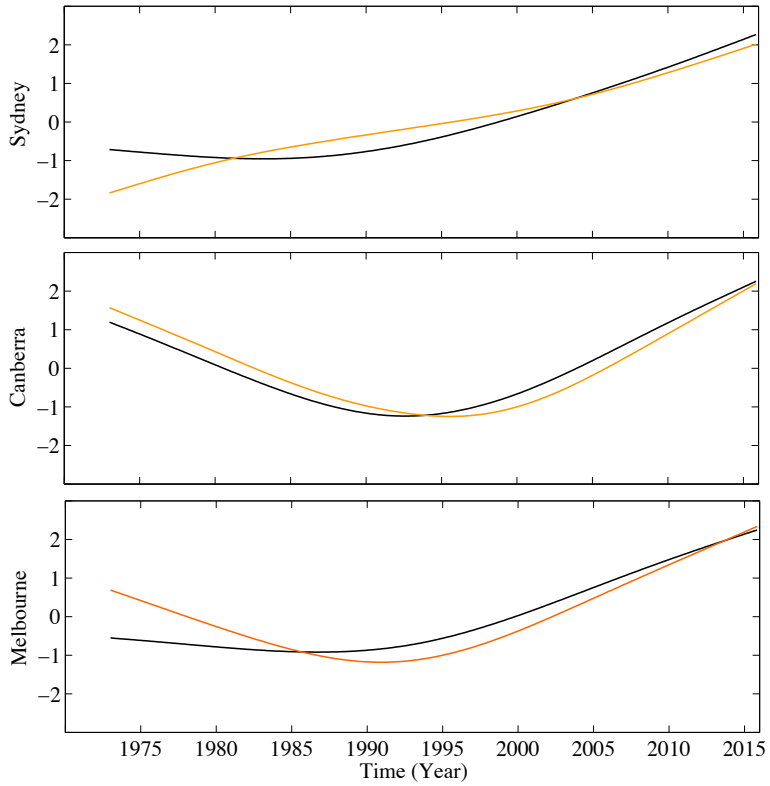

Figure 7.11: Long-term variability (trend) for FFDI (black) and drought factor (orange).

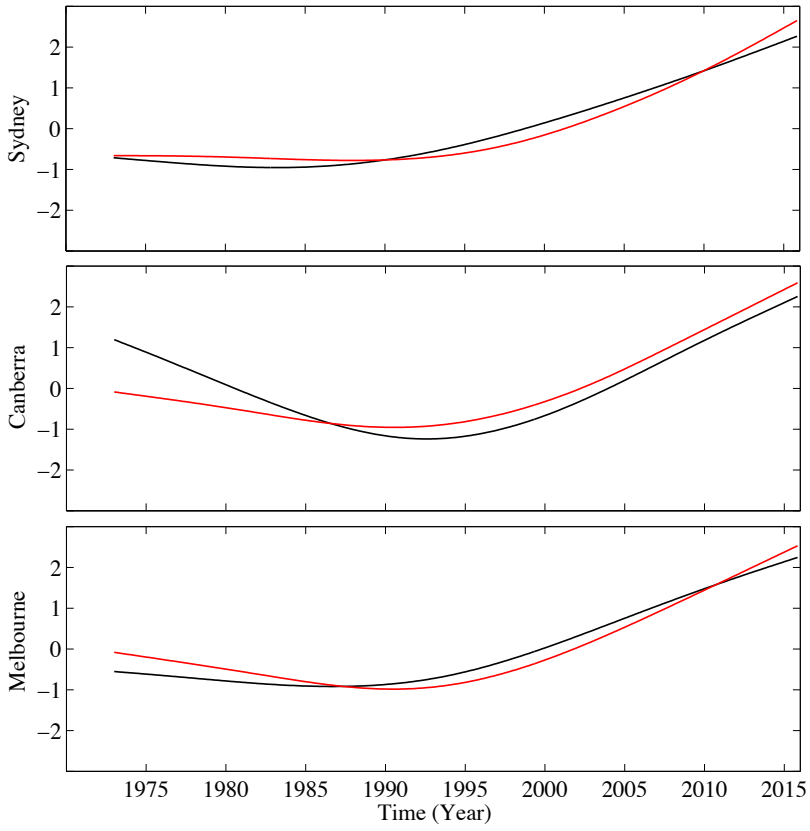

Figure 7.12: Long-term variability (trend) for FFDI (black) and temperature (red).

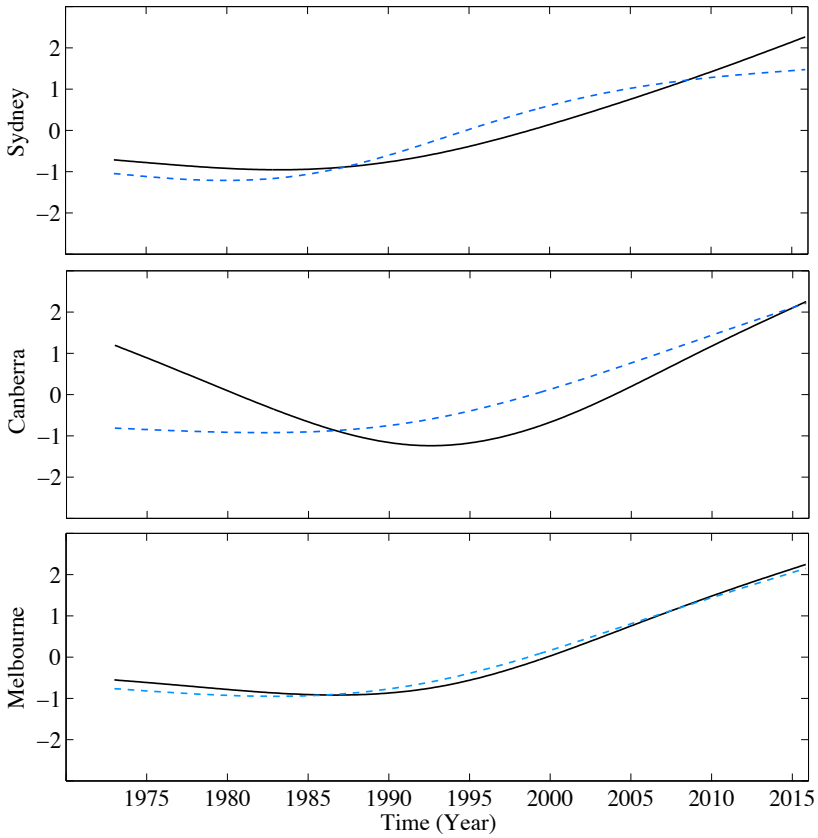

Figure 7.13: Long-term variability (trend) for FFDI (black) and wind speed (blue dashed line).

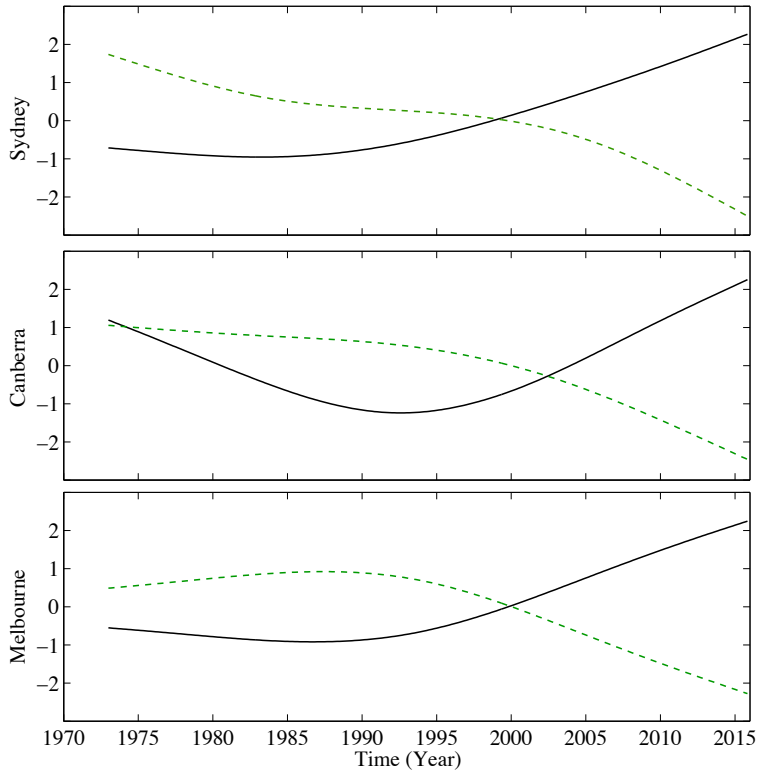

Figure 7.14: Long-term variability (trend) for FFDI (black) and relative humidity (green dashed line).

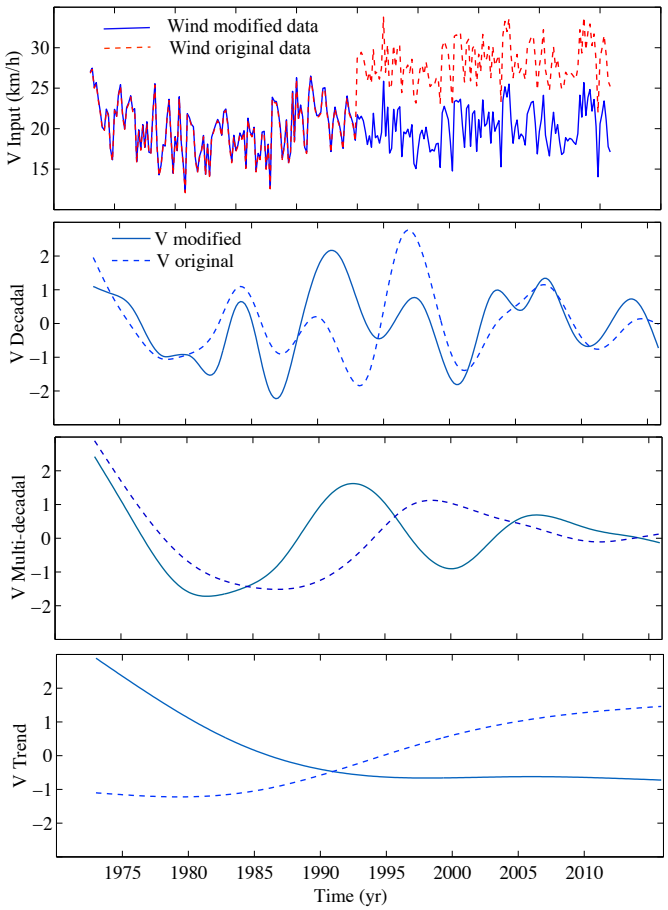

Figure 7.15: Wind-speed original and modified data (top panel) with the corresponding decadal and multi-decadal IMFs, and the trend.

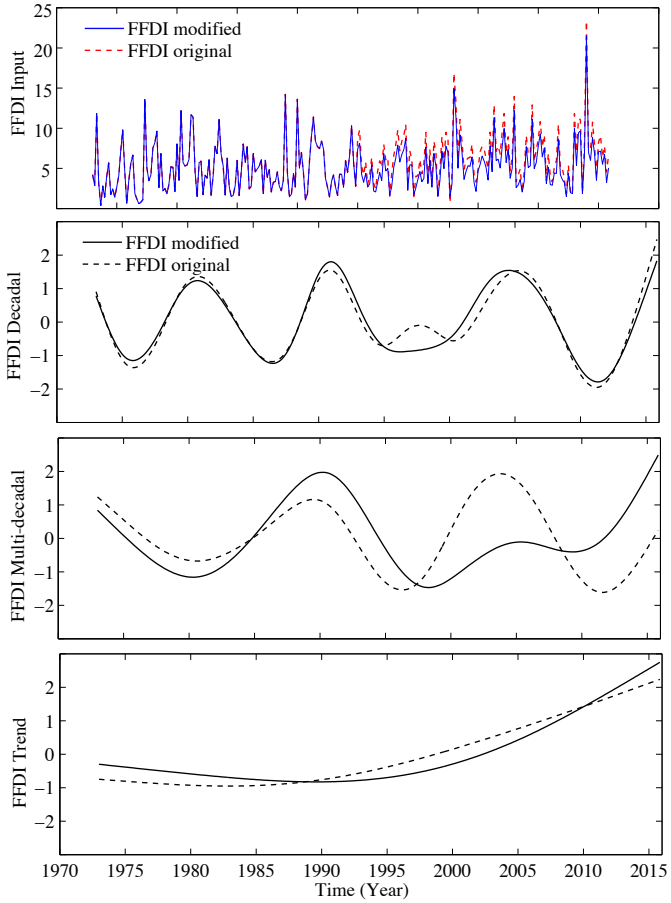

Figure 7.16: FFDI original and modified data (top panel) with the corresponding decadal and multi-decadal IMFs, and the trend.

Chapter 8

Conclusions and Future Work

This chapter presents a summary of this dissertation and its conclusions. In this dissertation, the focus was on the application and sensitivity of the Empirical Mode Decomposition (EMD) method, using three different interpolation methods, in decomposing a comprehensive selection of non-stationary data with the aim of improving the accuracy of this method and enhancing our ability to analyse environmental data.

8.1 EMD/EEMD sensitivity to data non-stationarity

Twenty different stationary and non-stationary synthetic datasets were used to determine the sensitivity of EMD (and the ensemble EMD (EEMD) used for noisy data) to non-stationarity in the data. In most cases, the method was able to find the correct frequency for the IMFs (modes), which is evidence for how powerful the method is in decomposing a complex (in the general sense) time series. However, the performance of EMD was strongly sensitive to abrupt changes in the mean, for example the 'hockey-stick' logarithmic trend, and the rise and fall of the Gaussian trend. This sensitivity was more obvious in the lower-frequency IMFs and in the residual (trend). Finding the correct shape of the residual is important for analysing long-term changes in the data. This can be critically important to the accurate study and interpretation of environmental data [Ezer et al., 2013].

8.2 EMD/EEMD sensitivity to the interpolation method

Three different interpolation methods were used: a simple piecewise-linear interpolation; a cubic-spline interpolation, used in the original EMD method; and a more sophisticated smoothing-spline interpolation. The results showed that, in the absence of noise, the EMD method was not sensitive to the interpolation method, but EEMD applied to noisy data was.

EEMD with linear interpolation suffered strongly from mode mixing and end effects in most cases, and consistently underestimated the amplitude of the sinusoidal signal components. Linear interpolation did however provide reasonable estimates the trend components (except for data with logarithmic and Gaussian trends).

EEMD with cubic interpolation provided better results than linear EEMD in the presence of noise, but still suffered from mode mixing and end effects. EEMD with smoothing-spline interpolation consistently provided the most accurate and robust results for the low-frequency IMFs. Because the smoothing parameter was set to favour of the low-frequency IMFs, the amplitude of the high-frequency IMFs in most cases was underestimated.

EEMD with both linear and cubic-spline interpolation showed more error in estimating the lower-frequency IMFs than smoothing-spline interpolation which was consistently more accurate and robust in obtaining the lower-frequency IMFs. The smoothing-spline EEMD results also showed considerably less sensitivity to noise, the endpoint issue and the nature of the data than EEMD using other two interpolation methods. Overall, the results support the use of smoothing-spline (SEEMD) interpolation in favour of the other interpolation methods considered.

8.3 EEMD implications for environmental time series

8.3.1 Temperature

In Chapter 5, the EEMD method with cubic-spline (CEEMD) and smoothing-spline (SEEMD) interpolations was used on temperature data. Regardless of the interpolation method, EEMD obtained good results overall for the temperature data. However, changing the interpolation method changed the degree of variability in the trend, and even the interpretation of the data. In addition, SEEMD found fewer modes, with less mode mixing, than CEEMD.

The results from SEEMD showed detail in the temperature variability that agreed with the literature. For example, SEEMD showed details in the decadal variability in the global temperature data, that were consistent with extreme events and El Niño effects reported in the literature.

The two interpolation methods (cubic spline and smoothing spline) both gave similar results for Sydney maximum temperature. For the Sydney data, the shape of the trend for the minimum temperature from SEEMD was quite different from that from CEEMD, although the increase in temperature was similar. The CEEMD trend for minimum temperature had a positive constant slope, whereas the SEEMD trend

showed a minimum temperature increasing at an increasing rate since 1920; the SEEMD result agreed with the literature.

8.3.2 Sea level

In Chapter 6, the implications of using SEEMD and CEEMD on sea-level data were investigated. The SEEMD results were again more resistant to mode mixing than the CEEMD results.

For the intra-decadal, decadal, and multi-decadal variability with periods more than 30 years, SEEMD and CEEMD obtained similar results. The sea-level trends obtained from the Newcastle data from SEEMD and CEEMD were similar, and agreed with the literature. For the Sydney trends, however, the results from the two interpolation methods disagreed at the endpoints of the data, giving different estimates of the sea-level rise. The SEEMD results agreed with the literature.

For Port Pirie, the SEEMD and CEEMD sea-level trends disagreed in both the shape and amount of sea-level rise. SEEMD gave a nonlinear sea-level rise, with the rate of rise increasing rapidly after 1990; CEEMD gave a uniform rise in sea level. Again the SEEMD results agreed with the literature. SEEMD and CEEMD gave different results in estimating sea-level acceleration, either in the amount of acceleration or the shape of the acceleration trend, for all three stations. SEEMD gave a result that could explain the relationship between the East Australian Current and sea level in Sydney, which also agreed with the literature.

