End-effect, stopping criterion, mode mixing and confidence limit for the Hilbert-Huang transform

JULIEN RÉMY DOMINIQUE GÉRARD LANDEL

NATIONAL UNIVERSITY OF SINGAPORE
2008

End-effect, stopping criterion, mode mixing and confidence limit for the Hilbert-Huang transform

JULIEN RÉMY DOMINIQUE GÉRARD LANDEL

(Eng. Deg., ÉCOLE POLYTECHNIQUE)

A THESIS SUBMITTED FOR THE DEGREE OF
MASTER OF ENGINEERING

DEPARTMENT OF MECHANICAL ENGINEERING
NATIONAL UNIVERSITY OF SINGAPORE

2008

Acknowledgments

The author would like to express his deep appreciation to his co-supervisor Professor Chew Yong Tian and Professor Lim Hock for giving the opportunity to work on this fascinating project. In particular, the author thanks them for their guidance, suggestions and recommendations throughout the project. The author also wish to thank his supervisor Associate Professor Christopher Yap for his constant support and patience during the research work.

Secondly, the author extends its gratitude to his friends Youcef Banouni and Benoit Mortgat for their thoughtful advice to improve this document.

Finally, the author expresses his love and gratitude to his parents, sister, brothers and other family members for their continuous support and encouragement throughout his study.

The author would like to acknowledge the financial support provided by the École Polytechnique.

Contents

Summary

The research reported in this thesis was undertaken from November 2007 to November 2008 at the Department of Mechanical Engineering of the National University of Singapore. This research focuses on the Hilbert-Huang transform, a new and powerful signal-processing technique, which has greater capability than all other existing methods in analysing any nonlinear and non-stationary signal. The Hilbert-Huang transform provides a time-frequency-amplitude representation of the data, which gives a very meaningful interpretation of the physical processes accounting for the phenomenon studied. Since its creation in 1998, scientists have successfully applied this method in many domains such as: biomedical applications, chemistry and chemical engineering, digital image analysis, financial applications, fluid mechanics, meteorological and atmospheric applications, ocean engineering, seismic studies, structural applications, health monitoring, and system identification.

The algorithm implementing the Hilbert-Huang transform is an empirical method with some mathematical and practical limitations. Firstly, the problem of the end-effect, which is inherent to the study of finite-length signals, can pose practical difficulties to the calculation of the envelopes of the signal, a fundamental step of the sifting process. Secondly, because of mathematical uncertainties, the sifting process has to be iterated several times before finding each mode of the signal. It becomes necessary to define at which iteration the sifting pro-

cess must be stopped. Thirdly, mode mixing can occur with a straightforward application of the algorithm. If this issue is not addressed, the results can be distorted.

After reviewing the basics of the Hilbert-Huang transform, solutions, comprising the source codes implemented in Matlab, addressing its flaws are presented under the form of control parameters of the original algorithm. Four end-point options are described: the clamped end-point option, the extrema extension technique, the mirror imaging extension method and a damped sinusoidal extension using an auto-regressive model. Then, a particular stopping criterion based on the two conditions defining an intrinsic mode function is chosen from a review of four criteria. Finally, the algorithm of an intermittency test handling the problem of mode mixing is provided. After that, a method evaluating the performances of the enhanced algorithm is described. It makes use of four indicators, from which the last three are newly introduced: the index of orthogonality, the number of IMFs, the number of iterations per IMF and the index of component separation. Next, a study of five test signals shows the abilities and the reliability of each indicator. Then, the choice of the control parameters based on a systematic study of the length-of-day data is discussed. It is found that the fourth end-point option combined with intermediate thresholds for the stopping criterion generally gives the best results. Finally, the efficiency of the intermittency test is demonstrated through the study of vortex-shedding signals. An unexpected discovery of periodical intra-wave frequency modulation with respect to the theoretical shedding frequency has been made from this analysis.

List of Tables

List of Figures

List of Hyperlinks

List of Source Codes

List of Symbols and Abbreviations

Symbols

$\mathcal{A}[x]$	defines the analytic signal of the variable x.
A	variable
a	instantaneous amplitude function
a_j	coefficient denoting the instantaneous amplitude of the j^{th} IMF or the amplitude of the j^{th} mode of the Fourier decomposition
a_{jk}	mode amplitude of the proto-IMF h_{jk}
b_1	first coefficient of the second-order auto-regressive model
b_2	second coefficient of the second-order auto-regressive model
C_i	set of control parameters for the case i, $C_i = (epo_i, \theta_{1,i}, \alpha_i)$
c_j	j^{th} IMF
$c_{j,i}$	j^{th} IMF of the i^{th} set
$c_{j,int}$	intermittent IMF of number j
$Cov(X, Y)$	designates the covariance, or auto-correlation function, of the variables X and Y.
d	diameter of a circular cylinder
dt	defines the differentiation of the variable t.

$E(X)$	designates the ensemble mean of the variable X.
$e^x, exp(x)$	denote the exponential function of the variable x.
$e_{max}[h]$	designates the upper envelope of the function h.
$e_{min}[h]$	designates the lower envelope of the function h.
F	cumulative distribution function
F_s	instantaneous vortex-shedding frequency
$F_{s,FT}$	vortex-shedding frequency given by the Fourier transform
$F_{s,HHT}$	instantaneous vortex-shedding frequency given by the HHT
$F_{s,T}$	theoretical vortex-shedding frequency
F_o	frequency of the periodically varying flow
$\mathcal{H}[x]$	defines the Hilbert transform of the variable x.
$H(\omega, t)$	Hilbert spectrum
$h(\omega)$	marginal spectrum
h_{jk}	proto-IMF at the k^{th} iteration of the j^{th} sifting process
i	integer or imaginary number
IEC	index of energy conservation
IO	index of orthogonality
ICS	instantaneous index of component separation
ICS_j	instantaneous index of component separation for the j^{th} and $(j+1)^{th}$ IMFs
J	subset of time instants with $J \subset T$
j	integer, which can designate the IMF number.
k	integer, which can designate the iteration number.
l	integer, which can designate the number of the last extremum.
L^p	the L^p class denotes the space of p-power integrable functions.
$log(x)$	denotes the decimal logarithm of the variable x.
M	maximum number of iterations for the third stopping criterion

m	integer
m_{jk}	mean of the envelopes at the k^{th} iteration of the j^{th} sifting process
$m_{j,int}$	mean of the intermittent residue of number j
$mean(X)$	designates the arithmetical mean of the discrete series X.
$min(X, Y)$	designates the minimum between the variables X and Y.
N	integer designating the size of a discrete-time series
n	integer, which can designate the total number of IMF
n_1	vector of intermittent criteria for a set of IMFs
N_{avg}	integer designating the length of averaging
N_{epl}	integer designating the length of extrapolation
N_{ext}	number of extrema
N_{IMF}	number of IMFs
$N_{ite,j}$	number of iterations of the j^{th} IMF
$N_{ite,T}$	total number of iterations, $N_{ite,T} = \sum_j N_{ite,j}$
$N_{j,k}$	denotes the number of points $H_i[t_i, \omega_i]$ belonging to the bin defined as: $t_j \leq t_i < t_{j+1}$ and $\omega_k \leq \omega_i < \omega_{k+1}$.
N_{set}	number of sets of IMFs
N_{zc}	number of zero-crossings
PV, P, CPV	Cauchy principal value
r_j	j^{th} residue
$\mathbf{R}$	set of all real numbers
$\Re[z]$	denotes the real part of the complex variable z.
Re	Reynolds number
S	optimal sifting number for the second stopping criterion
$S_i(t)$	i^{th} temporal test signal
SD	standard deviation
SD_{max}	maximum standard deviation for the first stopping criterion

$sd(C)$	squared deviation between the variable C and its mean $\overline{C}$, $sd(C) = \left(C - \overline{C}\right)^2$
T	time span or time vector
T_s	period of the vortex-shedding signal
T_{ω_i}	period of the instantaneous frequency of the i^{th} IMF
t	time
t_i	i^{th} time instant
t_{e_i}	time instant of the i^{th} extrema
$Var(X)$	designates the variance of the variable X.
$X(t)$	time-series data
x	real-valued variable or values of the discrete series X
x_{e_i}	value of the i^{th} extrema
X_{epl}	extrapolated extension of a discrete-time series
X_{shift}	shifted discrete-time series of X, $X_{shift} = X - \mu(N_{avg})$
x_{shift}	values of X_{shift}
y	real-valued variable or discrete series
z	complex-valued variable or discrete series
α	tolerance for the fourth stopping criterion
Δt	time step
$\Delta\omega$	frequency step
ε	infinitesimal variable
θ	instantaneous phase function
θ_1	first threshold for the fourth stopping criterion
θ_2	second threshold for the fourth stopping criterion
κ	damping coefficient of the second-order auto-regressive model
μ	arithmetic mean

ν	kinematic viscosity
σ	standard deviation
σ_{jk}	absolute value of the ratio of the mode amplitude to the mean of the proto-IMF h_{jk}, $\sigma_{jk} = \lvert m_{jk}/a_{jk} \rvert$
τ	variable of integration or constant
ω	instantaneous frequency function
ω_j	coefficient denoting the instantaneous frequency of the j^{th} IMF or the frequency of the j^{th} mode of the Fourier decomposition
ω_s	pulsation of the sinusoidal extension of the second-order auto-regressive model
$\overline{\cdot}$	$\overline{X}$ defines the arithmetic mean of the variable X.
$\langle \cdot, \cdot \rangle$	$\langle X, Y \rangle$ defines the scalar product of the variables X and Y.
$\#(\cdot)$	$\#(T)$ defines the cardinality, or the number of elements, of the set T.
$\lvert \cdot \rvert$	$\lvert X \rvert$ defines the absolute value of the variable X.
$\cdot * \cdot$	$f * g$ defines the convolution product of the two functions f and g.