8.3.3 Forest Fire Danger Index (FFDI)

In Chapter 7, SEEMD was used to analyse the long-term variability in the FFDI and its contributors. The SEEMD trend in FFDI and its contributors either had a minimum in about 1990 or started increasing then (except relative humidity which had a maximum in 1990). However the short length of the dataset means the accuracy of these trends is questionable.

At decadal time scales, the FFDI was mainly driven by temperature. This is likely to be due to the decadal variability in ENSO. The multi-decadal variability in FFDI showed a strong correlation with wind speed. In the long-term variability, both temperature and the drought factor had a strong influence on the FFDI, predicting hotter and drier conditions in the future, thus increasing the possibility of bushfires.

Sydney wind-speed data showed an abrupt inconsistency in 1990. This made the data highly non-stationary, and therefore affected the results of EMD. Improving the wind-speed data using the bias-mean method did not improve the outcome. Even

small changes in the FFDI data changed the low-frequency IMFs, and therefore our interpretation of FFDI behaviour at a multi-decadal scale.

8.4 Significance of this study and possible future studies

This study has shown that the use of smoothing-spline interpolation, rather than the usual cubic-spline interpolation, can significantly improve the accuracy of the EMD results. Using SEEMD reduced the sensitivity of the method to noise and also reduced mode mixing. For the data used in this study, the SEEMD results were more resistant to endpoint effects. The SEEMD results were also more robust regardless of the nature of the data. This robustness is evidence that, compared to EEMD using the other two interpolation methods, SEEMD was significantly less sensitive to data non-stationarity. All of these results were also confirmed when SEEMD and CEEMD were used on the temperature and sea-level data. SEEMD results consistently gave more physically meaningful detail than the original CEEMD.

However, despite all these advantages, SEEMD has its limitations, an important one of which is finding a suitable smoothing parameter to analyse the data. For the synthetic data used here, it was possible to find a suitable smoothing parameter that gave accurate results for all frequencies. There is a need for further study to determine whether a specific range of smoothing parameters can be found to give accurate results for all frequencies regardless of the nature of the input data. For example, using generalised cross validation may be a good method to find a suitable smoothing parameter. There are of course a number of other interpolation methods that could be used instead of the linear, cubic-spline or smoothing-spline interpolation methods.

Appendix A

All IMFs

In this Chapter, the EMD/EEMD results for all the 20 synthetic datasets are presented. Figures A.1 and A.2 are the EMD/EEMD results for stationary time series with and without noise. Figures A.3 to A.6 illustrate the IMFs for non-stationary data with time-dependent mean without noise. Figures A.7 and A.8 shows the IMFs for non-stationary data with time-dependent variance. Figures A.9 to A.12 are the IMFs for non-stationary data with time-dependent mean with constant noise and Figures A.13 to A.20 are the IMFs for non-stationary data with time-dependent mean and time-dependent variance.

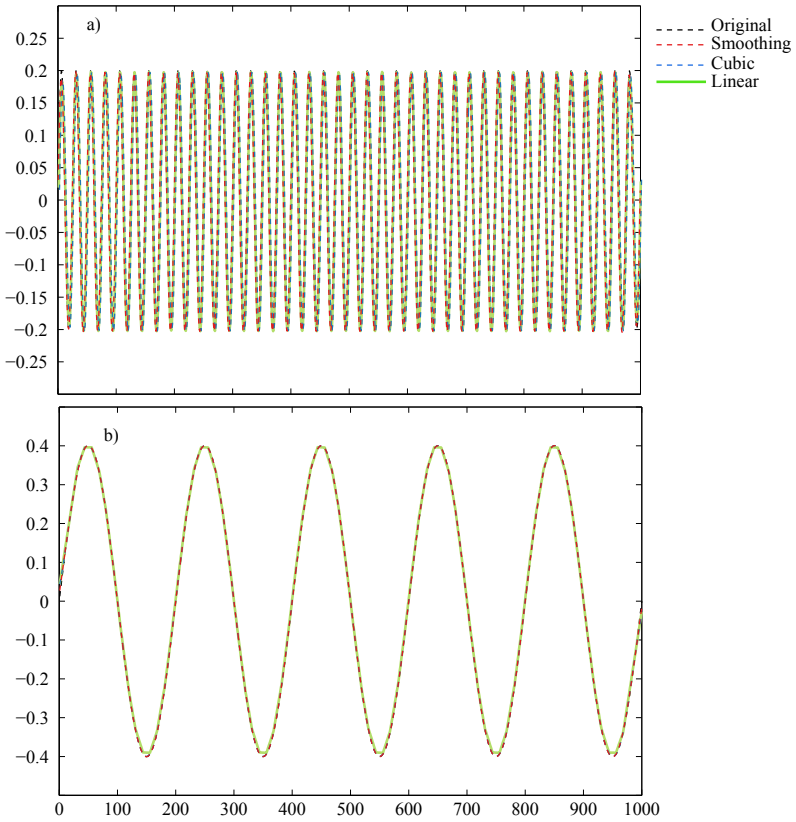

Figure A.1: EMD results for the noiseless stationary data x_0 (Equation 4.1, Chapter 4).

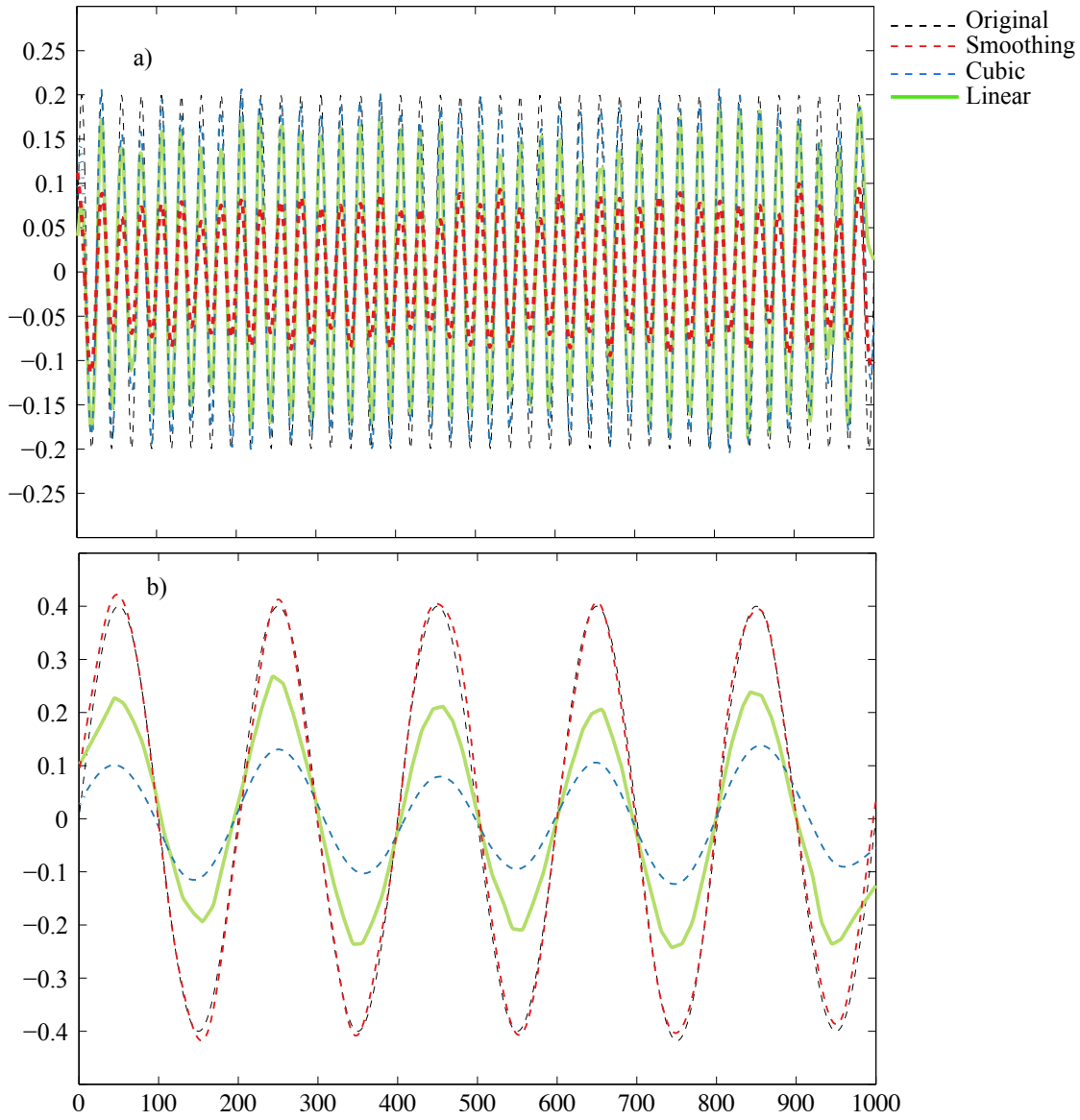

Figure A.2: EEMD results for the corresponding data with noise y_0 (Equation 4.3, Chapter 4).

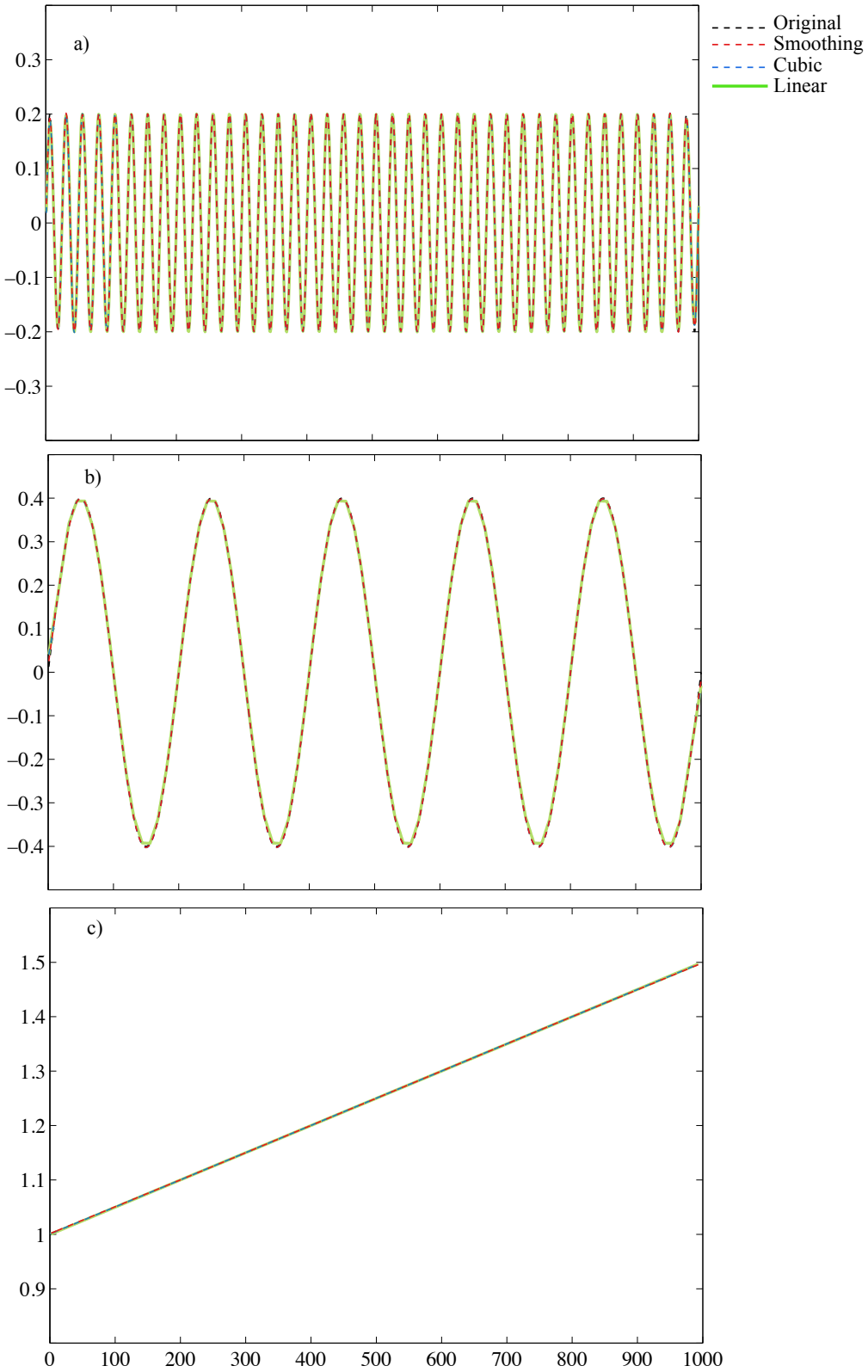

Figure A.3: EMD results for the non-stationary data with linear time-dependent mean, x_1 (Equation 4.4, Chapter 4).

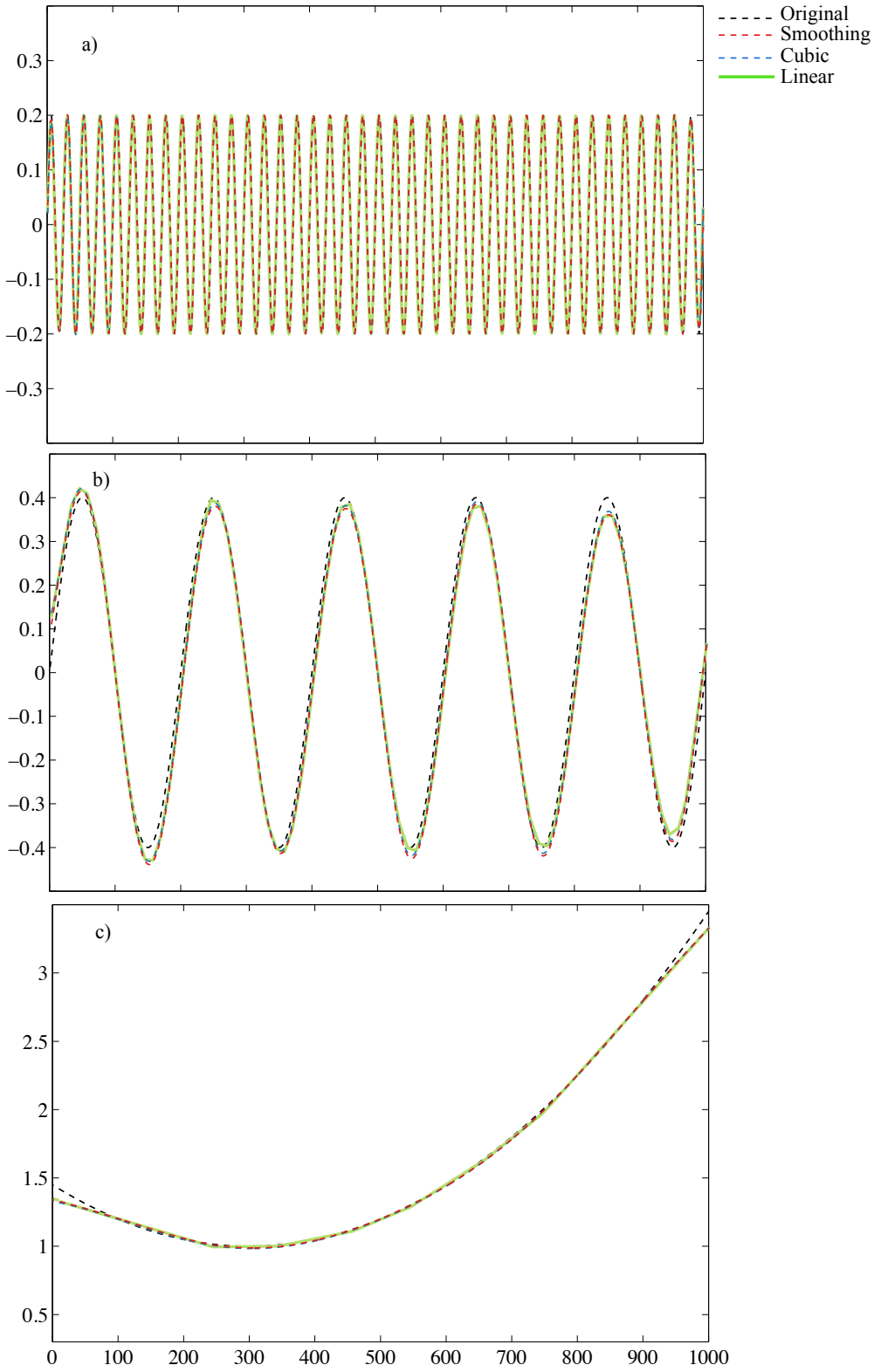

Figure A.4: EMD results for the non-stationary data with non-monotonic time-dependent mean, x_2 (Equation 4.4, Chapter 4).