Abbreviations

CL	confidence limit
`Dt`	time step of a time vector
EMD	empirical mode decomposition
`EMD`	EMD algorithm
`eo`	extension option
`epo`	end-point option
HHT	Hilbert-Huang transform
`HT`	Hilbert-transform algorithm
IMF	intrinsic mode function
`le`	length of extension
LOD	length-of-day
pchip	piecewise cubic Hermite interpolating polynomial
`t0`	first point of a time vector
`tN`	last point of a time vector

Main part

1. Introduction

1.1 The Hilbert-Huang transform

Analysing time-series data or signals is a very frequent task in scientific research and in practical applications. Among traditional data processing techniques, the Fourier transform is certainly the most well-known and powerful one. It has been frequently used in theoretical and practical studies since it was invented by Fourier in 1807. However, its application is limited to only linear and stationary signals, thus making it unsuitable for analysing some categories of real-world data. Then, several methods, based on joint time-frequency analysis, were developed during the last century to handle non-stationary processes and better explain local and transient variations: the windowed Fourier and Gàbor transforms, the Wigner-Ville distribution, and wavelet analysis and its derived techniques (see Cohen 1995 [10] for a detailed introduction to these techniques). Nevertheless, the main shortcoming of all these methods is their inability to study nonlinear signals, and their need of a predefined basis. Despite all the efforts of the scientific community to improve these techniques, none of them can correctly handle nonlinear and non-stationary data, which represent the most common data in real-world phenomena.

Recently, a new data-analysis method, named the Hilbert-Huang transform (HHT), has been introduced by Huang et al. (1998 and 1999) [27] [26] in order to study nonlinear and non-stationary signals. In addition, it aims at providing a physical understanding of the underlying processes represented in the signal, thus achieving the primary goal of signal processing. The HHT method proceeds in two steps: first, a signal is decomposed, following the Empirical Mode Decomposition (EMD) scheme, into Intrinsic Mode Functions (IMFs); second, the application of the Hilbert transform to each mode yields the complete time-frequency-energy representation of the signal. The algorithm actually relies on the ability of the Hilbert transform to reveal the local properties of time-series data and calculate the instantaneous frequency (Hahn 1995 [21]). However, due to theoretical limitations, a straightforward application of the Hilbert transform to the original signal would be very likely to lead to misleading results. For example, the instantaneous frequency could have negative values which is, of course, physically impossible. Therefore, the fundamental breakthrough of the HHT lies in the first step: the EMD prepares and decomposes the raw data into appropriate modes or IMFs, which can be subsequently analyzed by the Hilbert transform to eventually yield physically meaningful results.

The EMD is an empirical method based on the assumption that every signal consists of a superposition of narrow band-passed, quasi-symmetrical components. In order to retrieve these well-behaved components, the signal is decomposed by an ingenious method called the sifting process. Unlike all other techniques, the EMD has the distinctive feature of being adaptive, meaning that the decomposition depends only on the signal. There is no a priori defined basis such as the harmonics in the Fourier transform. This difference is very important because it ensures that all the information contained in the original signal are not distorted and that they can be fully recovered in the IMFs. Therefore, because of its adaptiveness and its ability to correctly analyze nonlinear and non-stationary data, the HHT proves to be the most powerful data-processing technique.

1.2 Applications of the HHT

Since the HHT was developed in 1998, many scientists and engineers have used this technique in various fields of science as well as in practical applications. In every case, the results given by the HHT are reported to be as good as or better than those obtained from other techniques such as the Fourier transform and the wavelet transform. We present here a few examples of the existing applications.

Biomedical applications: Huang et al. (1998) [30] analyzed the pulmonary blood pressure of rats with both the HHT and the classical Fourier analysis. A comparison of the results showed that the HHT could reveal more information on the blood pressure characteristics. Huang et al. (1999) [31] also studied the signals obtained from pulmonary hypertension. Their study investigated the linear and nonlinear influences of a step change of oxygen tension on the pulmonary blood pressure. Using the HHT, they found the analytic functions of both the mean blood pressure response, represented by the sum of the last IMFs, and the oscillations about the mean trend, represented by the sum of the first IMFs. Finally, from the mathematical formulations they were able to understand mechanisms related to blood pressure, which are crucial for applications in tissue remodeling of blood vessels.

Chemistry and chemical engineering: Phillips et al. (2003) [37] studied molecular dynamics simulation trajectories and conformational change in Brownian dynamics. Comparisons between HHT and wavelet analysis showed overall similar results; however, the HHT gave a better physical insight of conformational change events. Wiley et al. (2005) [50] investigated the internal motions and changes of conformations of proteins in order to understand their biological functions. Since these phenomena are wavelike in nature, they developed a technique called Reversible Digitally Filtered Molecular Dynamics to focus on low frequency motions, which correspond to large scale changes in structures.

The HHT proved to be a better tool than Fourier-based analysis to study these transient and non-stationary signals.

Financial applications: Huang et al. (2003) [29] demonstrated the usefulness of the EMD in statistical analysis of nonlinear and non-stationary financial data. They invented a new tool to quantify the volatility of the weekly mean of the mortgage rate over a thirty-year period. This tool, named the variability, was based on the ratio of the absolute value of the IMF to the signal. It offers a simple, direct and time-dependent measure of the market volatility, which proves to be more realistic than traditional methods based on standard deviation measurements.

Fluid Mechanics: Zeris and Prinos (2005) [55] performed a comparative analysis between wavelet transforms and the HHT in the domain of turbulent open channel flow. They managed to identify and study near wall characteristic coherent structures. They concluded that the HHT method should be prefered to the wavelet technique in any investigation on non-stationary flows because it gives more accurate results in joint time-frequency analysis, while the wavelet transform is strongly affected by smear effects. Hu et al. (2002) [22] conducted an experimental study of the instantaneous vortex-shedding frequency (F_s) in periodically varying flow (of frequency F_o). Using the HHT to decompose the streamwise velocity signal in the wake of a stationary T-shaped cylinder, they found three different regimes depending on the ratio F_s/F_o. Firstly, for $F_s/F_o > 4.37$, the variations of the instantaneous vortex-shedding frequency are correlated to the variations of the incoming flow without phase lag. Secondly, for $1.56 < F_s/F_o < 4.37$, the same behaviour is observed but with a phase lag linearly related to the frequency ratio. Furthermore, they observed a hysteresis vortex-shedding behaviour. Thirdly, for $0.29 < F_s/F_o < 1.56$, they found no interactions between F_s and F_o. Moreover, this regime features two occurrences of lock-on at $F_s/F_o \approx 1$ and $F_s/F_o \approx 0.5$.

Image analysis: Long (2005) [33] showed that it was possible to use the HHT in image analysis because rows and columns can be seen as discrete-space series. The study of inverse wavelengths and energy values as functions of time or distance for the case of water-wave images made possible the measurement of characteristic features. In conclusion, the author emphasized the great perspectives offered by the HHT in the domain of image processing. Nunes et al. (2005) [36] went further by applying the HHT to 2D data such as images. They developed a bidimensional version of the EMD and replaced the Hilbert transform by the Riesz transform, which can be applied on multidimensional signals. Finally, they demonstrated that their enhanced version of the HHT was efficient to detect texture changes in both synthetic and natural images. Later, Damerval et al. (2005) [12] improved the bidimensional EMD by using Delaunay triangulation and piecewise cubic interpolation. They showed, through an application on white noise, that their improvements on the algorithm significantly increased the computational speed of the sifting process.

Noise detection: After discovering that the EMD behaved like a dyadic filter, Flandrin et al. (2005) [16] suggested its use to denoise-detrend signals containing noise. Coughlin and Tung (2005) [11] also demonstrated how noise could be identified in atmospheric signals. They defined a statistical test of confidence to discriminate noise from the signals according to their respective energy spectra. Huang et al. (1998) [27] showed that the EMD could serve as a filtering tool by simply retaining and summing the IMFs of desired bandwidth.

Ocean engineering: Many studies have been conducted in this domain after Huang et al. (1999) [26] analysed nonlinear water waves in 1999. For example, Schlurmann and Dätig (2005) [46] were interested in rogue waves. Understanding how they are generated is very important for designing offshore structures and ships that will not be damaged by these waves. Yan et al. (2005) [54] also showed that the HHT could help assessing the health of marine eco-systems by

analyzing ocean color data.

Structural engineering applications: Salvino et al. (2005) [44] used the HHT as a means to identify internal mechanical failures of structures. The locations of failure were identified by analyzing the instantaneous phase of structural waves. A similar application was developed by Huang et al. (2005) [25] to diagnose the health of bridges. The HHT analyzed responses of vibration-tests, and two criteria based on the instantaneous frequency were defined to assess the state of the structure.

Needless to say, this list is not exhaustive, and other interesting applications can be found in Attoh-Okine (2005) [1]. Finally, since its creation a decade ago, the HHT has been developed in various applications with successful results, indicating the great potential for this novel data-processing technique.

1.3 Objectives of the study

The main objective of this study is to serve as a guide for understanding, implementing and using the Hilbert-Huang transform. Explanations about the underlying motivations of the development of the HHT, i.e. how to retrieve the instantaneous frequency, are given along with details about the algorithm. The main flaws of the algorithm, namely the end-effect, the stopping criterion and the mode mixing phenomenon, are thoroughly discussed. Then, different solutions to these limitations are proposed under the form of control parameters in the algorithm. Finally, these control parameters are tested with different signals. Meanwhile, four quantitative indexes, which aim at assessing the results of the HHT, are presented and it is shown how they can help finding the most adapted control parameters for the study of a signal with the HHT algorithm. Precisely, the purpose of the present work is to ease some of the tedious and lengthy tasks that users of the HHT could encounter during the implementation or the appli-

cation of this technique, which, however, deserves to be considered as the first and most powerful method to analyse real-world phenomena.

Chapter 2 begins with the description of the empirical mode decomposition and how the Hilbert transform can retrieve the instantaneous frequency and amplitude from the intrinsic mode functions. Then, a literature review of the critical points of the HHT is conducted. The fundamental concept of instantaneous frequency is reviewed. The main flaws of the algorithm are described and the concept of confidence limit for the HHT is presented. Finally, the implementation of the HHT algorithm is detailed. Firstly, the EMD and the Hilbert-transform algorithms are introduced. Secondly, four end-point options handling the problem of end-effect as well as an efficient stopping criterion for the sifting process are described. Thirdly, the implementation of the intermittency test, a necessary test to prevent mode mixing, is given. Fourthly, four quantitative indexes evaluating the decomposition and the Hilbert spectrum are introduced. Fifthly, the algorithm to calculate the confidence limit for the HHT is provided. All the source codes, implemented in Matlab, of the HHT algorithm and its control parameters can be found in Appendix B.

Chapter 3 presents three studies of computed and experimental signals performed with the HHT. The first study shows the behaviour of the HHT algorithm with five simple test signals. The influence of each control parameter on the results is assessed by the quantitative indexes. Second, a systematic study of the end-point options and of the stopping criterion is conducted with the length-of-day data. Third, the phenomenon of vortex-shedding is investigated to show how the HHT algorithm can be successfully used to interpret a physical nonlinear phenomenon.

2. HHT algorithm

2.1 Basics of the HHT

2.1.1 Empirical mode decomposition

As Huang et al. (1998) [27] explained, the empirical mode decomposition method is an empirical sifting process aiming at decomposing any nonlinear and non-stationary signal into a set of IMF components. In order to have well-behaved Hilbert transforms of the IMFs, i.e. a meaningful instantaneous frequency, the components must have the following characteristics: firstly, they must have a unique time scale; secondly, they must be quasi-symmetric. The characteristic time scale is determined with the distance between successive extrema. Therefore, an IMF can be defined as follows:

1. Its number of extrema and zero-crossings must be equal or differ at most by one.

2. At any point, the mean value of its envelopes defined by the local maxima and the local minima should be zero.

The sifting process, which reveals the intrinsic oscillations of a time-series data, $X(t)$, has been described by Huang et al. (1998) [27] as "intuitive , direct, a pos-

teriori and adaptive, with the basis of the decomposition based on and derived from the data". However, since it is a very recent method, its whole mathematical validation has yet to be proved, the mathematical issues related with the HHT will be discussed in Section 2.2.

The first step of the sifting process is to identify the local extrema of the signal, then the upper and lower envelopes are calculated as the cubic spline interpolations of the local maxima and minima respectively. Next, the first component h_1, designated as the first proto-IMF, is the difference between the data and the mean of the envelopes m_1:

$$X(t) - m_1 = h_1. \tag{2.1}$$

Figure 2.1 illustrates these steps. h_1 should ideally represent the first IMF. However, due to several mathematical approximations in the sifting process, this first proto-IMF may not exactly satisfy the two conditions of IMF. Since neither a mathematical definition of an envelope nor a mathematical definition of the mean exist, the use of cubic spline interpolations can lead to some imperfections. For example, an inflexion point or a riding wave in the original data, which certainly has a physical meaning and represents the finest time-scale, may not be correctly sifted and new local extrema can appear after subtracting the mean from the signal. In addition, the mean may not be exactly zero at the end of the first step. Therefore, to eliminate riding waves and to make the profile more symmetric, the sifting process must be repeated several times, using the resulting proto-IMF as the data in the following iteration. Finally, k iterations may be necessary to get the first IMF h_{1k}

$$h_{1(k-1)} - m_{1k} = h_{1k}. \tag{2.2}$$

The first IMF of the test data is displayed on Figure 2.2; it has been obtained after 40 sifting iterations and it shows the finest scale of the signal. Then, it is

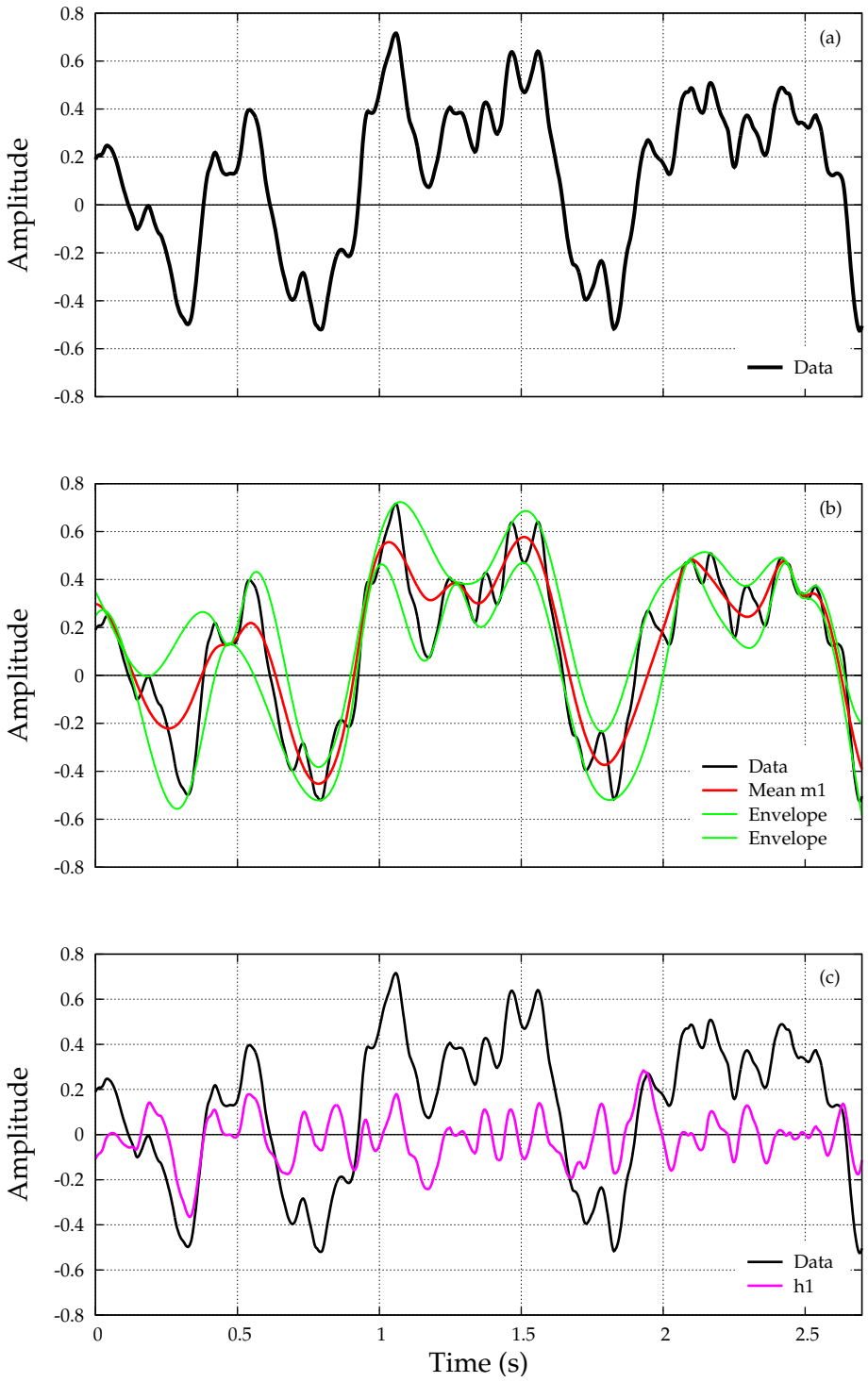

Figure 2.1: Illustration of the sifting process: (a) test data (blue); (b) test data, upper and lower envelopes (green), and mean m_1 (red); (c) test data and first proto-IMF h_1 (pink). We can see on Figure (c) that the inflexion point at $t = 1.7$ s in the data has become a new oscillation in h_1, which is not symmetric. Therefore, the sifting process must be iterated to eliminate this kind of imperfection.