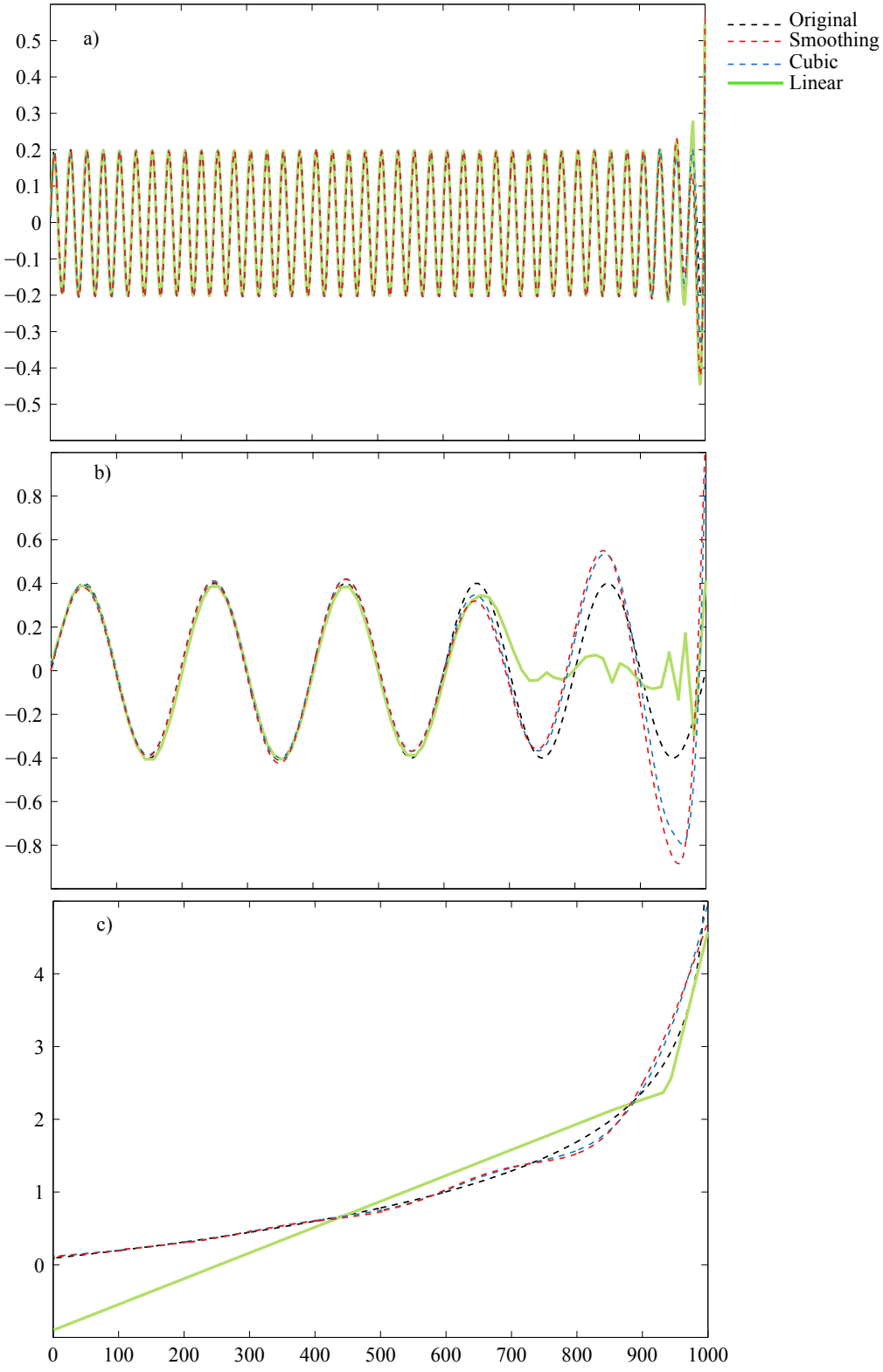

Figure A.5: EMD results for the non-stationary data with logarithmic time-dependent mean, x_3 (Equation 4.4, Chapter 4).

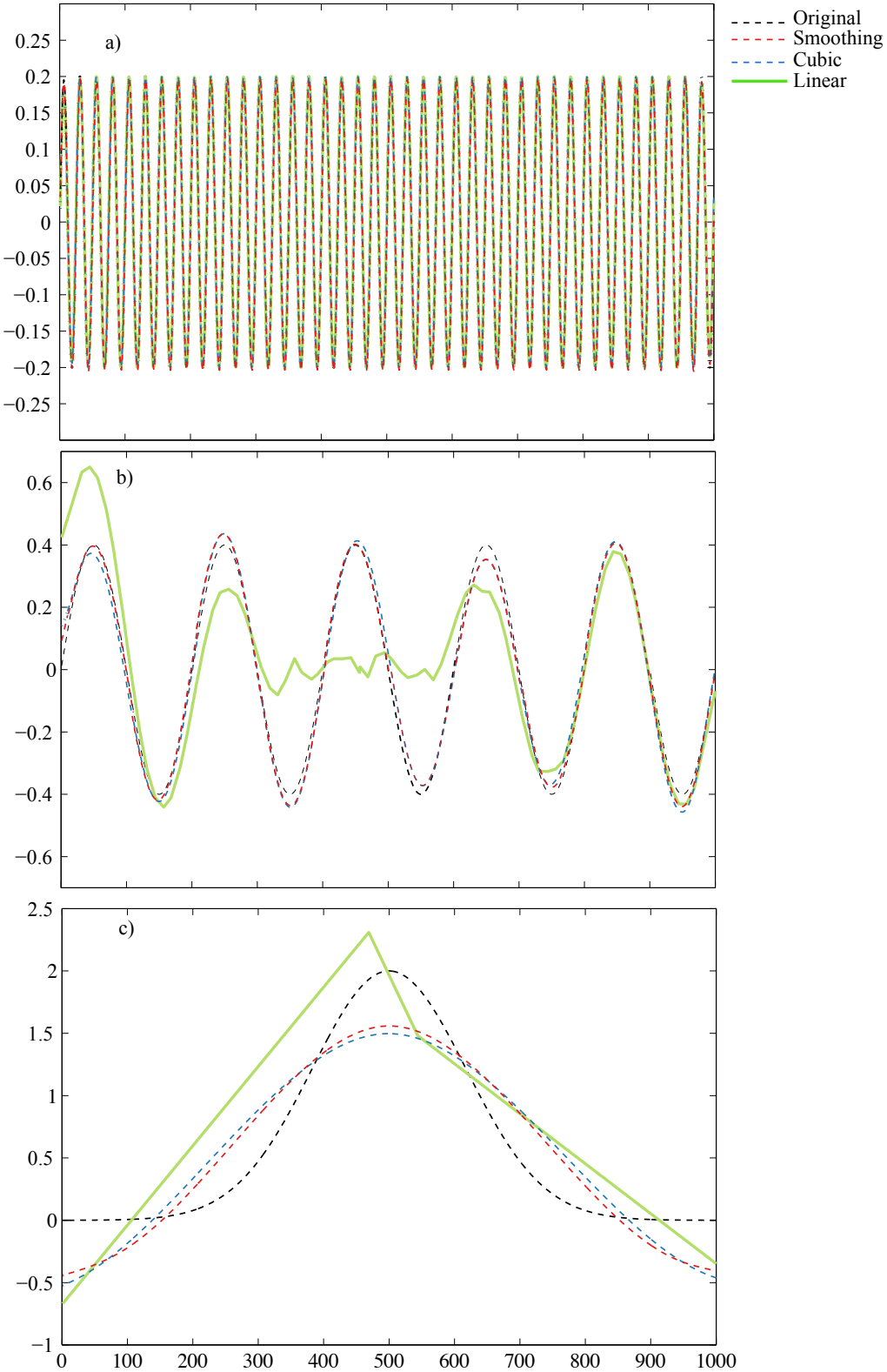

Figure A.6: EMD results for the non-stationary data with Gaussian time-dependent mean, x_4 (Equation 4.4, Chapter 4).

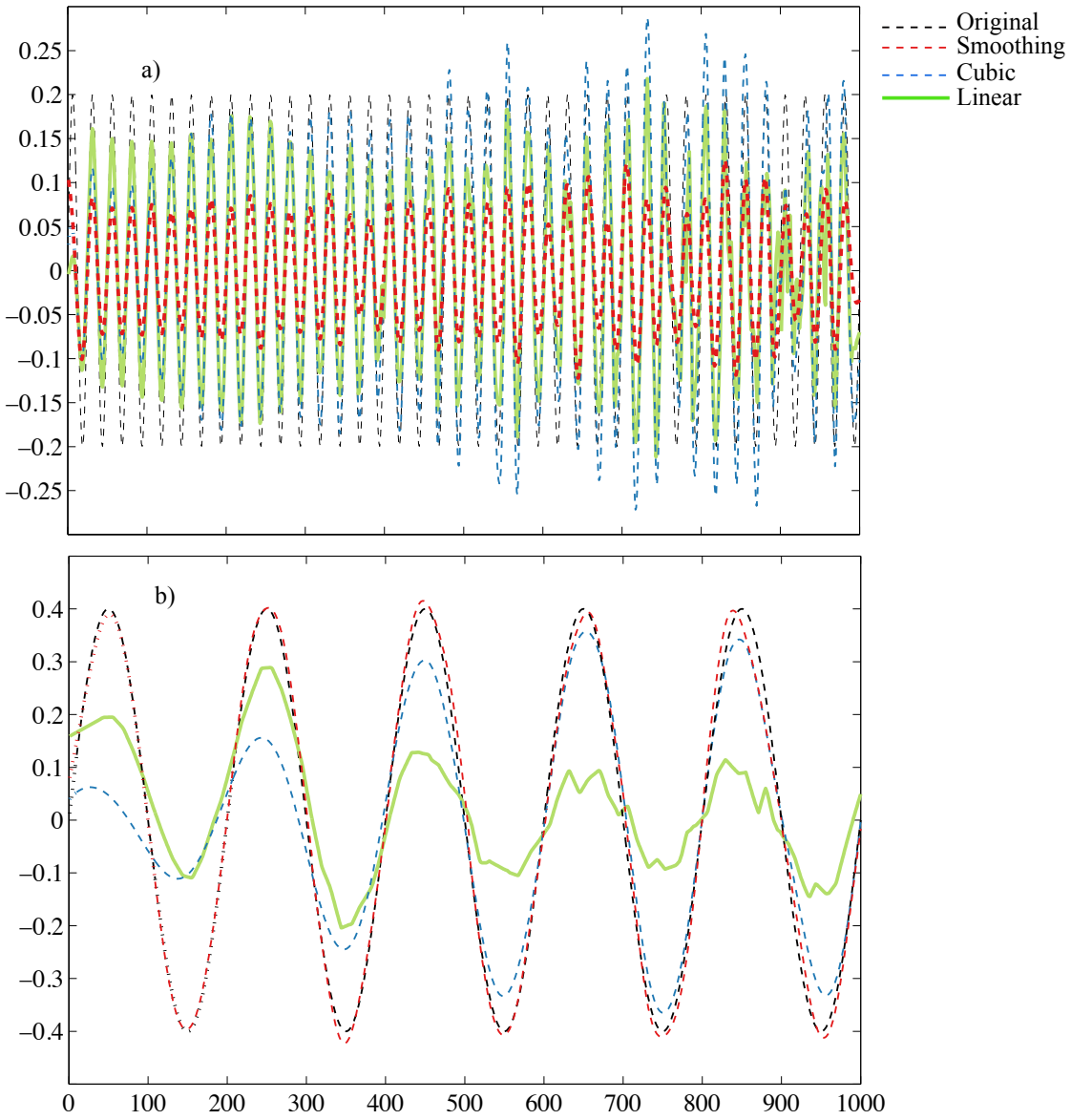

Figure A.7: EEMD results for the non-stationary data with time-dependent variance, z_{05} (Equation 4.10, Chapter 4).

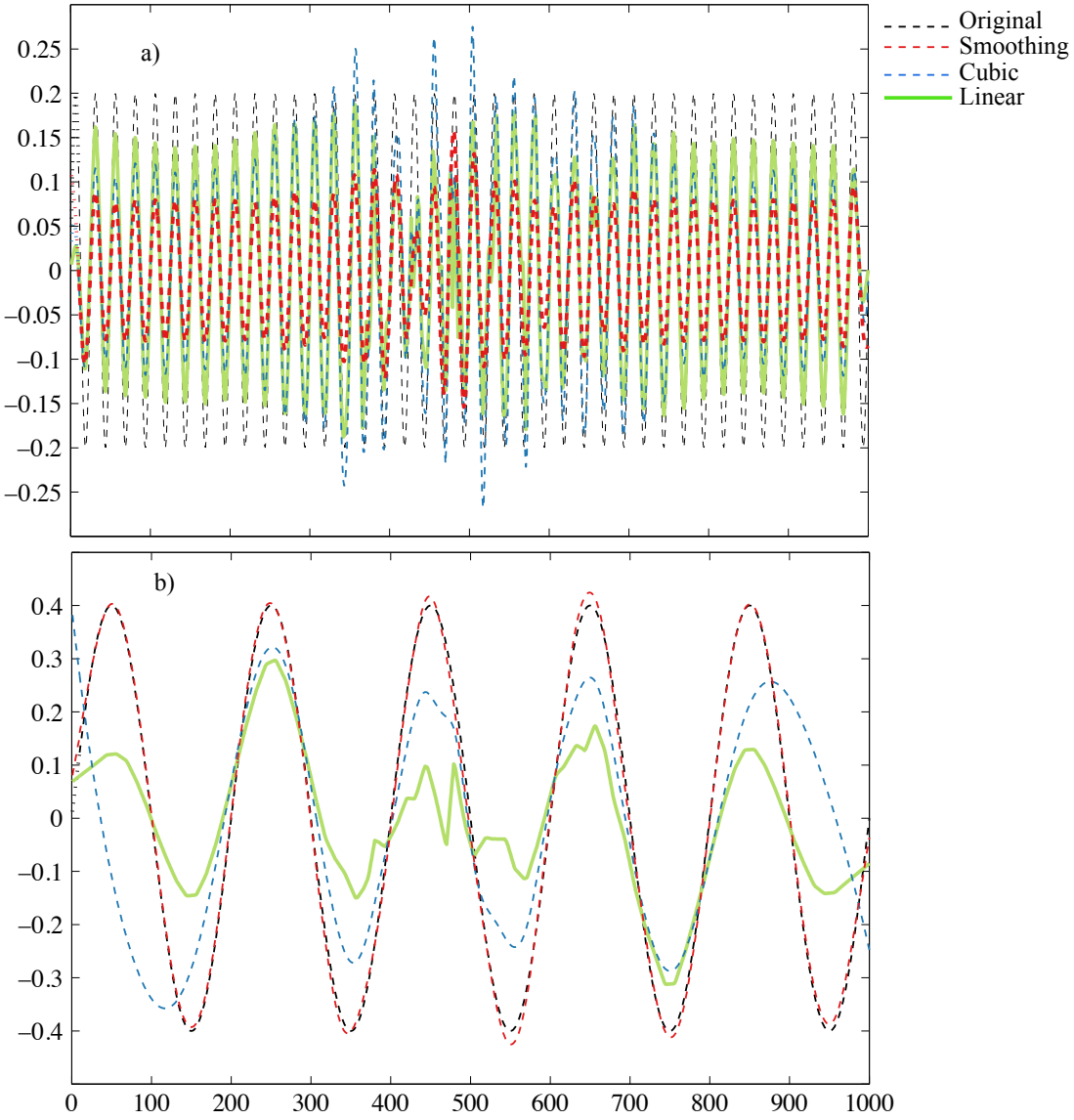

Figure A.8: EEMD results for the non-stationary data with time-dependent variance, z_{06} (Equation 4.10, Chapter 4).