recorded as:

$$h_{1k} = c_1. \tag{2.3}$$

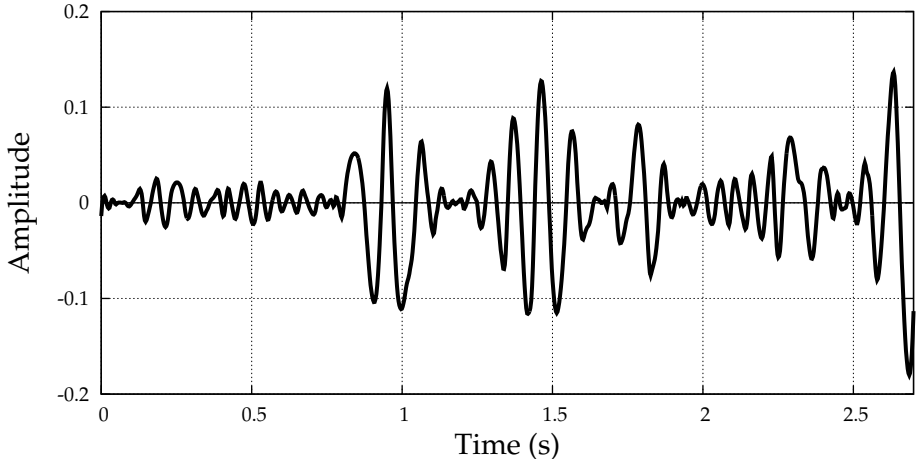

Figure 2.2: The first IMF component c_1 of the test data, after 40 iterations.

The stoppage of the sifting process can be difficult to determine in practice. Although the first condition can be easily implemented, a clear definition of the second one is somewhat cumbersome since converging toward a zero numerical mean is almost impossible. Consequently, a stopping criterion must be adapted to determine the degree of approximation for the implementation of the second condition. Four different stopping criteria are introduced and discussed in Section 2.2.5. This criterion is a critical point because it must ensure that the signal has been sufficiently sifted so that all the hidden oscillations have been retrieved; on the other hand, too many iterations can flatten the wave amplitude, thus affecting the original physical sense.

Once the first IMF c_1 has been obtained, the sifting process is repeated with the first residue r_1 resulting from the difference between c_1 and the signal:

$$X(t) - c_1 = r_1. \tag{2.4}$$

Finally, the last IMF c_n, after n sifting processes, is reached when the last residue r_n has either a too low amplitude or becomes a monotonic function. It can be remarked that the frequency range of the successive IMFs decreases with increas-

ing IMF number. Indeed, the first IMFs capture the finest scales of the signal while the subsequent residues keep only the oscillations of larger time scales. In addition, the choice to base the time scale on the distance between successive extrema has the non-negligible benefit of requiring no zero reference. For example, in the case of a signal with a non-zero trend, this trend will eventually be recovered in the last residue. Finally, the original signal is:

$$X(t) = \sum_{i=1}^{n} c_i + r_n. \tag{2.5}$$

Therefore, the signal has been decomposed into n modes or IMFs and one residue r_n. Now, the Hilbert transform can be applied to these modes since they all possess the adequate characteristics: they contain a single time scale, and their wave-profile is symmetric.

2.1.2 Hilbert spectral analysis

Hilbert transform

The second phase of the HHT consists of applying the Hilbert transform to all the IMFs in order to determine their instantaneous frequency as well as their instantaneous amplitude. Though the EMD has already given meaningful information about the data by showing the time evolution of its intrinsic modes, the Hilbert transform can reveal the frequency and the amplitude of each IMF and at each time instant. This is a step further in understanding the physical mechanisms represented in the original signal.

The Hilbert transform (see Appendix A.2) of an IMF $c(t)$ is simply the principal value (PV) of its convolution with $1/t$:

$$\mathcal{H}[c(t)] = \frac{1}{\pi} PV \int_{-\infty}^{\infty} \frac{c(\tau)}{(t-\tau)} d\tau. \tag{2.6}$$

Then, we can deduce the analytic signal of $c(t)$:

$$\mathcal{A}[c(t)] = c(t) + i\mathcal{H}[c(t)] = a(t)e^{i\theta(t)}, \tag{2.7}$$

with a the instantaneous amplitude and θ the phase function defined as:

$$a(t) = \sqrt{c^2(t) + \big(\mathcal{H}[c(t)]\big)^2}, \quad \text{and} \quad \theta(t) = \arctan\left(\frac{\mathcal{H}[c(t)]}{c(t)}\right), \tag{2.8}$$

hence we can immediately compute the instantaneous frequency

$$\omega(t) = \frac{d\theta(t)}{dt}. \tag{2.9}$$

The Hilbert transform can be applied to each IMF component so that the original data can be expressed in the following form:

$$X(t) = \Re\left[\sum_{j=1}^{n} a_j(t)\exp\left(i\int \omega_j(t)dt\right)\right] + r_n, \tag{2.10}$$

where $\Re$ denotes the real part. The last residue r_n has been left on purpose because its frequency is infinite. The Fourier representation of the same signal would be

$$X(t) = \Re\left[\sum_{j=1}^{n} a_j e^{i\omega_j t}\right], \tag{2.11}$$

with a_j and ω_j constant. Therefore, comparing Equation (2.10) with Equation (2.11), the HHT can be seen as a generalization of the Fourier transform. This form accounts for the ability of the HHT to handle nonlinear and non-stationary signals.

Hilbert spectrum and marginal spectrum

The expansion (2.10) of the signal can yield a very meaningful time-frequency-amplitude distribution, or a time-frequency-energy distribution (where the energy is the square of the amplitude) if prefered. This representation is desig-

nated as the Hilbert spectrum $H(\omega, t)$. Basically, $H(\omega, t)$ is formed by the data points $(t, w_j(t), a_j(t))$, directly obtained from (2.10), for all t and for $1 \leq j \leq n$. As an example, the Hilbert spectrum of the test data has been plotted on Figure 2.3 in its three-dimensional form, and on Figure 2.4 in its two-dimensional form and with the amplitude based on a color scale. In this example, the discrete Hilbert transform was applied to each IMF using the embedded function '`hilbert`' of Matlab. This function provides with the instantaneous amplitude and the instantaneous phase. To obtain the instantaneous frequency, the discrete derivation described in Equation (2.20) is used. A smoothed Hilbert spectrum can also be plotted to obtain a more qualitative representation; however, the original Hilbert spectrum is more accurate. Then, the integration over time of

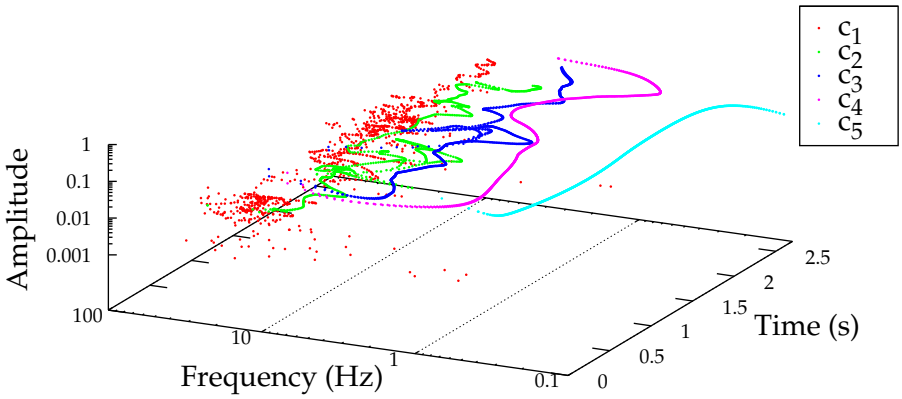

Figure 2.3: 3D Hilbert spectrum of the test data. Each point represents a given array $(t, w_j(t), a_j(t))$ for t and j fixed. Each color corresponds to a specific IMF (i.e. a given j).

the Hilbert spectrum can be calculated. It yields the marginal spectrum $h(\omega)$:

$$h(\omega) = \int_0^T H(\omega, t)dt. \tag{2.12}$$

As an example, the marginal spectrum of the data has been plotted on Figure 2.5.

Although it is possible to compare the Fourier spectrum with the marginal spectrum, there is a fundamental difference between the two representations.

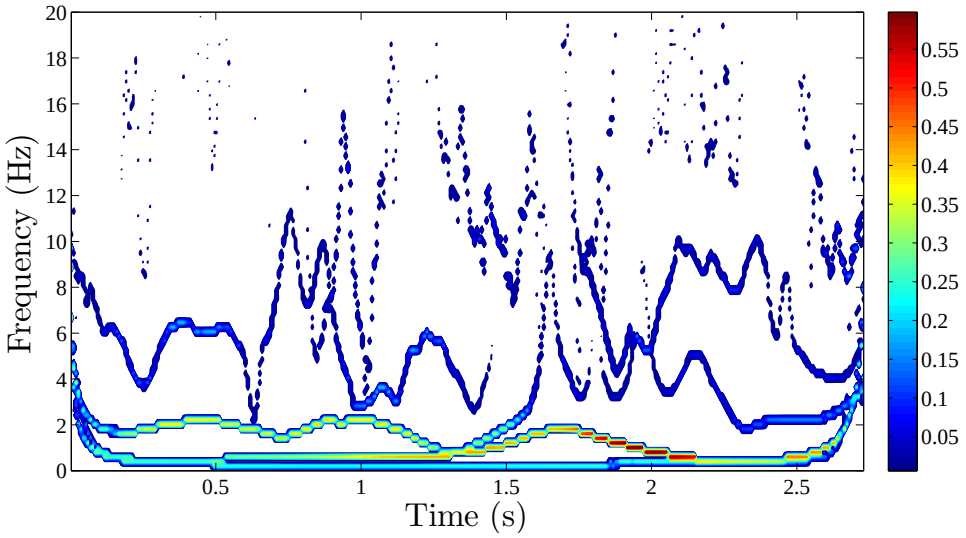

Figure 2.4: 2D Hilbert spectrum of the test data. The color scale corresponds to the instantaneous amplitude.

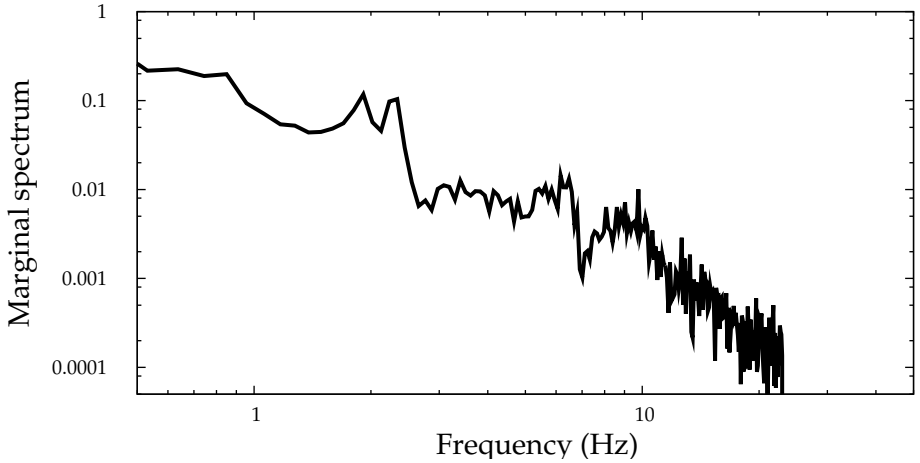

Figure 2.5: Marginal spectrum of the test data.

If a certain frequency has a high energy, in the Fourier spectrum it means that there is the corresponding harmonic (a sinusoidal wave) with a high amplitude over the whole time span. On the other hand, in the marginal spectrum it means that, over the time span, local oscillations with this frequency occur more often. Finally, because the Hilbert spectrum can give time information, it should be prefered to the marginal spectrum. In particular, the marginal spectrum cannot be used in the case of non-stationary data because it would fail to describe the instantaneous and transient characteristics of the signal.

2.2 Literature review

2.2.1 Meaningful instantaneous frequency

The concept of instantaneous frequency is essential in the Hilbert-Huang transform. Indeed, the key motivation behind this new data analysis technique stems from generations of scientists who have sought to grasp, not only the mathematical meaning, but also the physical essence of this concept. Since the works of Fourier and Hilbert, many researchers have attempted to develop joint time-frequency analysis. The main reason why so many mathematicians and physicists have continuously striven for a good definition of this concept is simple: if the time evolution of physical phenomena is of prime importance, the knowledge of its frequency is also necessary for their complete understanding. Although the Fourier transform is the first great tool which puts forward the idea of time-frequency duality, it actually fails to predict the evolution in time of the frequency. Indeed, the Fourier spectrum can show us the energy distribution of a signal in the frequency domain, but it cannot give the precise timing at which each frequency appears. Yet, this information is crucial to study accurately non-stationary [1] and transient phenomena. From our own experience, we know that

[1] A definition of stationarity can be found in Appendix A.1. Briefly, a time series is stationary if its mean, variance and autocorrelation function do not change over time (see, for example, Brockwell and Davis 1996 [7]).

most of the natural processes are rarely stationary, hence the need for a means of handling this kind of signals. For instance, in daily life experiences, music demonstrates the importance of frequency variation: melodies are based upon the variations of pitch, or frequency, of the sound produced by the instruments. In this example, the knowledge of which frequency appears in the melody, as the Fourier spectrum could give, is not very useful. That is why in many situations, and not only in signal processing, we want to know precisely the timing of each frequency. For this purpose, techniques such as the short-time Fourier transform (see, for example, Cohen 1995 [10], Prasad and Iyengar 1997 [40], or Gàbor 1946 [18]), the Wigner-Ville distribution (see, for example, Boashash (1992) [5], Mecklenbräuker and Hlawatsch (1997) [34], or Cohen 1995 [10]), and the wavelet analysis (Prasad and Iyengar 1997 [40] or Daubechies 1992 [13]) have been developed [2]. However, they suffer from either poor time resolution or poor frequency resolution. For instance, the principle of the short-time Fourier transform is to decrease the width of the window in order to focus on local variations of the frequency; however, doing so results in the broadening of the frequency bandwidth, thus worsening the frequency resolution. This inherent limitation is known as the uncertainty principle (see, for example, Skolnik 2001 [48] or Prasad and Iyengar 1997 [40]); it has first been derived by Heisenberg in 1927 while he was studying the nascent quantum mechanics. As Cohen (1995) [10] explains, the uncertainty principle states that "the densities of time and frequency cannot both be made narrow" arbitrarily.

However, Huang et al. (1998) [27] and Cohen (1995) [10] remark that the problem in the calculus of the frequency may actually stem from the method itself. Indeed, it seems paradoxical that to estimate the local frequency, one must perform an integration over the whole time domain. Thus, a new method, different from any existing technique, should be found. So, attempting to give a new definition, Cohen suggests calculating the frequency as the derivative of

[2] A more exhaustive overview of data-analysis techniques can be found in Huang et al. (1998) [27].

the phase of the signal. But then, the problem is to retrieve the phase, and a first method, namely the quadratic model, proved to be difficult to apply in most cases. Hopefully, a second solution, derived by Gàbor in 1946 [18] from the concept of analytic signal and using the Hilbert transform [3], eases this issue. However, as Cohen 1995 [10] further explains this definition is not yet perfect because many paradoxes can arise. For example, the instantaneous frequency can have negative values although the spectrum of the analytic signal is, by definition, equal to zero for negative frequencies. In fact, a good definition cannot be simply mathematical, but it must also ensure that it is physically meaningful.

In the research for a correct definition of the instantaneous frequency, the work of Huang et al. (1998) [27] has been definitive. They explain that, contrary to what Hahn claims, the Hilbert transform cannot be directly applied to any time series. A straightforward application can actually lead to the following problems for the phase function: firstly, it may not be differentiable; secondly, it can have unbounded derivatives; and thirdly, it can lead to non-physical results (such as negative derivatives). Furthermore, Shen et al. (2005) [47] state that, in order to retrieve a physically meaningful instantaneous frequency after applying the Hilbert transform, the signal has to be in a self-coherent form. In other words, it should be quasi-periodic and quasi-symmetric (or quasi-monotone)—these properties actually corresponds to the conditions of IMF described in Section 2.1.1. Another way to verify whether an analytic signal is self-coherent is to study its representation in the complex space. In polar coordinates, the instantaneous amplitude and the instantaneous frequency are represented by the radius of rotation and the time evolution of the phase angle. Salvino et al. (2005) [44] report that, to be self-coherent, a signal must have "a definite evolving direction (e.g., either clockwise or counterclockwise) and a unique center or rotation at any time" in the complex space. Actually, if a system did not follow these con-

[3] The mathematical formulations of the Hilbert transform, the analytic signal and the derivation of the instantaneous frequency are detailed in Appendix A.2, and further explanations can be found in Hahn (1995) [21].

ditions, then there would be infinite ways of describing its time evolution, and the instantaneous frequency would have the problems mentioned previously. Huang (2005a) [23] has very well illustrated this problem with the case of a simple sine wave; he has noticed that the addition of a constant to this function can influence its Hilbert transform so that the instantaneous frequency is eventually affected. Moreover, when this constant is superior than the amplitude of the signal, the results can even yield negative frequencies. However, we can intuitively understand that a change in the trend should, by no means, affect the frequency of the signal. On the other hand, he has showed that when the signal is self-coherent, the instantaneous frequency is always meaningful. Therefore, this condition seems to be a requirement before applying the Hilbert transform.