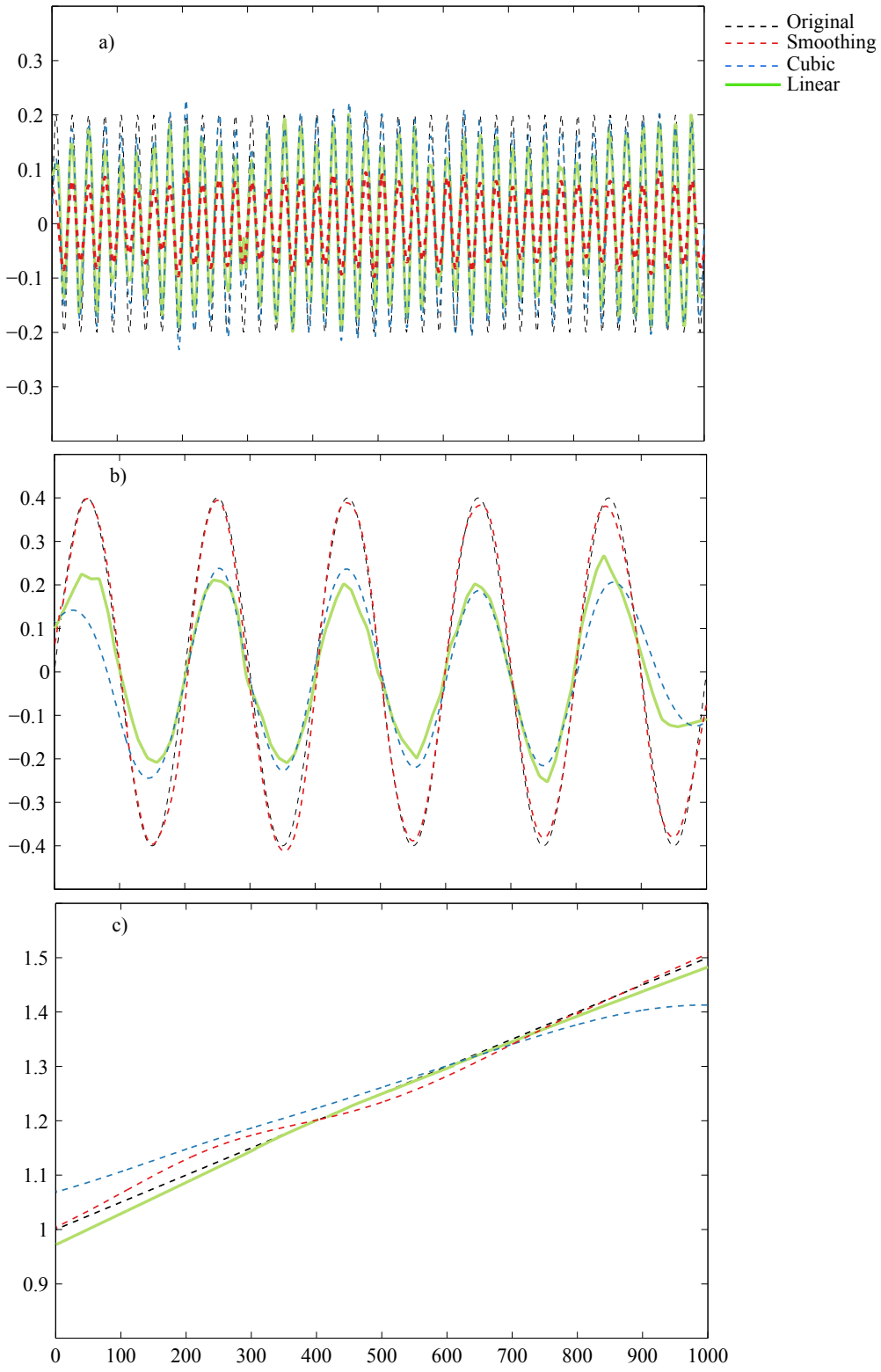

Figure A.9: EEMD results for the non-stationary data with time-dependent mean and constant noise, y_1 (Equation 4.9, Chapter 4).

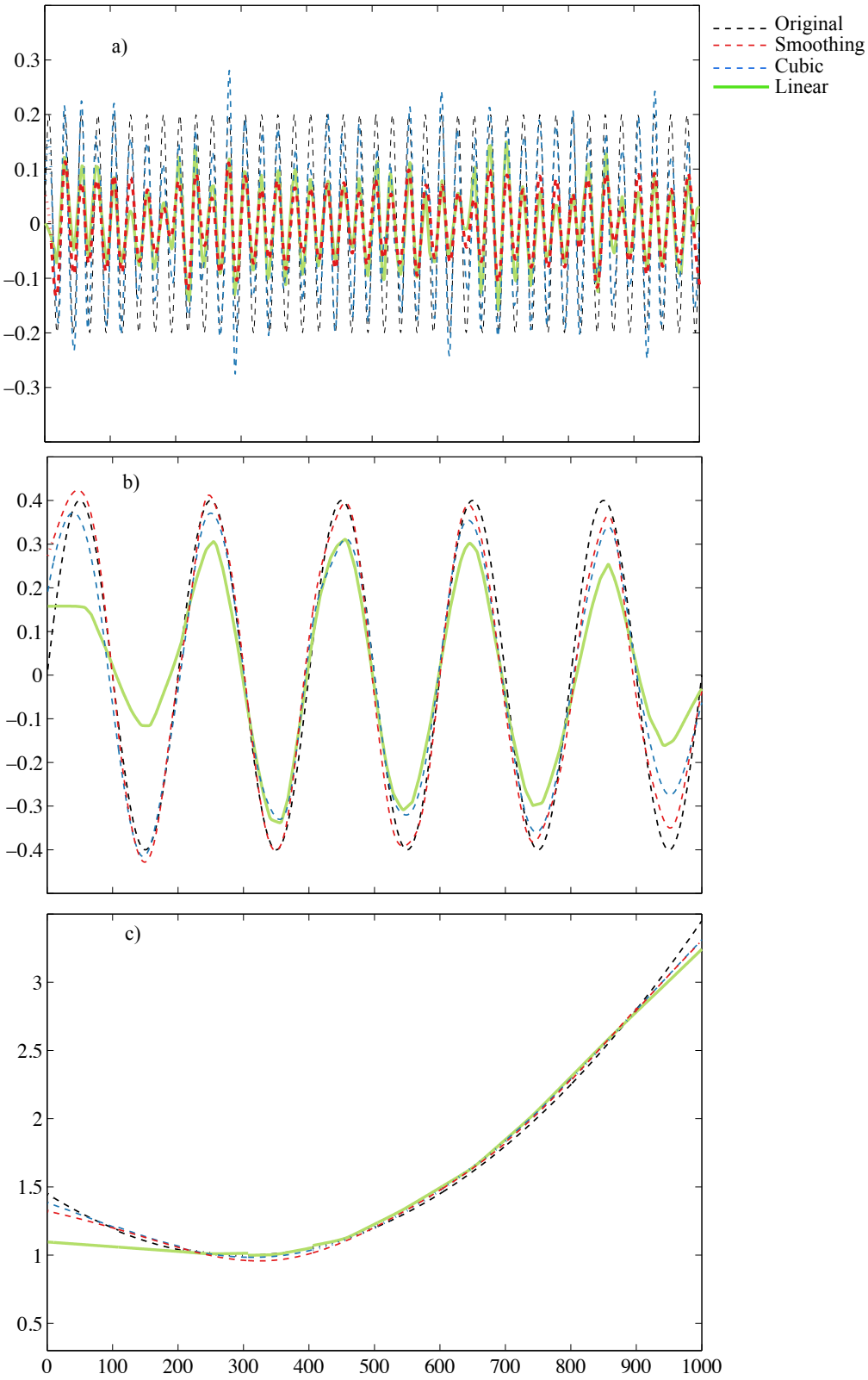

Figure A.10: EEMD results for the non-stationary data with time-dependent mean and constant noise, y_2 (Equation 4.9, Chapter 4).

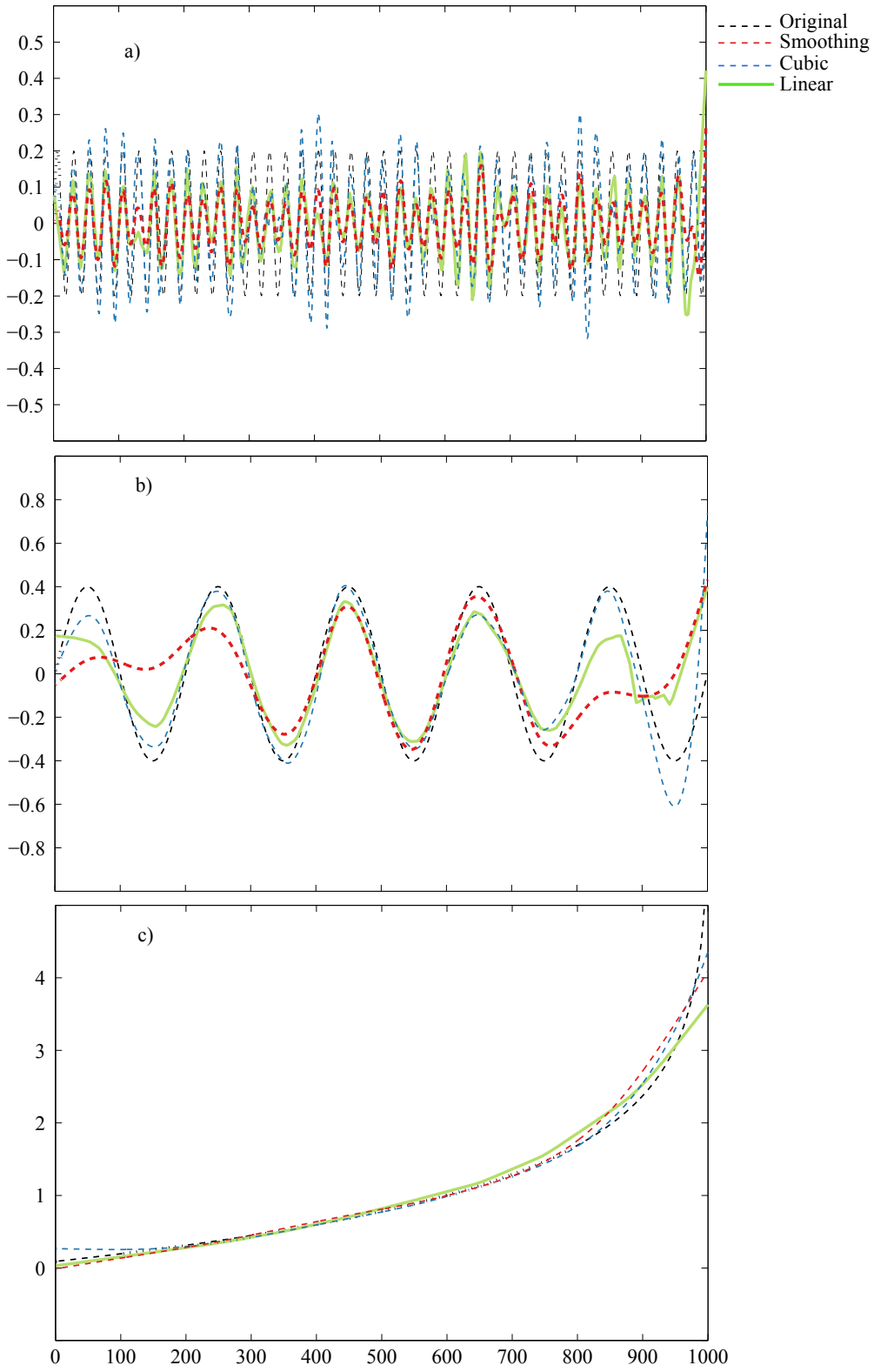

Figure A.11: EEMD results for the non-stationary data with time-dependent mean and constant noise, y_3 (Equation 4.9, Chapter 4).

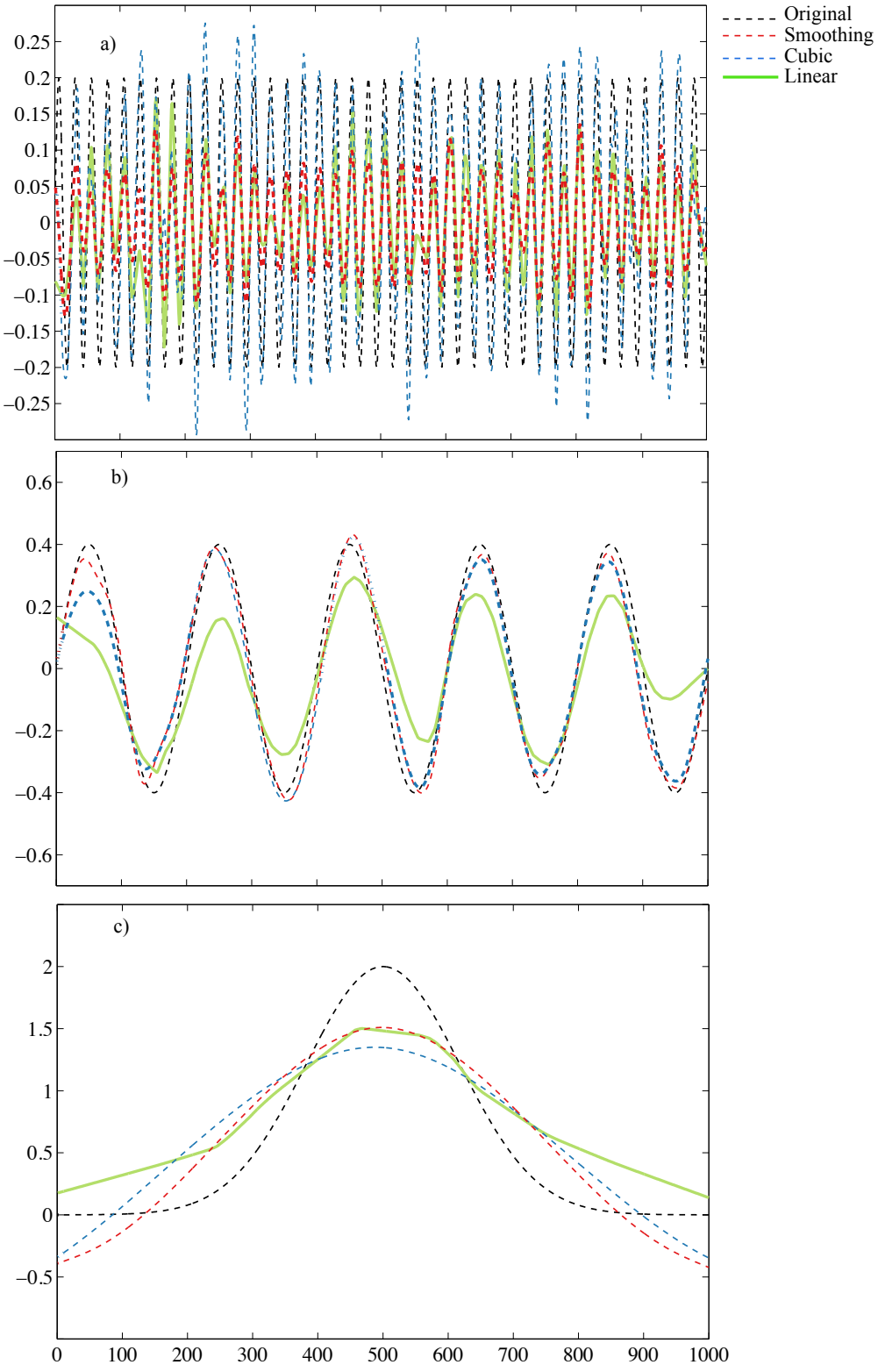

Figure A.12: EEMD results for the non-stationary data with time-dependent mean and constant noise, y_4 (Equation 4.9, Chapter 4).

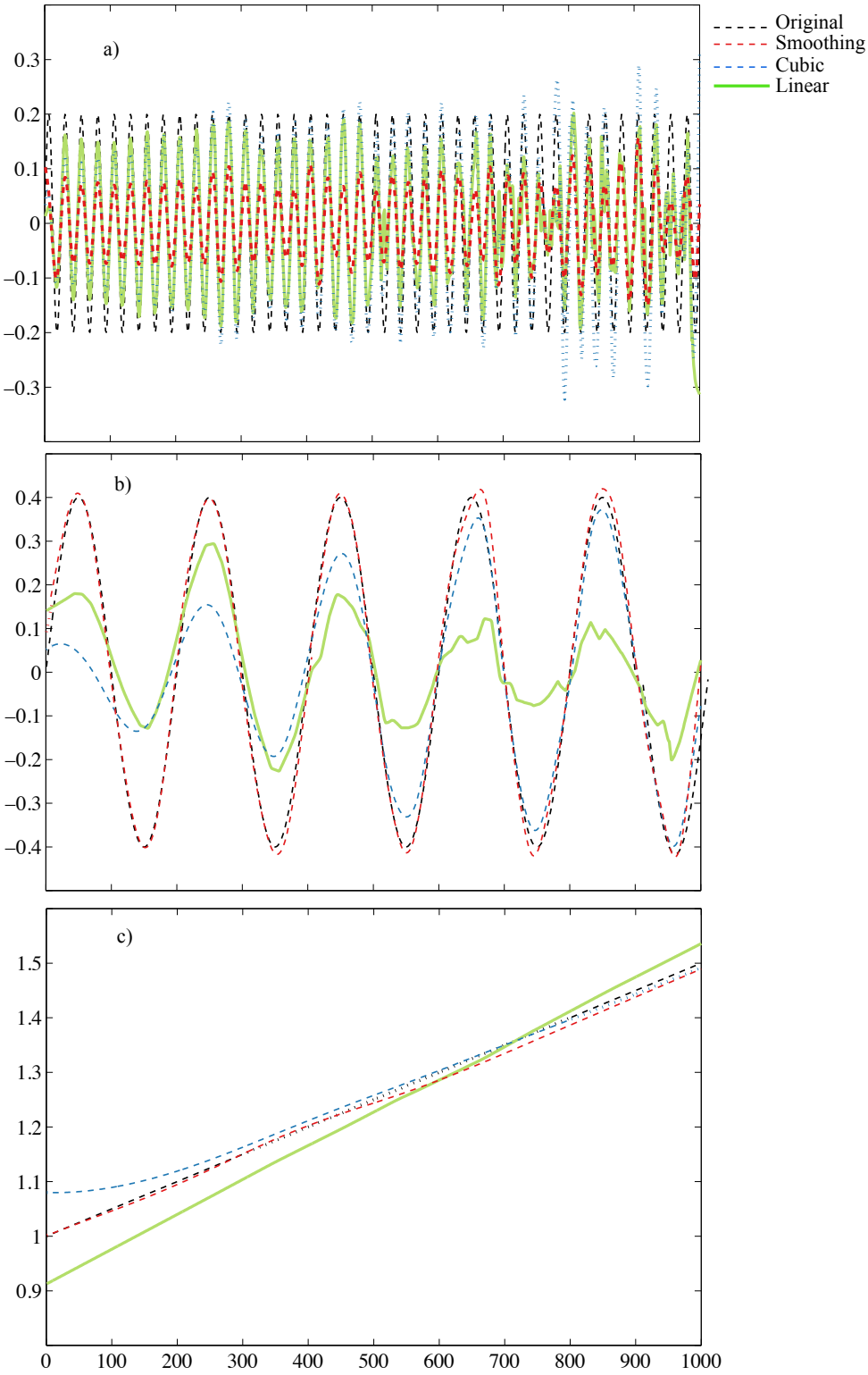

Figure A.13: EEMD results for the non-stationary data with time-dependent mean and variance, z_{15} (Equation 4.13, Chapter 4).