Although the idea of Huang et al. (1998) [27] to define the instantaneous frequency only for self-coherent signals is reasonable from a physical point of view, a proper theoretical definition is still an unsettled question. However, this hypothesis has been the basis of the Hilbert-Huang transform. Likewise the Fourier theory invented in 1807 but not fully proved until 1933 by Plancherel (1933) [39], it may need some years before achieving the complete and rigorous mathematical proof of the HHT. Then, assuming that the instantaneous frequency could not be directly retrieved from the signal, Huang et al. (1998) [27] invented the EMD method which precisely decomposes the signal into a set of self-coherent components. The key idea behind this approach is the concept of multicomponentness described by Cohen (1995) [10]. First, he explains that a monocomponent signal is a signal with a unique and well-defined instantaneous frequency (derived from the phase function of the analytic signal). Then, by generalization, he defines a multicomponent signal as the sum of monocomponent signals whose instantaneous bandwidth are well separated. Finally, we can see that the HHT expansion, presented in Equation (2.10) and rewritten hereafter, has effectively achieved this goal: to retrieve all the monocomponent signals

entangled in a single signal.

$$X(t) = \Re\left[\sum_{j=1}^{n} a_j(t) \exp\left(i \int \omega_j(t) dt\right)\right] + r_n, \tag{2.13}$$

Likewise it seems natural that common phenomena are seldom stationary or linear, it seems plausible that real-world signals can mingle various processes at the same time. Furthermore, it is very unlikely that these intrinsic components can be decomposed on a predefined basis, hence the importance of the HHT to be adaptive. In conclusion, the HHT, which starts by retrieving the mono-components and then calculates the instantaneous frequency and amplitude of a signal, is a powerful method revealing the underlying physical mechanisms contained in any phenomenon.

2.2.2 Completeness and orthogonality

Completeness: As Huang et al. (1998) [27] explain, the completeness of the decomposition is automatically satisfied according to Equation (2.5). Furthermore, they report that numerical tests conducted with different data sets confirm this property of the EMD. In fact, the difference between the sum of all the IMFs, including the last residue, and the signal is found to be inferior than the roundoff error of the computer.

Orthogonality: According to Huang et al. (1998) [27], the decomposition procedure should ensure the local orthogonality of the IMFs. From Equations (2.1) to (2.3) we can see that an IMF is obtained from the difference between the signal $X(t)$ and its mean $\overline{X(t)}$, hence

$$\overline{\langle (X(t) - \overline{X(t)}), \overline{X(t)} \rangle} = 0, \tag{2.14}$$

in which $\langle \ . \ , \ . \ \rangle$ designates the scalar product. However, this equation is not exact because first, the mean is not the true mean since it is calculated from

computed cubic spline envelopes; and second, an IMF does not entirely correspond to $\overline{X(t)}$ since several sifting iterations are often needed. But Huang et al. (1998) [27] further report that the leakage is often very small in practice: around 1% in most cases, and inferior than 5% for very short data. Finally, they add that othogonality should not be a requirement for nonlinear decompositions because it is not physically sensical.

In this study, the orthogonality between the IMFs will be used as a means to assess the quality of the decomposition. Moreover, an index of orthogonality will be presented in Section 2.3.6 to quantify the overall orthogonality of the EMD.

2.2.3 Mean and envelopes

The calculation of the mean of a signal is another crucial issue in the HHT. As can be seen in the decomposition process presented in Section 2.1.1, it is a key phase in the sifting process; nevertheless, a mathematical definition of the mean of a signal does not exist. So, Huang et al. (1998) [27] originally suggested to identify it as the average of the upper and lower envelopes. But this hypothesis does not truly resolve the problem of the mean because, as Riemenschneider et al. (2005) [41] underlined, "a good mathematical description of envelopes remains an unsolved issue". However, different practical solutions, as regards the envelopes, have been investigated: low- and high-order polynomial interpolations have been tested. Finally, Huang (2005b) [24] concluded that cubic spline interpolations offered the best solution because they did not require too much computation processing, and they needed very few predetermined parameters (only two extrapolated points at the edges of the signal), thus preserving the adaptive character of the EMD.

In this study, different polynomial interpolations have been tested, for example linear interpolations and the so-called *pchip* interpolation; however, none produced as good results as the cubic spline. Therefore, the solution of two cubic

spline envelopes to calculate the mean has been adopted.

2.2.4 End-effect

End-effect is a common issue in data-processing of finite-length signals. In the HHT, it occurs in the sifting process for the calculation of the cubic spline interpolations, and then, in the application of the Hilbert transform to the IMFs. In the first case, the problem is to terminate the cubic spline interpolations at the edges of the signal. Actually, if the ends of the envelopes were left unconstrained, the resulting IMFs would display large swings with spurious energy levels at their ends. Therefore, a solution must be adopted to extend the data and terminate the envelopes, so that the propagated error is minimized. Various solutions have been presented in the literature and Shen et al. (2005) [47] categorize them as:

- signal extension approaches with or without damping;
- and extrema extension techniques. These methods require two predicted extrema at both ends in the case of cubic spline envelopes.

In addition, we must keep in mind that the issue of forecasting time series can be particularly difficult for non-stationary data since they are unpredictable by essence.

A first solution, stated by Duffy (2005) [15], consists of extending the signal with sinusoidal curves of the size of the signal. Coughlin and Tung (2005) [11] also used this method, but they added only two or three oscillations in order to flatten the envelopes; they reported that longer extensions could affect low-frequency IMFs. In these two studies, the authors also reported that they did not seek for more complicated techniques since these sine extensions allowed sufficiently good qualitative results. Hwang et al. (2005) [32] chose a mirror imaging extension (possibly including windowing with an exponential decay) method: 30% of the signal was mirrored beyond the end-points. This solution

raises the question of the length of the extension, the authors noticed that one third of the data length gave the best results. However, we can wonder whether this solution can be effective for every signal. Another interesting approach of signal extension was adopted by Pinzón et al. (2005) [38], who extended their signals with similar experimental data without trends. This solution shows that a strong knowledge of the phenomenon can actually be very useful to predict more data points.

A simple and effective method of extrema extension was described by Shen et al. (2005) [47]. Compared to the previous techniques, the addition of only two extrema can be very useful because it consumes very few computation resources.

Finally, all these solutions can greatly alleviate end problems for periodic or quasi-periodic signals; however, they may not be as effective for non-stationary and transient signals. In this regard, Cheng et al. (2007) [9] performed a comparative analysis between three sophisticated forecasting techniques. A study of nonlinear and non-stationary data with intermittent signals showed that a method based on support vector regression machines was superior than a technique based on neural networks as well as an auto-regressive model. In particular, the first method was less time consuming and had usually smaller experimental errors. Moreover, it needed much less a priori knowledge of the phenomenon than the second forecasting technique, which required several control parameters.

In this study, four techniques have been tested and compared: a clamped end-point option, a mirror imaging technique, an extrema extension approach and an auto-regressive model. More details of their implementation are presented in Section 2.3.3.

2.2.5 Stopping criteria for the sifting process

Basically, the purpose of the stopping criterion is to end the sifting process when a proto-IMF verifies the two conditions of IMF. This issue, summarized in Equation (2.3), is critical because the success of the whole decomposition, and then of obtaining a physically meaningful instantaneous frequency entirely depends upon the correct enforcement of these two requirements. As we have seen, the proto-IMF resulting after the first iteration may not be an IMF because of imperfections in the sifting process due to the calculation of the mean with the envelopes. Therefore, more iterations are needed to ensure that riding waves and inflexion points have been correctly sifted, and that the local mean is almost equal to zero. On the other hand, too many iterations can also be damaging for the IMFs because, as Huang et al. (1998) [27] observed, it tends to flatten intrinsic oscillations thus distorting and affecting the original information. In addition, Rilling et al. (2003) [42] state that over-sifting can lead to over-decomposition, meaning that after too many iterations a single monocomponent can be spread on several successive IMFs.

First stopping criterion: A first idea for the implementation of the stopping criterion was suggested by Huang et al. (1998) [27], it is based on the standard deviation, SD, computed from two consecutive sifting results

$$SD(h_{j(k-1)}, h_{jk}) = \sum_{t=0}^{T} \left[\frac{|(h_{j(k-1)}(t) - h_{jk}(t))|^2}{h_{j(k-1)}^2(t)} \right], \tag{2.15}$$

in which j designates the sifting process number or the IMF number. Then, the sifting process is stopped and the j^{th} IMF is found if SD is inferior than a predetermined threshold SD_{max} (typical values lie between 0.2 and 0.3 [27]). Set in a mathematical formulation:

$$h_{jk} = c_j \quad \text{if} \quad SD(h_{j(k-1)}, h_{jk}) \leq SD_{max}. \tag{2.16}$$

However, this stopping criterion has several shortcomings according to Huang (2005b) [24]: for instance, even though the standard deviation is small, the first condition of equal numbers of extrema and zero-crossings may not be guaranteed.