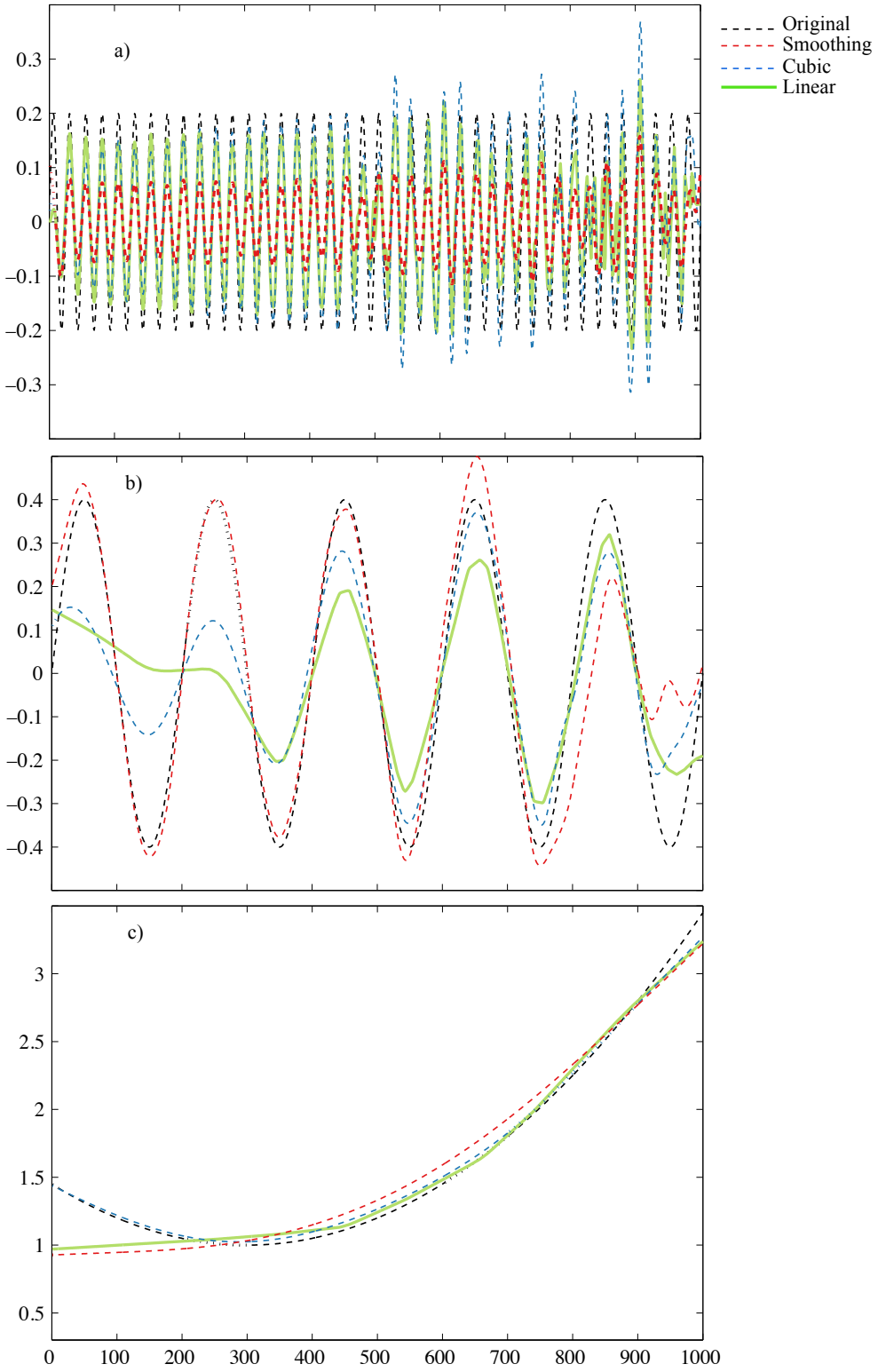

Figure A.14: EEMD results for the non-stationary data with time-dependent mean and variance, z_{25} (Equation 4.13, Chapter 4).

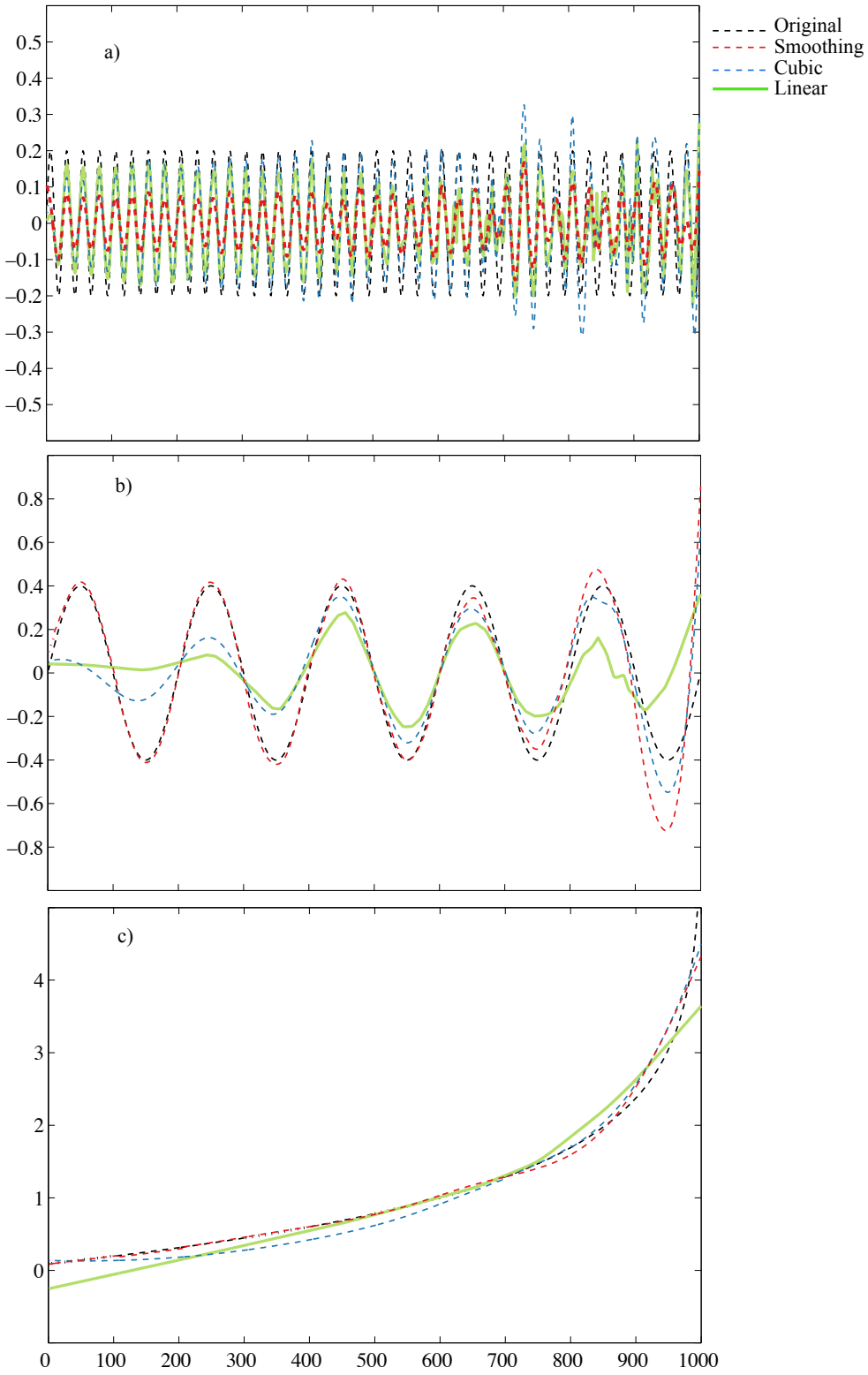

Figure A.15: EEMD results for the non-stationary data with time-dependent mean and variance, z_{35} (Equation 4.13, Chapter 4).

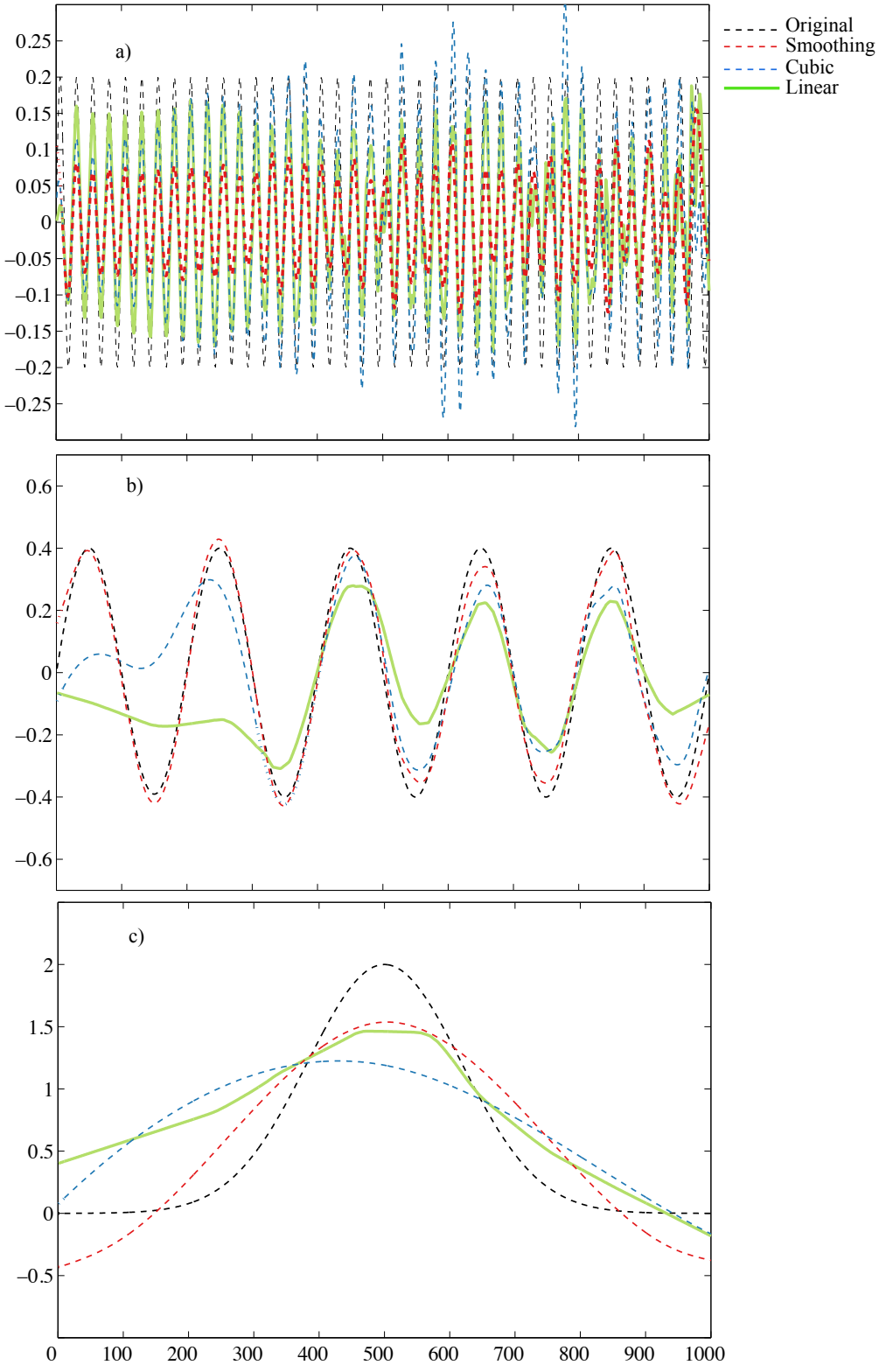

Figure A.16: EEMD results for the non-stationary data with time-dependent mean and variance, z_{45} (Equation 4.13, Chapter 4).

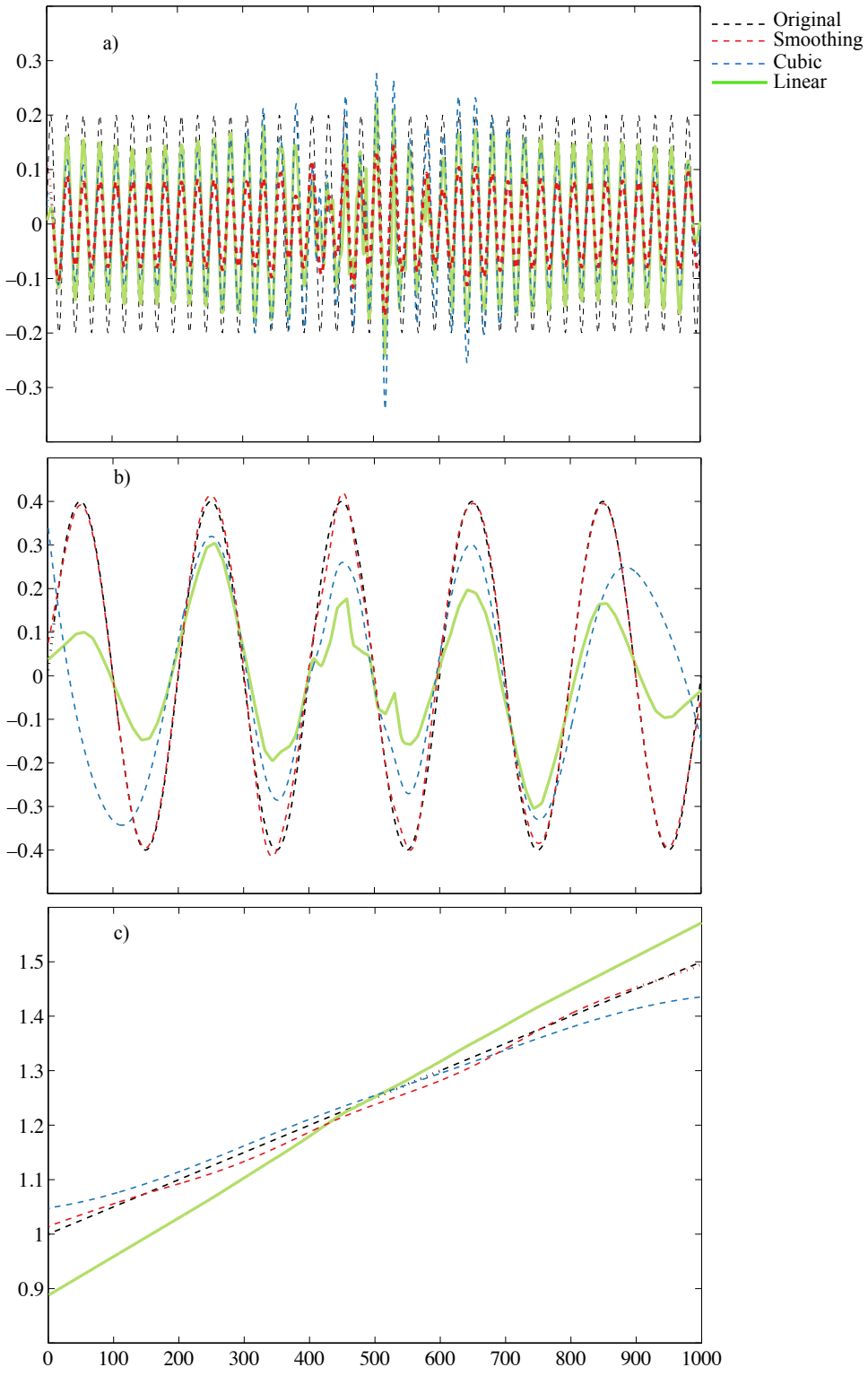

Figure A.17: EEMD results for the non-stationary data with time-dependent mean and variance, z_{16} (Equation 4.13, Chapter 4).