Second stopping criterion: Afterwards, another stopping criterion, more related to the definition of the IMFs, has been presented by Huang et al. (1999, 2003) [26] [28]: the IMF is chosen as the first proto-IMF of a series of S consecutive iterations which successfully verify the first IMF-requirement. Set in a mathematical formulation:

$$\begin{aligned} h_{jk} = c_j \quad \text{if} \quad & |N_{zc}(h_{jk}) - N_{ext}(h_{jk})|, \dots \\ & \dots, |N_{zc}(h_{j(k+S-1)}) - N_{ext}(h_{j(k+S-1)})| \leq 1, \end{aligned} \tag{2.17}$$

in which N_{zc} designates the number of zero-crossings and N_{ext} the number of extrema. The S-number is a predetermined parameter which should be set between 4 and 8 according to Huang et al. (2003) [28]. This simple criterion not only guarantees the first condition, but the S successful iterations also ensure that all the extrema have been sifted and that the mean is approximately zero. Moreover, the first proto-IMF is chosen in the series in order to limit the problem of over-sifting already mentioned. Finally, as the S-number increases, the stopping criterion becomes stricter, and the number of iterations needed to obtain the IMF increases as well. Therefore, S must be chosen with care in order to obtain a meaningful decomposition.

Third stopping criterion: A third and simpler stopping criterion has sometimes been suggested [28] [42]. The sifting process is stopped after a predetermined number M of iterations, regardless of the two requirements. Set in a

mathematical formulation:

$$h_{jk} = c_j \quad \text{if} \quad k = M. \tag{2.18}$$

It can be either combined with the previous criterion, or it can be used alone. It is meant to prevent over-sifting and also to avoid a never-ending sifting loop [4]. However, this solution does not guarantee any of the two IMF-requirements, therefore it can be unsatisfying since the number of iterations depends very much on the data and it can also vary between IMFs of a same decomposition.

Fourth stopping criterion: A fourth stopping criterion handling the two IMF-requirements has been enunciated by Rilling et al. (2003) [42]: the sifting process is stopped if both the two following conditions are satisfied,

- the numbers of zero-crossings and extrema of the proto-IMF h_{jk} differ at most by one. (This is simply the first condition of IMF.)
- the absolute value of the ratio of the mean $m_{jk}(t)$ of h_{jk} to its mode amplitude (defined as $a_{jk}(t) = (e_{max}[h_{jk}(t)] - e_{min}[h_{jk}(t)])/2$, where e_{max} and e_{min} designate the upper and lower envelopes respectively) is lower than a predetermined threshold θ_1 for a fraction of the total signal size, say $(1 - \alpha)$; *and*, this ratio is lower than a second threshold θ_2.

Set in a mathematical formulation for a discrete-time series of length T:

$$h_{jk} = c_j \quad \text{if} \quad \begin{cases} |N_{zc}(h_{jk}) - N_{ext}(h_{jk})| \leq 1, \\ \text{and} \quad \sigma_{jk}(t) < \theta_1 \quad \forall\, t \in J \subset T \quad : \quad \#(J) \geq (1-\alpha)\#(T), \\ \text{and} \quad \sigma_{jk}(t) < \theta_2 \quad \forall\, t \in T, \end{cases} \tag{2.19}$$

in which $\sigma_{jk}(t) = |m_{jk}(t)/a_{jk}(t)|$, and $\#(J)$ and $\#(T)$ designate the cardinality (size) of the sets J and T respectively. As Rilling et al. (2003) [42] detail: the

[4] The problem of convergence of the sifting process is another unsolved mathematical issue of the HHT (see Huang 2005b [24]).

second condition imposes "*globally* small fluctuations in the mean" with the first threshold θ_1 and a small tolerance α, "while taking into account *locally* large excursions" with the second threshold θ_2 in the third condition. They further suggest to set $\theta_1 \approx 0.05$, $\alpha \approx 0.05$ and $\theta_2 \approx 10\theta_1$. As an example, the second condition of this stopping criterion (2^{nd} and 3^{rd} equations in (2.19)) for $(\theta_1 = 0.05, \theta_2 = 0.5, \alpha = 0.05)$ can be interpreted as follows. The relative mean of the IMF ($|m_{jk}/a_{jk}|$) has to be lower than $\theta_1 = 0.05$ for at least $(1 - \alpha) = 95\%$ of the data over the time span, *while* the relative mean of the remaining 5% of the data has to be only lower than $\theta_2 = 0.5$.

In the present work, the fourth stopping criterion has been used in the HHT algorithm because it seems to be the most complete one as it clearly accounts for the two conditions of IMF. Moreover, the influence of the thresholds and the tolerance on the results of the decomposition and on the Hilbert spectrum is thoroughly investigated in Chapter 3. The aim is to provide a good evaluation of the appropriate values to choose for the two thresholds and the tolerance.

2.2.6 Mode mixing in the decomposition

According to Huang et al. (1999) [26], the problem of mode mixing is inherent to a straightforward application of the EMD algorithm—and more precisely to the sifting process, as it has been described in Section 2.1.1. It can be caused by intermittent signals or noisy data, and the main consequence is the spread of modes between the IMFs. This problem must be prevented, and various solutions, such as the intermittency test presented by Huang et al. (1999, 2003) [26] [28], have been proposed to tackle it.

This phenomenon, likewise turbulence in fluid dynamics, is the mixing of different time scales in a single component. It can occur intermittently, meaning that it is not regular and therefore difficult to predict and interpret. Huang et al. (1999) [26] further explain that mode mixing is actually not physically pos-

sible because "no process can engender very different time scales in the same response"; consequently, it must be identified and the mixed components must be separated. In the EMD, the problem of intermittency is very important since it can severely affect the shape of an IMF as it tends to mingle different mono-components in the same mode. Normally, as we have seen in Section 2.2.1, IMFs should contain only one range of frequencies since each represents a monocomponent signal whose bandwidth clearly differs from others.

Figure 2.6 depicts the effect of mode mixing on the IMFs when the original EMD algorithm is used. As can be seen on Figures (e) to (h) the intermittent high-frequency component has strongly affected the decomposition. The first IMF contains two components of very different frequency bandwidth; in this example, the carrier frequency is actually one tenth of the intermittent frequency. In addition, it can also be observed that the next IMFs are affected by the problem of mode mixing in c_1. In fact, any problem encountered in one mode is transmitted to the subsequent modes as a result of Equation (2.4) at the end of the sifting process. So, to obtain the correct IMFs, from (b) to (d), the signal has first undergone an intermittency test. Finally, the intermittent signal has been entirely extracted in the first IMF, while the carrier has then been properly recovered in the second.

The intermittency test prescribed by Huang et al. (1999, 2003) [26] [28] works as follows: in signals or residues to be analyzed, if the distance between two successive extrema is greater than a predetermined value n_1, then all the data between these two extrema must be discarded from the resulting IMF. In other words, n_1 corresponds to the maximum half-period which the IMF can possess. The aim is to discriminate intermittent components, which are either noisy data or whose frequency is unexpected in the mode, from the signal so that the sifting process will not mingle very different frequency scales in the same IMF.

Other techniques tackling mode mixing can be found in Gao et al. (2008) [17]. Moreover, Gao et al. (2008) [17] offer an alternative to the intermittency test