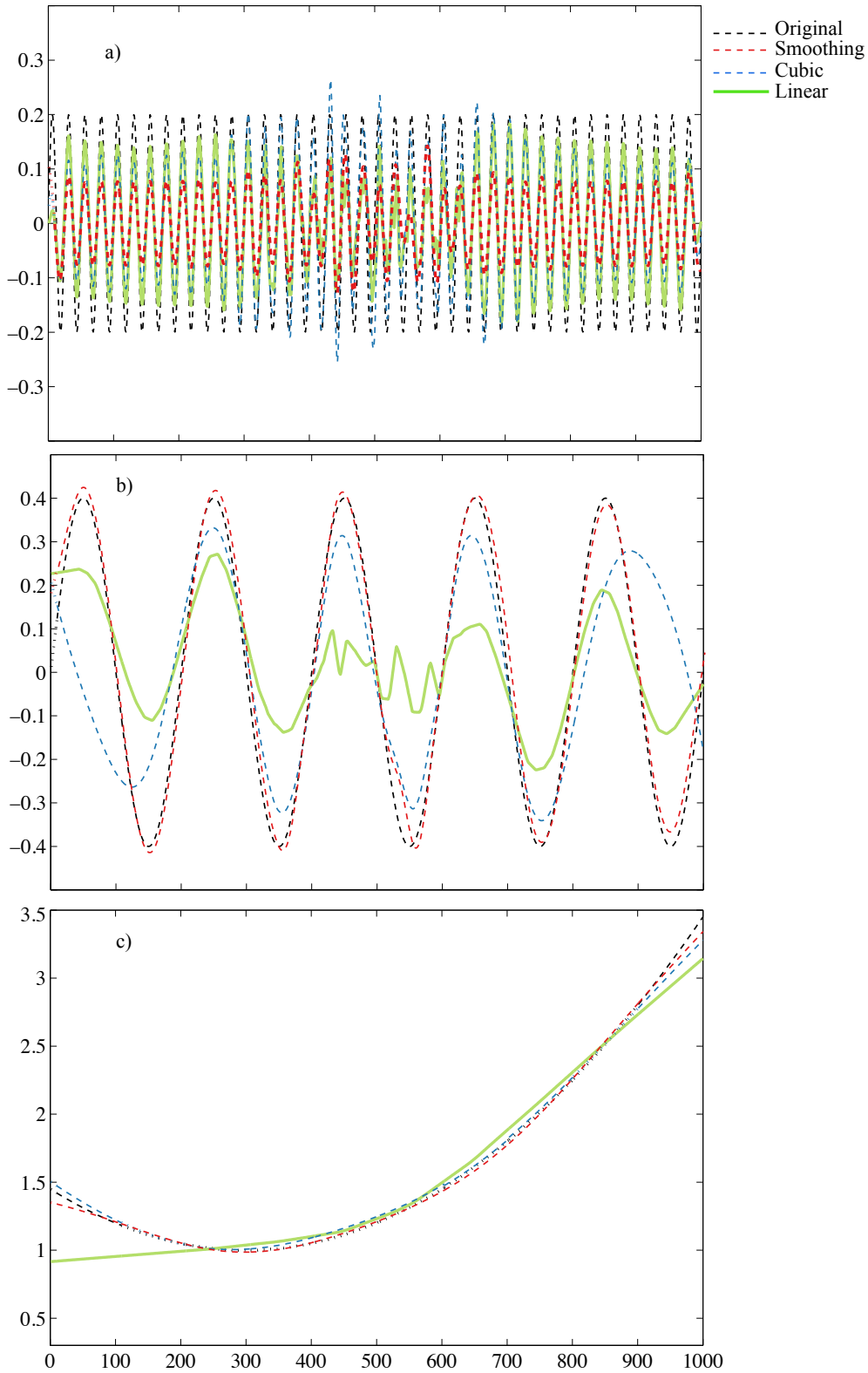

Figure A.18: EEMD results for the non-stationary data with time-dependent mean and variance, z_{26} (Equation 4.13, Chapter 4).

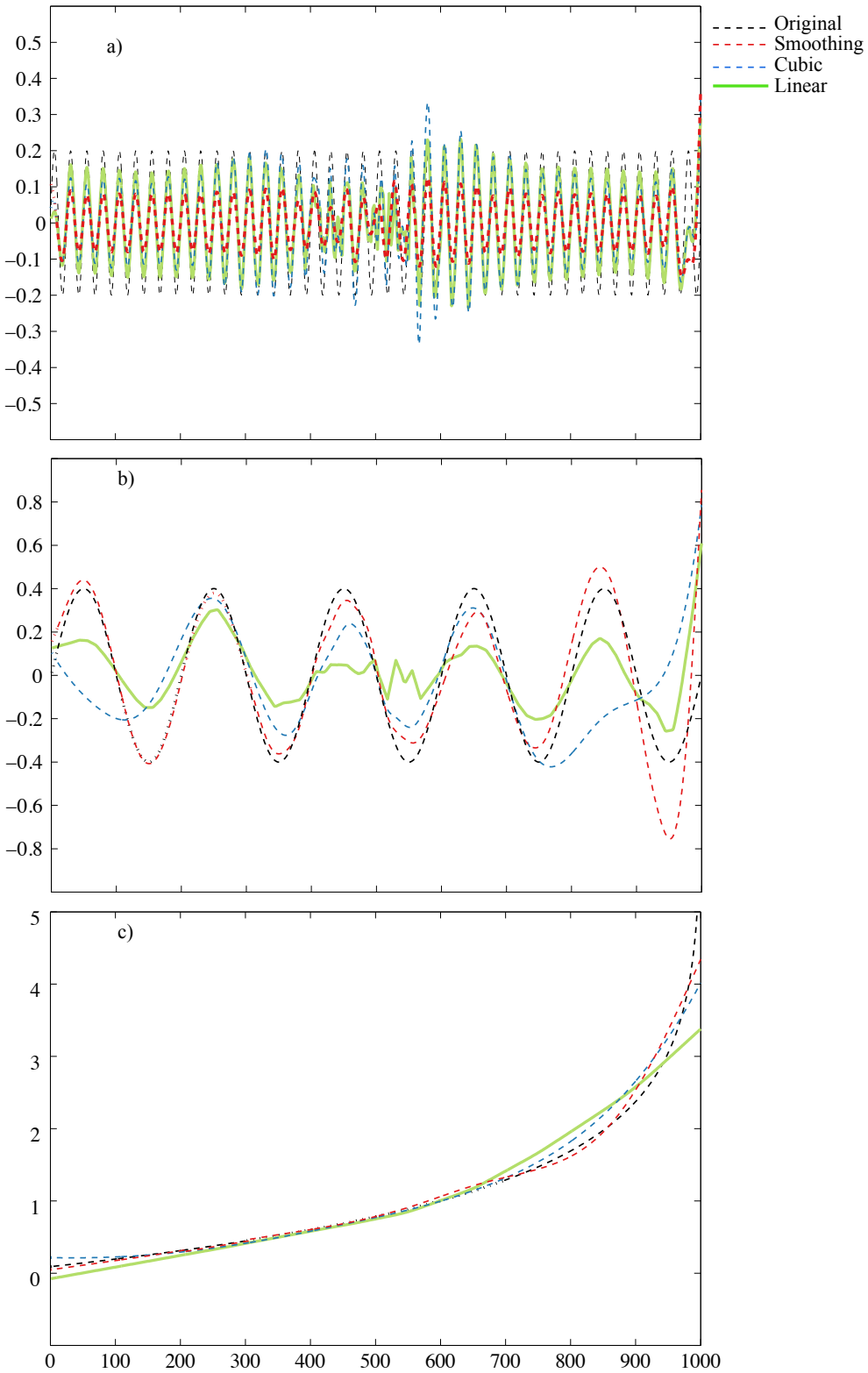

Figure A.19: EEMD results for the non-stationary data with time-dependent mean and variance, z_{36} (Equation 4.13, Chapter 4).

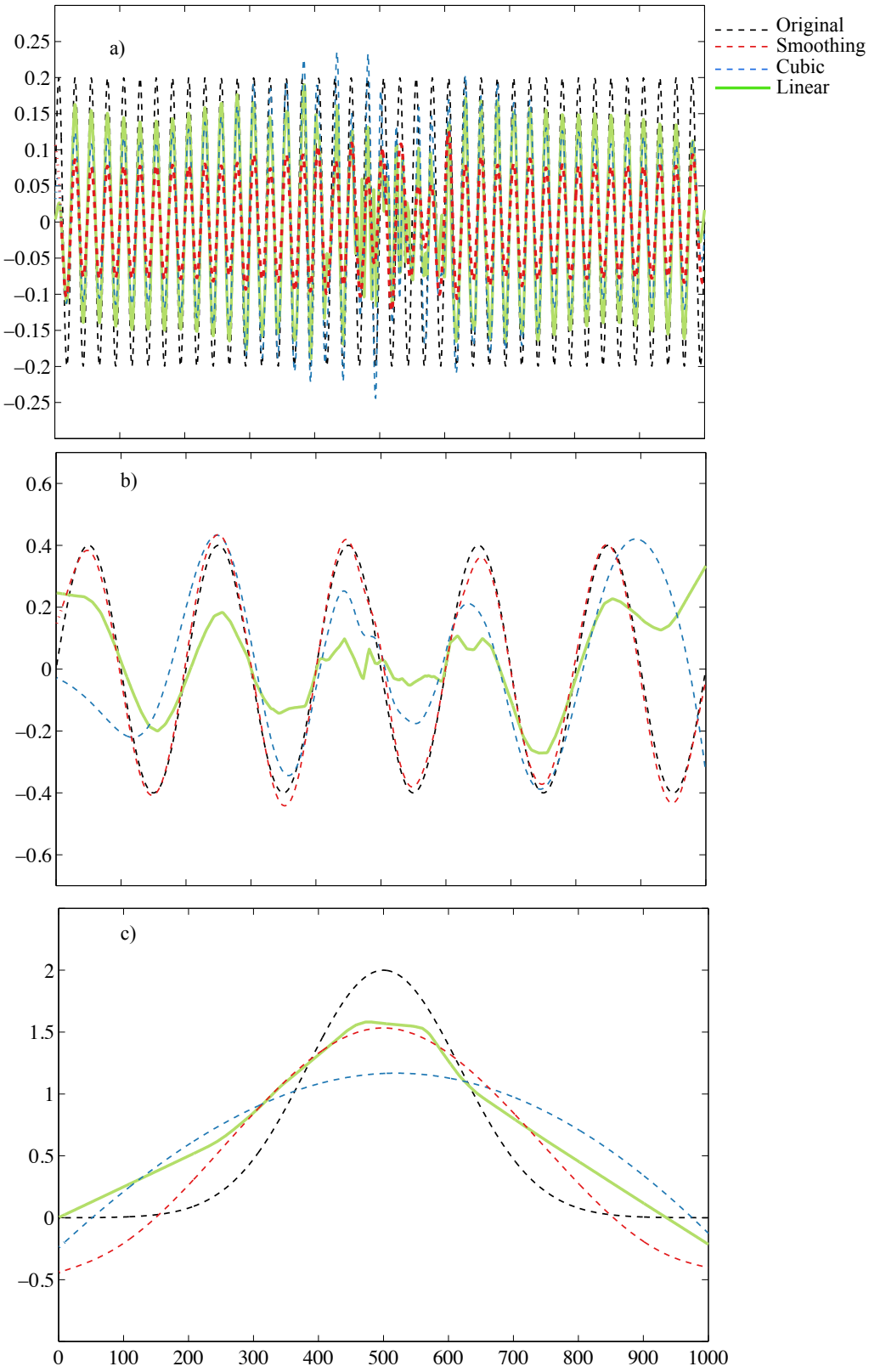

Figure A.20: EEMD results for the non-stationary data with time-dependent mean and variance, z_{46} (Equation 4.13, Chapter 4).

References

M. Ahmed, K. J. Anchukaitis, A. Asrat, H. P. Borgaonkar, M. Braida, B. M. Buckley, U. Büntgen, B. M. Chase, D. A. Christie, E. R. Cook, et al. Continental-scale temperature variability during the past two millennia. *Nature geoscience*, 6(5): 339, 2013.

J. I. Antonov, S. Levitus, and T. P. Boyer. Steric sea level variations during 1957–1994: Importance of salinity. *Journal of Geophysical Research: Oceans*, 107(C12), 2002.

F. Bahri and J. Sharples. Sensitivity of the Hilbert-Huang transform to interpolation methodology: examples using synthetic and ocean data. *MODSIM2015, 21st International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand*, pages 1324–1330, 2015.

R. Barbero, J. Abatzoglou, E. A. Steel, and N. K. Larkin. Modeling very large-fire occurrences over the continental United States from weather and climate forcing. *Environmental Research Letters*, 9(12):124009, 2014.

M. J. Bastiaans. A sampling theorem for the complex spectrogram, and gabor's expansion of a signal in gaussian elementary signals. *Optical Engineering*, 20(4): 204597–204597, 1981.

T. Beer and A. Williams. Estimating Australian forest fire danger under conditions of doubled carbon dioxide concentrations. *Climatic Change*, 29(2):169–188, 1995.

A.-O. Boudraa and J.-C. Cexus. Emd-based signal filtering. *Instrumentation and Measurement, IEEE Transactions on*, 56(6):2196–2202, 2007.

R. Bracewell. The Fourier Transform and IIS Applications. *New York*, 1965.

R. Bradstock, T. Penman, M. Boer, O. Price, and H. Clarke. Divergent responses of fire to recent warming and drying across south-eastern Australia. *Global change Biology*, 20(5):1412–1428, 2014.

R. J. Burgette, C. S. Watson, J. A. Church, N. J. White, P. Tregoning, and R. Coleman. Characterizing and minimizing the effects of noise in tide gauge time series: relative and geocentric sea level rise around Australia. *Geophysical Journal International*, page 131, 2013.

W. Cai. Antarctic ozone depletion causes an intensification of the Southern Ocean super-gyre circulation. *Geophysical Research Letters*, 33(3), 2006.

G. J. Cary, R. A. Bradstock, A. M. Gill, and R. J. Williams. *Global Change and Fire Regimes in Australia*. CSIRO Publishing: Collingwood, VIC, Canada, 2012.

P. Cetina-Heredia, M. Roughan, E. Van Sebille, and M. Coleman. Long-term trends in the East Australian Current separation latitude and eddy driven transport. *Journal of Geophysical Research: Oceans*, 119(7):4351–4366, 2014.

D. P. Chambers, M. A. Merrifield, and R. S. Nerem. Is there a 60-year oscillation in global mean sea level? *Geophysical Research Letters*, 39(18), 2012.

C. Chandler, P. Cheney, P. Thomas, L. Trabaud, and D. Williams. *Fire in Forestry. Volume 1. Forest fire behavior and effects. Volume 2. Forest fire management and Organization.* John Wiley & Sons, Inc., 1983.

R. Chandler and M. Scott. *Statistical methods for trend detection and analysis in the environmental sciences*. John Wiley & Sons, 2011.

Y. Changwei, Z. Jianjing, F. Gang, H. Zhiyi, and Z. Cheng. Time-frequency analysis of seismic response of a high steep hill with two side slopes when subjected to ground shaking by using HHT. 2012.

Q. Chen, N. Huang, S. Riemenschneider, and Y. Xu. A B-spline approach for empirical mode decompositions. *Advances in Computational Mathematics*, 24(1-4):171–195, 2006.

Y. Cheng, T. Ezer, and B. D. Hamlington. Sea level acceleration in the China Seas. *Water*, 8(7):293, 2016.

J. A. Church and N. J. White. Sea-level rise from the late 19th to the early 21st century. *Surveys in Geophysics*, 32(4-5):585–602, 2011.

J. A. Church, J. R. Hunter, K. L. McInnes, and N. J. White. Sea-level rise around the Australian coastline and the changing frequency of extreme sea-level events. *Australian Meteorological Magazine*, 55(4):253–260, 2006.

J. A. Church, N. J. White, T. Aarup, W. S. Wilson, P. L. Woodworth, C. M. Domingues, J. R. Hunter, and K. Lambeck. Understanding global sea levels: past, present and future. *Sustainability Science*, 3(1):9–22, 2008.

J. A. Church, P. U. Clark, A. Cazenave, J. M. Gregory, S. Jevrejeva, A. Levermann, M. A. Merrifield, G. A. Milne, R. S. Nerem, and P. D. Nunn. Sea level change. Technical report, PM Cambridge University Press, 2013.

J. G. Cogley. Geodetic and direct mass-balance measurements: comparison and joint analysis. *Annals of Glaciology*, 50(50):96–100, 2009.

D. Coumou and S. Rahmstorf. A decade of weather extremes. *Nature climate change*, 2(7):491, 2012.

T. J. Crowley. Causes of climate change over the past 1000 years. *Science*, 289 (5477):270–277, 2000.

S. Dangendorf, C. Mudersbach, T. Wahl, and J. Jensen. Characteristics of inter-annual and decadal sea-level variability and the role of meteorological forcing: the long record of Cuxhaven. *Ocean Dynamics*, 63(2-3):209–224, 2013.

M. Dätig and T. Schlurmann. Performance and limitations of the Hilbert–Huang transformation (HHT) with an application to irregular water waves. *Ocean Engineering*, 31(14):1783–1834, 2004.

C. De Boor, C. De Boor, E.-U. Mathématicien, C. De Boor, and C. De Boor. *A practical guide to splines*, volume 27. Springer-Verlag New York, 1978.

J. E. Deeming. *National fire-danger rating system*, volume 84. Rocky Mountain Forest and Range Experiment Station, Forest Service, US Dept. of Agriculture, 1972.

R. Deering and J. F. Kaiser. The use of a masking signal to improve empirical mode decomposition. In *IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005. Proceedings.(ICASSP'05).*, volume 4, pages iv–485. IEEE, 2005.

E. Deléchelle, J. Lemoine, and O. Niang. Empirical mode decomposition: an analytical approach for sifting process. *IEEE Signal Processing Letters,*, 12(11):764–767, 2005.

Y. Deng, W. Wang, C. Qian, Z. Wang, and D. Dai. Boundary-processing-technique in EMD method and Hilbert transform. *Chinese Science Bulletin*, 46(11):954–960, 2001.

D. Donnelly. The Fast Fourier and Hilbert-Huang Transforms: a comparison. In *Computational Engineering in Systems Applications, IMACS Multiconference on*, volume 1, pages 84–88. IEEE, 2006.

B. Douglas, M. T. Kearney, and S. P. Leatherman. *Sea level rise: History and consequences*, volume 75. Academic Press, 2000.

B. C. Douglas. Global sea level rise. *Journal of Geophysical Research: Oceans*, 96 (C4):6981–6992, 1991.

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

D. G. Duffy. The application of Hilbert-Huang transforms to meteorological datasets. *Hilbert-Huang Transform and Its Applications, edited by Huang, NE and Shen, SP*, pages 129–147, 2005.

M. H. England, S. McGregor, P. Spence, G. A. Meehl, A. Timmermann, W. Cai, A. S. Gupta, M. J. McPhaden, A. Purich, and A. Santoso. Recent intensification of wind-driven circulation in the Pacific and the ongoing warming hiatus. *Nature Climate Change*, 4(3):222–227, 2014.

T. Ezer, L. P. Atkinson, W. B. Corlett, and J. L. Blanco. Gulf Stream's induced sea level rise and variability along the US Mid-Atlantic coast. *Journal of Geophysical Research: Oceans*, 118(2):685–697, 2013.

M. Feng, Y. Li, and G. Meyers. Multidecadal variations of Fremantle sea level: Footprint of climate variability in the tropical Pacific. *Geophysical Research Letters*, 31(16), 2004.

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

M. D. Flannigan, M. A. Krawchuk, W. J. de Groot, B. M. Wotton, and L. M. Gowman. Implications of changing climate for global wildland fire. *International Journal of Wildland Fire*, 18(5):483–507, 2009.

G. Foster and S. Rahmstorf. Global temperature evolution 1979? 2010. *Environmental Research Letters*, 6(4):044022, 2011.

P. Fox-Hughes, R. Harris, G. Lee, M. Grose, and N. Bindoff. Future fire danger climatology for Tasmania, Australia, using a dynamically downscaled regional climate model. *International Journal of Wildland Fire*, 23(3):309–321, 2014.

M. Ghil and R. Vautard. Interdecadal oscillations and the warming trend in global temperature time series. *Nature*, 350:324–327, 1991.

M. Ghil, M. Allen, M. Dettinger, K. Ide, D. Kondrashov, M. Mann, A. W. Robertson, A. Saunders, Tian, Y, and F. Varadi. Advanced spectral methods for climatic time series. *Reviews of Geophysics*, 40(1):3–1, 2002.

J. Godfrey, G. Cresswell, T. Golding, A. Pearce, and R. Boyd. The separation of the East Australian Current. *Journal of Physical Oceanography*, 10(3):430–440, 1980.

J. A. Gregory and M. Sarfraz. A rational cubic spline with tension. *Computer Aided Geometric Design*, 7(1-4):1–13, 1990.

D. Griffiths. *Improved formulae for the McArthur forest fire danger meter.* Bureau of Meteorology, 1998.

M. R. Grose, P. Fox-Hughes, R. M. Harris, and N. L. Bindoff. Changes to the drivers of fire weather with a warming climate–a case study of southeast Tasmania. *Climatic Change*, 124(1-2):255–269, 2014.

S. L. Hahn. *Hilbert Transforms in Signal Processing.* Artech House on Demand, 1996.

B. Hamon. *A century of tide records: Sydney (Fort Denison) 1886-1986.* Flinders Institute for Atmospheric and Marine Sciences, Flinders University of South Australia, 1987.

J. Hansen, M. Sato, R. Ruedy, K. Lo, D. W. Lea, and M. Medina-Elizade. Global temperature change. *Proceedings of the National Academy of Sciences*, 103(39): 14288–14293, 2006.

N. Harvey, E. J. Barnett, R. P. Bourman, and A. P. Belperio. Holocene sea-level change at Port Pirie, South Australia: a contribution to global sea-level rise estimates from tide gauges. *Journal of Coastal Research*, pages 607–615, 1999.

A. Hasson, G. Mills, B. Timbal, and K. Walsh. Assessing the impact of climate change on extreme fire weather events over Southeastern Australia. *Climate Research*, 39(2):159–172, 2009.

C. Hay, J. X. Mitrovica, N. Gomez, J. R. Creveling, J. Austermann, and R. E. Kopp. The sea-level fingerprints of ice-sheet collapse during interglacial periods. *Quaternary Science Reviews*, 87:60–69, 2014.

N. Herold and A. Santoso. Indian ocean warming during peak el niño cools surrounding land masses. *Climate Dynamics*, pages 1–16, 2017.

K. Hill, S. Rintoul, R. Coleman, and K. Ridgway. Wind forced low frequency variability of the east australia current. *Geophysical Research Letters*, 35(8), 2008.

F. Hlawatsch and F. Auger. *Time-frequency Analysis*. John Wiley & Sons, 2013.

N. J. Holbrook, I. D. Goodwin, S. McGregor, E. Molina, and S. B. Power. ENSO to multi-decadal time scale changes in East Australian Current transports and Fort Denison sea level: Oceanic Rossby waves as the connecting mechanism. *Deep Sea Research Part II: Topical Studies in Oceanography*, 58(5):547–558, 2011.

S. Holgate. On the decadal rates of sea level change during the twentieth century. *Geophysical Research Letters*, 34(1), 2007.

P. Holmes, J. L. Lumley, and G. Berkooz. *Turbulence, coherent structures, dynamical systems and symmetry*. Cambridge University Press, 1998.

J. Hong, J. Kim, H. Ishikawa, and Y. Ma. Surface layer similarity in the nocturnal boundary layer: the application of Hilbert-Huang transform. *Biogeosciences*, 7 (4):1271–1278, 2010.

L.-C. Hong. *Super El Niño*. Springer, 2016.

J. R. Houston and R. G. Dean. Sea-level acceleration based on us tide gauges and extensions of previous global-gauge analyses. *Journal of Coastal Research*, 27(3): 409–417, 2011.

J. R. Houston and R. G. Dean. Effects of sea-level decadal variability on acceleration and trend difference. *Journal of Coastal Research*, 29(5):1062–1072, 2013.

H. Huang and J. Pan. Speech pitch determination based on hilbert-huang transform. *Signal Processing*, 86(4):792–803, 2006.

N. E. Huang and S. S. Shen. *Hilbert-Huang transform and its applications*, volume 5. World Scientific, 2005.

N. E. Huang and Z. Wu. A review on Hilbert-Huang transform: Method and its applications to geophysical studies. *Reviews of Geophysics*, 46(2), 2008.

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

spectrum for nonlinear and non-stationary time series analysis. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, volume 454, pages 903–995. The Royal Society, 1998a.

N. E. Huang, Z. Shen, and S. R. Long. A new view of nonlinear water waves: The Hilbert Spectrum 1. *Annual Review of Fluid Mechanics*, 31(1):417–457, 1999.

N. E. Huang, M.-L. Wu, W. Qu, S. R. Long, S. S. Shen, and J. E. Zhang. Applications of Hilbert-Huang Transform to non-stationary financial time series analysis. *Applied Stochastic Models in Business and Industry*, 19(3):245–268, 2003.

W. Huang, Z. Shen, N. E. Huang, and Y. C. Fung. Engineering analysis of biological variables: an example of blood pressure over 1 day. *Proceedings of the National Academy of Sciences*, 95(9):4816–4821, 1998b.

L. Hughes. Climate change and Australia: trends, projections and impacts. *Austral Ecology*, 28(4):423–443, 2003.

M. Hulme, J. Turnpenny, and G. J. Jenkins. *Climate change scenarios for the United Kingdom: the UKCIP02 Briefing Report*. Tyndall Centre for Climate Change Research, School of Environmental Sciences, University of East Anglia Norwich, UK, 2002.

M. Imani, R.-J. You, and C.-Y. Kuo. Analysis and prediction of Caspian Sea level pattern anomalies observed by satellite altimetry using autoregressive integrated moving average models. *Arabian Journal of Geosciences*, 7(8):3339–3348, 2014.

S. Jevrejeva, A. Grinsted, J. Moore, and S. Holgate. Nonlinear trends and multiyear cycles in sea level records. *Journal of Geophysical Research: Oceans*, 111(C9), 2006.

S. Jevrejeva, J. Moore, and A. Grinsted. Relative importance of mass and volume changes to global sea level rise. *Journal of Geophysical Research: Atmospheres*, 113(D8), 2008.

D.-j. H. Jin-Ping, ZHAO. Mirror extending and circular spline function for empirical mode decomposition method. *Journal of Zhejiang University Science*, 2(3):247–252, 2001.

P. Jones. Recent warming in global temperature series. *Geophysical Research Letters*, 21(12):1149–1152, 1994.

P. D. Jones, M. New, D. E. Parker, S. Martin, and I. G. Rigor. Surface air temperature and its changes over the past 150 years. *Reviews of Geophysics*, 37(2): 173–199, 1999.

R. K. Kaufmann, H. Kauppi, M. L. Mann, and J. H. Stock. Reconciling anthropogenic climate change with observed temperature 1998–2008. *Proceedings of the National Academy of Sciences*, 108(29):11790–11793, 2011.

J. J. Keetch and G. M. Byram. A drought index for forest fire control. *Res. Pap. SE-38. Asheville, NC: US Department of Agriculture, Forest Service, Southeastern Forest Experiment Station. 35 p.*, 38, 1968.

A. C. Kemp, B. P. Horton, J. P. Donnelly, M. E. Mann, M. Vermeer, and S. Rahmstorf. Climate related sea-level variations over the past two millennia. *Proceedings of the National Academy of Sciences*, 108(27):11017–11022, 2011.

J. Kenigson and W. Han. Detecting and understanding the accelerated sea level rise along the east coast of the United States during recent decades. *Journal of Geophysical Research: Oceans*, 119(12):8749–8766, 2014.

R. Khoury and D. W. Harder. *Numerical Methods and Modelling for Engineering*. Springer, 2015.

K. J. King, R. M. de Ligt, and G. J. Cary. Fire and carbon dynamics under climate change in south-eastern Australia: insights from FullCAM and FIRESCAPE modelling. *International Journal of Wildland Fire*, 20(4):563–577, 2011.

A. Köhl and D. Stammer. Decadal sea level changes in the 50-year gecco ocean synthesis. *Journal of Climate*, 21(9):1876–1890, 2008.

Z. Kolláth and K. Oláh. Multiple and changing cycles of active stars-i. methods of analysis and application to the solar cycles. *Astronomy & Astrophysics*, 501(2): 695–702, 2009.

R. E. Kopp, A. C. Kemp, K. Bittermann, B. P. Horton, J. P. Donnelly, W. R. Gehrels, C. C. Hay, J. X. Mitrovica, E. D. Morrow, and S. Rahmstorf. Temperature-driven global sea-level variability in the common era. *Proceedings of the National Academy of Sciences*, page 201517056, 2016.

Y. Kopsinis and S. S. McLaughlin. Enhanced empirical mode decomposition using a novel sifting-based interpolation points detection. In *IEEE/SP 14th Workshop on Statistical Signal Processing, 2007. SSP'07.*, pages 725–729. IEEE, 2007.

A. Levermann, A. Griesel, M. Hofmann, M. Montoya, and S. Rahmstorf. Dynamic sea level changes following changes in the thermohaline circulation. *Climate Dynamics*, 24(4):347–354, 2005.

D.-C. Lin, Z.-L. Guo, F.-P. An, and F.-L. Zeng. Elimination of end effects in empirical mode decomposition by mirror image coupled with support vector regression. *Mechanical Systems and Signal Processing*, 31:13–28, 2012.

B. Liu, S. Riemenschneider, and Y. Xu. Gearbox fault diagnosis using empirical mode decomposition and Hilbert spectrum. *Mechanical Systems and Signal Processing*, 20(3):718–734, 2006.

Y. Liu, J. Stanturf, and S. Goodrick. Trends in global wildfire potential in a changing climate. *Forest Ecology and Management*, 259(4):685–697, 2010.

D. M. Livingstone. Lake oxygenation: Application of a one-box model with ice cover. *International Review of Hydrobiology*, 78(4):465–480, 1993.

A. Lombard, G. Garric, and T. Penduff. Regional patterns of observed sea level change: insights from a 1/4 global ocean/sea-ice hindcast. *Ocean Dynamics*, 59 (3):433–449, 2009.

C. Lucas. On developing a historical fire weather data-set for Australia. *Australian Meteorological and Oceanographic Journal*, 60(1):1, 2010.

C. Lucas, K. Hennessy, G. Mills, and J. Bathols. Bushfire weather in southeast australia: recent trends and projected climate change impacts. 2007.

K. Lyu, X. Zhang, J. A. Church, A. B. Slangen, and J. Hu. Time of emergence for regional sea-level change. *Nature Climate Change*, 4(11):1006–1010, 2014.

H. Madsen. *Time Series Analysis*. CRC Press, 2007.

M. E. Mann and J. Park. Global-scale modes of surface temperature variability on interannual to century timescales. *Journal of Geophysical Research: Atmospheres*, 99(D12):25819–25833, 1994.

M. E. Mann, E. Gille, J. Overpeck, W. Gross, R. S. Bradley, F. T. Keimig, and M. K. Hughes. Global temperature patterns in past centuries: An interactive presentation. *Earth Interactions*, 4(4):1–1, 2000.

M. Mariani and M.-S. Fletcher. The Southern Annular Mode determines interannual and centennial-scale fire activity in temperate southwest Tasmania, Australia. *Geophysical Research Letters*, 2016.

M. D. Mastrandrea, C. B. Field, T. F. Stocker, O. Edenhofer, K. L. Ebi, D. J. Frame, H. Held, E. Kriegler, K. J. Mach, and P. R. Matschoss. Guidance note for lead authors of the IPCC fifth assessment report on consistent treatment of Uncertainties. 2010.

S. Matthews. A comparison of fire danger rating systems for use in forests. *Australian Meteorological and Oceanographic Journal*, 58(1):41, 2009.

S. Matthews, A. L. Sullivan, P. Watson, and R. J. Williams. Climate change, fuel and fire behaviour in a eucalypt forest. *Global Change Biology*, 18(10):3212–3223, 2012.

A. G. McArthur. Fire Behaviour in Eucalypt Forests. 1967.

G. A. Meehl, W. M. Washington, W. D. Collins, J. M. Arblaster, A. Hu, L. E. Buja, W. G. Strand, and H. Teng. How much more global warming and sea level rise? *Science*, 307(5716):1769–1772, 2005.

G. A. Meehl, T. F. Stocker, W. D. Collins, P. Friedlingstein, A. T. Gaye, J. M. Gregory, A. Kitoh, R. Knutti, J. M. Murphy, and A. Noda. Global climate projections. *Climate Change*, 3495:747–845, 2007.

S. Meignen and V. Perrier. A new formulation for empirical mode decomposition based on constrained optimization. *IEEE Signal Processing Letters,*, 14(12):932–935, 2007.

B. Metz. *Climate change 2001: mitigation: contribution of Working Group III to the third assessment report of the Intergovernmental Panel on Climate Change*, volume 3. Cambridge University Press, 2001.

B. Meyssignac and A. Cazenave. Sea level: a review of present-day and recent-past changes and variability. *Journal of Geodynamics*, 58:96–109, 2012.

W. Mitchell, J. Chittleborough, B. Ronai, and G. Lennon. Sea level rise in Australia and the Pacific. *The South Pacific Sea Level and Climate Change, Quarterly Newsletter*, 5:10–19, 2000.

J. X. Mitrovica, M. E. Tamisiea, J. L. Davis, and G. A. Milne. Recent mass balance of polar ice sheets inferred from patterns of global sea-level change. *Nature*, 409 (6823):1026–1029, 2001.

P. Moin. *Fundamentals of Engineering Numerical Analysis*. Cambridge University Press, 2010.

M. K. I. Molla, A. Sumi, and M. S. Rahman. Analysis of temperature change under global warming impact using empirical mode decomposition. *International Journal of Information Technology*, 3(2):131–139, 2006.

S. Mooney, S. Harrison, P. Bartlein, A.-L. Daniau, J. Stevenson, K. Brownlie, S. Buckman, M. Cupper, J. Luly, and M. Black. Late Quaternary fire regimes of Australasia. *Quaternary Science Reviews*, 30(1):28–46, 2011.

M. Mudelsee. *Climate Time Series Analysis*. Springer, 2013.

B. F. Murphy and B. Timbal. A review of recent climate variability and climate change in southeastern Australia. *International journal of Climatology*, 28(7): 859–879, 2008.

R. Neely, O. Toon, S. Solomon, J.-P. Vernier, C. Alvarez, J. English, K. Rosenlof, M. Mills, C. Bardeen, and J. Daniel. Recent anthropogenic increases in SO_2 from Asia have minimal impact on stratospheric aerosol. *Geophysical Research Letters*, 40(5):999–1004, 2013.

R. M. Nelson Jr. Prediction of diurnal change in 10-h fuel stick moisture content. *Canadian Journal of Forest Research*, 30(7):1071–1087, 2000.

N. Nicholls. The changing nature of Australian droughts. *Climatic change*, 63(3): 323–336, 2004.

R. J. Nicholls and A. Cazenave. Sea-level rise and its impact on coastal zones. *Science*, 328(5985):1517–1520, 2010.

I. Noble, A. Gill, and G. Bary. Mcarthur's fire-danger meters expressed as equations. *Australian Journal of Ecology*, 5(2):201–203, 1980.

J. Oerlemans, M. Dyurgerov, and R. Van de Wal. Reconstructing the glacier contribution to sea-level rise back to 1850. *The Cryosphere*, 1(1):59–65, 2007.

S. Okumura. *The short time Fourier transform and local signals*. PhD thesis, Carnegie Mellon University, 2011.

M. Olivieri and G. Spada. Intermittent sea-level acceleration. *Global and planetary change*, 109:64–72, 2013.

I. P. on Climate Change. *Climate Change 2014 Impacts, Adaptation and Vulnerability: Regional Aspects*. Cambridge University Press, 2014.

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

A. Parker. Sea level trends at locations of the United States with more than 100 years of recording. *Natural Hazards*, 65(1):1011–1021, 2013.

E. Parzen. *Stochastic processes*, volume 24. SIAM, 1999.

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

M. Peel, T. McMahon, R. Srikanthan, and K. Tan. 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*, page 702. Engineers Australia, 2011.

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

W. Peltier, D. Argus, and R. Drummond. Space geodesy constrains ice age terminal deglaciation: The global ice-6g_c (vm5a) model. *Journal of Geophysical Research: Solid Earth*, 120(1):450–487, 2015.

Z. Peng, W. T. Peter, and F. Chu. A comparison study of improved Hilbert–Huang transform and wavelet transform: application to fault diagnosis for rolling bearing. *Mechanical Systems and Signal Processing*, 19(5):974–988, 2005.

S. C. Phillips, M. T. Swain, A. P. Wiley, J. W. Essex, and C. M. Edge. Reversible digitally filtered molecular dynamics. *The Journal of Physical Chemistry B*, 107 (9):2098–2110, 2003.

A. Pigorini, A. G. Casali, S. Casarotto, F. Ferrarelli, G. Baselli, M. Mariotti, M. Massimini, and M. Rosanova. Time–frequency spectral analysis of TMS-evoked EEG oscillations by means of Hilbert–Huang transform. *Journal of Neuroscience Methods*, 198(2):236–245, 2011.

D. J. Pines and L. W. Salvino. Health monitoring of one-dimensional structures using empirical mode decomposition and the Hilbert-Huang Transform. In *SPIE's*

9th Annual International Symposium on Smart Structures and Materials, pages 127–143. International Society for Optics and Photonics, 2002.

A. Pitman, G. Narisma, and J. McAneney. The impact of climate change on the risk of forest and grassland fires in Australia. *Climatic Change*, 84(3):383–402, 2007.

Y. N. Pokhrel, N. Hanasaki, P. J. Yeh, T. J. Yamada, S. Kanae, and T. Oki. Model estimates of sea-level change due to anthropogenic impacts on terrestrial water storage. *Nature Geoscience*, 5(6):389–392, 2012.

M. B. Priestley. Non-linear and Non-stationary Time Series Analysis. 1988.

D. Pugh. *Changing Sea Levels: Effects of Tides, Weather and Climate*. Cambridge University Press, 2004.

R. G. Quayle, T. C. Peterson, A. N. Basist, and C. S. Godfrey. An operational near-real-time global temperature index. *Geophysical Research Letters*, 26(3):333–335, 1999.

S. Rahmstorf. A semi-empirical approach to projecting future sea-level rise. *Science*, 315(5810):368–370, 2007.

S. Rahmstorf. Sea-level rise: towards understanding local vulnerability. *Environmental Research Letters*, 7(2):021001, 2012a.

S. Rahmstorf. Modeling sea level rise. *Nature Education Knowledge*, 3(3):4, 2012b.

V. Rai and A. Mohanty. Bearing fault diagnosis using FFT of intrinsic mode functions in Hilbert–Huang transform. *Mechanical Systems and Signal Processing*, 21 (6):2607–2615, 2007.

W. J. Randel, R. R. Garcia, N. Calvo, and D. Marsh. Enso influence on zonal mean temperature and ozone in the tropical lower stratosphere. *Geophysical Research Letters*, 36(15), 2009.

E. M. Rasmusson, X. Wang, and C. F. Ropelewski. The biennial component of ENSO variability. *Journal of Marine Systems*, 1(1-2):71–96, 1990.

R. D. Ray and B. C. Douglas. Experiments in reconstructing twentieth-century sea levels. *Progress in Oceanography*, 91(4):496–515, 2011.

D. Ren, S. Yang, Z. Wu, and G. Yan. Evaluation of the emd end effect and a window based method to improve emd. In *Technology and Innovation Conference, 2006. ITIC 2006. International*, pages 1568–1572. IET, 2006.

G. Richards. The fullcam carbon accounting model: Development, calibration and implementation. *Carbon accounting and emissions trading related to bioenergy, wood products and carbon sequestration*, 1, 2001.

K. Ridgway. Long-term trend and decadal variability of the southward penetration of the East Australian Current. *Geophysical Research Letters*, 34(13), 2007.

K. Ridgway and J. Dunn. Mesoscale structure of the mean East Australian Current System and its relationship with topography. *Progress in Oceanography*, 56(2): 189–222, 2003.

K. Ridgway and J. Godfrey. Mass and heat budgets in the East Australian Current: A direct approach. *Journal of Geophysical Research: Oceans*, 99(C2):3231–3248, 1994.

K. Ridgway, R. Coleman, R. Bailey, and P. Sutton. Decadal variability of East Australian Current transport inferred from repeated high-density XBT transects, a CTD survey and satellite altimetry. *Journal of Geophysical Research: Oceans (1978–2012)*, 113(C8), 2008.

E. Rignot, I. Velicogna, M. Van den Broeke, A. Monaghan, and J. Lenaerts. Acceleration of the contribution of the Greenland and Antarctic ice sheets to sea level rise. *Geophysical Research Letters*, 38(5), 2011.

J. Salisbury and M. Wimbush. Using modern time series analysis techniques to predict ENSO events from the SOI time series. *Nonlinear Processes in Geophysics*, 9(3/4):341–345, 2002.

A. H. Sallenger Jr, K. S. Doran, and P. A. Howd. Hotspot of accelerated sea-level rise on the Atlantic coast of North America. *Nature Climate Change*, 2(12):884–888, 2012.

S. Saramul and T. Ezer. Spatial variations of sea level along the coast of Thailand: Impacts of extreme land subsidence, earthquakes and the seasonal monsoon. *Global and Planetary Change*, 122:70–81, 2014.

N. Scafetta. Multi-scale dynamical analysis (MSDA) of sea level records versus PDO, AMO, and NAO indexes. *Climate Dynamics*, 43(1-2):175–192, 2014.

T. Schlurmann. Spectral analysis of nonlinear water waves based on the Hilbert-Huang transformation. *Transactions–American Society of Mechanical Engineers Journal of Offshore Mechanics and Arctic Engineering*, 124(1):22–27, 2002.

J. H. Scott and E. D. Reinhardt. Assessing crown fire potential by linking models of surface and crown fire behavior. 2001.

J. Sharples, R. McRae, R. Weber, and A. M. Gill. A simple index for assessing fuel moisture content. *Environmental Modelling & Software*, 24(5):637–646, 2009.

J. J. Sharples. An overview of mountain meteorological effects relevant to fire behaviour and bushfire risk. *International Journal of Wildland Fire*, 18(7):737–754, 2009.

J. J. Sharples, R. H. McRae, and S. R. Wilkes. Wind terrain effects on the propagation of wildfires in rugged terrain: fire channelling. *International Journal of Wildland Fire*, 21(3):282–296, 2012.

T. M. Smith and R. W. Reynolds. A global merged land-air-sea surface temperature reconstruction based on historical observations (1880-1997). *Journal of Climate*, 18(12):2021–2036, 2005.

T. M. Smith, R. W. Reynolds, T. C. Peterson, and J. Lawrimore. Improvements to NOAA's historical merged land-ocean surface temperature analysis (1880-2006). *Journal of Climate*, 21(10):2283–2296, 2008.

R. Sneeuwjagt and G. Peet. *Forest Fire Behaviour Tables for Western Australia*. WA Department Conservation & Land Management, 1985.

J. Sniderman and S. G. Haberle. Fire and vegetation change during the Early Pleistocene in southeastern Australia. *Journal of Quaternary Science*, 27(3):307–317, 2012.

S. Solomon. *Climate change 2007 the Physical Science Basis: Working group I contribution to the fourth assessment report of the IPCC*, volume 4. Cambridge University Press, 2007.

H. Späth. *One Dimensional Spline Interpolation Algorithms*. AK Peters/CRC Press, 1995.

R. W. Spencer. Precise monitoring of global temperature trends. *Science*, 247: 1558–1558, 1990.

C. Sun, M. Feng, R. J. Matear, M. A. Chamberlain, P. Craig, K. R. Ridgway, and A. Schiller. Marine downscaling of a future climate scenario for Australian boundary currents. *Journal of Climate*, 25(8):2947–2962, 2012.

I. M. Suthers, J. W. Young, M. E. Baird, M. Roughan, J. D. Everett, G. B. Brassington, M. Byrne, S. A. Condie, J. R. Hartog, and C. S. Hassler. The strengthening East Australian Current, its eddies and biological effects–an introduction and overview, 2011.

T. Thayaparan and A. Yasotharan. Limitations and Strengths of the Fourier Transform Method to Detect Accelerating Targets. Technical report, DTIC Document, 2000.

D. W. Thompson, J. M. Wallace, P. D. Jones, and J. J. Kennedy. Identifying signatures of natural climate variability in time series of global-mean surface temperature: Methodology and insights. *Journal of Climate*, 22(22):6120–6141, 2009.

A. Timmermann, S. McGregor, and F.-F. Jin. Wind effects on past and future regional Sea level trends in the Southern Indo-Pacific. *Journal of Climate*, 23(16): 4429–4437, 2010.

K. Tolhurst, W. Street, and V. Creswick. Report on fire danger ratings and public warning. *Victorian Bushfires Royal Commission, Melbourne, Victoria, Australia*, 2010.

M. E. Torres, M. Colominas, G. Schlotthauer, P. Flandrin, et al. A complete ensemble empirical mode decomposition with adaptive noise. In *IEEE International*

Conference on Acoustics, Speech and Signal Processing (ICASSP), 2011, pages 4144–4147. IEEE, 2011.

K. E. Trenberth. Changes in precipitation with climate change. *Climate Research*, 47(1-2):123–138, 2011a.

K. E. Trenberth. Attribution of climate variations and trends to human influences and natural variability. *Wiley Interdisciplinary Reviews: Climate Change*, 2(6): 925–930, 2011b.

K. E. Trenberth and J. T. Fasullo. An apparent hiatus in global warming? *Earth's Future*, 1(1):19–32, 2013.

K. E. Trenberth, A. Dai, G. Van Der Schrier, P. D. Jones, J. Barichivich, K. R. Briffa, and J. Sheffield. Global warming and changes in drought. *Nature Climate Change*, 4(1):17–22, 2014.

A. Troccoli, K. Muller, P. Coppin, R. Davy, C. Russell, and A. L. Hirsch. Long-term wind speed trends over Australia. *Journal of Climate*, 25(1):170–183, 2012.

C. Van Wagner, P. Forest, et al. Development and Structure of the Canadian Forest FireWeather Index System. In *Can. For. Serv., Forestry Tech. Rep.* Citeseer, 1987.

A. Veltcheva and C. G. Soares. Identification of the components of wave spectra by the Hilbert Huang transform method. *Applied Ocean Research*, 26(1):1–12, 2004.

D. C. Verdon, A. S. Kiem, and S. W. Franks. Multi-decadal variability of forest fire riskeastern australia. *International Journal of Wildland Fire*, 13(2):165–171, 2004.

M. Vermeer and S. Rahmstorf. Global sea level linked to global temperature. *Proceedings of the National Academy of Sciences*, 106(51):21527–21532, 2009.

D. Viegas. Innovations and solutions in fire behaviour prediction issues. In *Teoksessa: Proceedings of the International Symposium Forest Fires: Needs & Innovations Athens, Greece*, pages 18–19, 1999.

C. L. Vincent, G. Giebel, P. Pinson, and H. Madsen. Resolving nonstationary spectral information in wind speed time series using the Hilbert-Huang transform. *Journal of Applied Meteorology and Climatology*, 49(2):253–267, 2010.

N. R. Viney. A review of fine fuel moisture modelling. *International Journal of Wildland Fire*, 1(4):215–234, 1991.

H. Visser, S. Dangendorf, and A. C. Petersen. A review of trend models applied to sea level data with reference to the acceleration-deceleration debate. *Journal of Geophysical Research: Oceans*, 120(6):3873–3895, 2015.

R. S. Vose, D. R. Easterling, and B. Gleason. Maximum and minimum temperature trends for the globe: An update through 2004. *Geophysical Research Letters*, 32 (23), 2005.

G. Wahba and Y. Wang. Behavior near zero of the distribution of gcv smoothing parameter estimates. *Statistics & probability letters*, 25(2):105–111, 1995.

P. Watson. Is there evidence yet of acceleration in mean sea level rise around mainland Australia? *Journal of Coastal Research*, 27(2):368–377, 2011.

M. Wenzel and J. Schröter. Reconstruction of regional mean sea level anomalies from tide gauges using neural networks. *Journal of Geophysical Research: Oceans*, 115 (8), 2010.

M. Wenzel and J. Schröter. Global and regional sea level change during the 20th century. *Journal of Geophysical Research: Oceans*, 119(11):7493–7508, 2014.

N. J. White, J. A. Church, and J. M. Gregory. Coastal and global averaged sea level rise for 1950 to 2000. *Geophysical Research Letters*, 32(1), 2005.

N. J. White, I. D. Haigh, J. A. Church, T. Koen, C. S. Watson, T. R. Pritchard, P. J. Watson, R. J. Burgette, K. L. McInnes, Z.-J. You, et al. Australian sea levelstrends, regional variability and influencing factors. *Earth-Science Reviews*, 136:155–174, 2014.

A. A. Williams, D. J. Karoly, and N. Tapper. The sensitivity of Australian fire danger to climate change. *Climatic Change*, 49(1):171–191, 2001.

P. Woodworth, N. J. White, S. Jevrejeva, S. Holgate, J. Church, and W. Gehrels. Evidence for the accelerations of sea level on multi-decade and century timescales. *International Journal of Climatology*, 29(6):777–789, 2009.

P. L. Woodworth, M. Menéndez, and W. R. Gehrels. Evidence for century-timescale acceleration in mean sea levels and for recent changes in extreme sea levels. *Surveys in geophysics*, 32(4-5):603–618, 2011.

F. Wu and L. Qu. 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(3):586–602, 2008.

Z. Wu and N. E. Huang. A study of the characteristics of white noise using the Empirical Mode Decomposition Method. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, volume 460, pages 1597–1611. The Royal Society, 2004.

Z. Wu and N. E. Huang. Ensemble empirical mode decomposition: a noise-assisted data analysis method. *Advances in adaptive data analysis*, 1(01):1–41, 2009.

Z. Wu, N. E. Huang, and X. Chen. Some considerations on physical analysis of data. *Advances in Adaptive Data Analysis*, 3(01n02):95–113, 2011.

C. Wunsch, R. M. Ponte, and P. Heimbach. Decadal trends in sea level patterns: 1993-2004. *Journal of Climate*, 20(24):5889–5911, 2007.

K. Zeng and M.-X. He. A simple boundary process technique for empirical mode decomposition. In *IEEE International Geoscience and Remote Sensing Symposium, 2004. IGARSS'04. Proceedings. 2004*, volume 6, pages 4258–4261. IEEE, 2004.

R. R. Zhang, S. Ma, and S. Hartzell. Signatures of the seismic source in EMD-based characterization of the 1994 Northridge, California, earthquake recordings. *Bulletin of the Seismological Society of America*, 93(1):501–518, 2003.

X. Zhang and J. A. Church. Sea level trends, interannual and decadal variability in the Pacific Ocean. *Geophysical Research Letters*, 39(21), 2012.

Z. Zhidong and W. Yang. A new method for processing end effect in empirical mode decomposition. In *International Conference on Communications, Circuits and Systems, 2007. ICCCAS 2007.*, pages 841–845. IEEE, 2007.