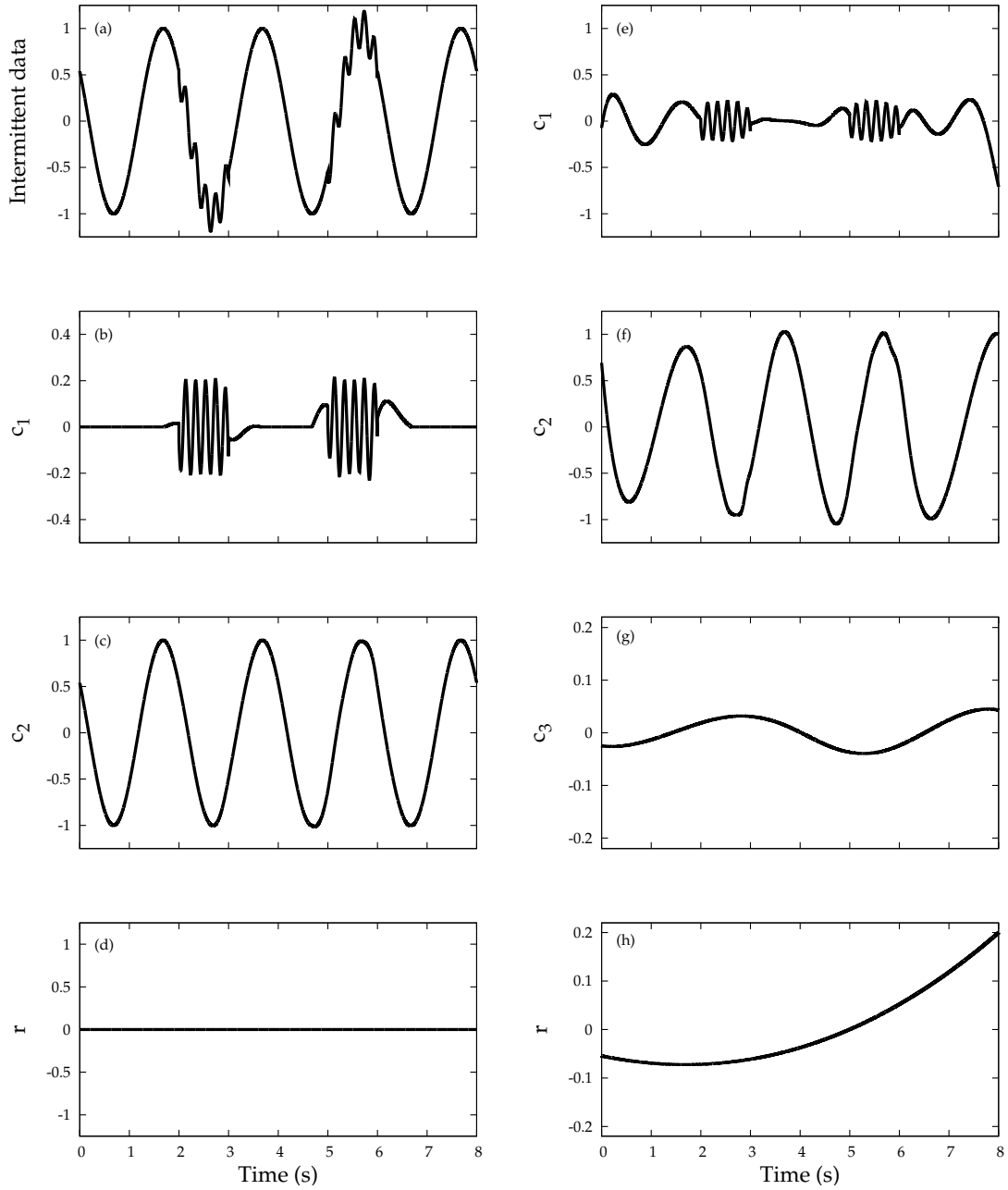

Figure 2.6: Illustration of mode mixing in the decomposition of an intermittent signal, Figure (a). Figures from (e) to (h) show the IMFs from a straightforward decomposition and using the algorithm presented in Section 2.1.1; we can see that mode mixing occurs in the first IMF c_1 and also that c_2, c_3, and r are affected. Figures from (b) to (d) display the IMFs of a decomposition of (a) using first the intermittency test and second the EMD algorithm; the intermittent low amplitude signal is completely retrieved in c_1 and does not mingle anymore with the lower-frequency sine wave successfully retrieved in c_2.

developed by Huang et al. (1999, 2003) [26] [28]. First, they suggest to use the Teager Kaiser Energy Operator to locate the intermittent components of the signal. They prefer this operator to the Hilbert transform since it is not subjected to the Gibbs effect [5] and since its computation is also slightly faster. Second, the mingled components are separated using a difference operator, the EMD and cumulative sums.

In conclusion, these two algorithms seem to be effective to prevent mode mixing, tests with LOD data show similar satisfying results. However, it must be noted that both of them need one predetermined parameter to discriminate either the critical half-period in the intermittency test, or the critical energy level in the second technique. Finally, Huang et al. (1999) [26] caution about the utilisation of such tests because any manipulation to the data increases the risk to affect the decomposition. Indeed, some information could be lost or the IMFs could be distorted by forcing the signal to behave in a particular way. In fact, the adaptive aspect of the HHT could be compromised by too many manipulations; however, intermittency and mode-mixing are not physical, therefore they must be prevented.

In the present work, we have chosen to implement the intermittency test of Huang et al. (1999, 2003) [26] [28]. Details regarding the algorithm will be presented in Section 2.3.5.

2.2.7 Confidence limit

Huang et al. (2003) [28] established a method to determine the confidence limit for the results of the HHT. Their method is based on the calculus of the ensemble mean of different sets of IMFs derived from a unique signal. The particularity of their approach is that they could not invoke the ergodic assumption [6] since

[5] The Gibbs effect, also known as 'ringing phenomenon', describes the overshooting of the Fourier series, or other eigenfunctions such as the Hilbert transform, at a jump discontinuity (see Weisstein (no date) [49]).

[6] The ergodic assumption is applicable for linear and stationary data and allows to substitute the ensemble mean by the temporal mean (see, for example, Gray and Davisson (1977) [19]).

most signals studied with the HHT are neither linear nor stationary, two necessary conditions for this assumption. Therefore, they suggested to generate different decompositions from the same data by varying the control parameters of the EMD. For example, by adjusting the second stopping criterion with various values for the S-number and M-number, they obtained slightly different sets of IMFs. In fact, each decomposition is statistically near from the ideal decomposition, and, as we have seen, the differences are related to the practical implementation of the EMD algorithm. Therefore, the ensemble mean and the standard deviation can be computed for each IMF (see Section 2.3.7 for calculation details), and the results yield the confidence limit of the data set without any loss in time and frequency resolution, a problem which frequently occurs under the ergodic assumption. In addition to providing with a standard measure of the accuracy of the marginal spectrum and the Hilbert spectrum, their method revealed the optimal range for the second stopping criterion (i.e. the stopping criterion that is likely to lead to a meaningful decomposition). In particular, Huang et al. (2003) [28] found that, in the case of the LOD data, the optimum S-number should be chosen between 4 and 8.

At first, the HHT can be rather difficult to monitor since several parameters such as the stopping criterion, the end-point option and the intermittency test can be adapted. In this regard, one objective of this study is to give some indications about these control parameters; and, likewise Huang et al. (2003) [28]'s study on the S-number, we will investigate the optimum choice for the fourth stopping criterion and for the end-point options throughout Chapter 3.

2.3 Implementation of the HHT algorithm

In this section are first described the crucial and adjustable control parameters of the HHT algorithm whose parametrization can affect the results. Second, means to assess the data are introduced. Third, the confidence-limit algorithm

is presented. Finally, the different parts of the HHT algorithm can be found in Appendix B.

2.3.1 Empirical mode decomposition

The source code showed in Section B.1 is a basic implementation of the empirical mode decomposition; it returns the IMFs and the last residue of an input signal. The end-point option can be chosen, and the thresholds of the fourth stopping criterion can be adjusted. A last option can be used to perform an intermittency test for some IMFs during the sifting process (see Section 2.3.5 for details on the algorithm of the intermittency test).

2.3.2 Hilbert transform

The source code presented in Section B.2 computes the analytic signal using the Hilbert transform, then the instantaneous amplitude and instantaneous frequency are calculated. The computation of the amplitude is a straightforward application of Equation (2.8). However, the computation of the frequency is not simply the derivative of the phase function, and the formula used in the algorithm is based on a method developed by Barnes (1992) [2]. In fact, the computation of the derivative of a discrete-time function can be difficult, so a good representation of the discrete-time instantaneous frequency is

$$w[t] = \frac{1}{2\Delta t} \tan^{-1}\left(\frac{x[t-\Delta t]y[t+\Delta t] - x[t+\Delta t]y[t-\Delta t]}{x[t-\Delta t]x[t+\Delta t] + y[t+\Delta t]y[t-\Delta t]}\right), \tag{2.20}$$

in which x and y denote respectively the real part and the imaginary part of a discrete-time analytic signal $z[t] = x[t] + iy[t]$, and Δt is the time step. A second method using the central difference scheme has been described by Boashash (1992) [4], it gives also satisfying results. The discrete-time instantaneous frequency is defined as

$$w[t] = \frac{\theta[t+\Delta t] - \theta[t-\Delta t]}{2\Delta t}, \tag{2.21}$$

where $\theta[t]$ is the discrete-time phase function of the analytic signal. Furthermore, the Matlab embedded function `unwrap` has been used in the computation of the instantaneous frequency in order to prevent 2π-periodic strong discontinuities.

Finally, the algorithm features three different extension options [7] to extend the data in order to deal with the Gibbs effect:

1. With the first option, there is no extension.
2. With the second option, the signal is mirrored anti-symmetrically at the edges.
3. The third option extends the data with a damped sinusoidal curve using an auto-regressive model. If one IMF contains less than one extrema, which can occur for intermittent IMFs, this option cannot be used so the first option will automatically be chosen for this IMF.

The extension has to be as continuous as possible, so that a smooth transition can alleviate the end-effect in the instantaneous amplitude and frequency curves.

2.3.3 End-point options[8]

Clamped end-points

The clamped end-point option is the simplest technique to terminate the cubic spline interpolations. The first and last points of the data are considered as both maxima and minima for every iteration of the sifting process. In other words, the IMFs are forced to be zero at their ends. Figure 2.7 depicts the upper and lower cubic spline envelopes using this option for the test data.

Though the risk of having large spurious swings in the IMFs has disappeared, this option imposes a strong constraint on the cubic spline envelopes. By

[7] These three *extension options* must not be confused with the four *end-point options* detailed in Section 2.3.3. The main difference is: the end-point options are applied to the signal, the residues or the proto-IMFs in the sifting process; whereas the extension options are used when applying the Hilbert transform to each IMF in the second step of the HHT algorithm.

[8] The source code of the four different end-point options can be found in Appendix B.1 at the end of the EMD algorithm.

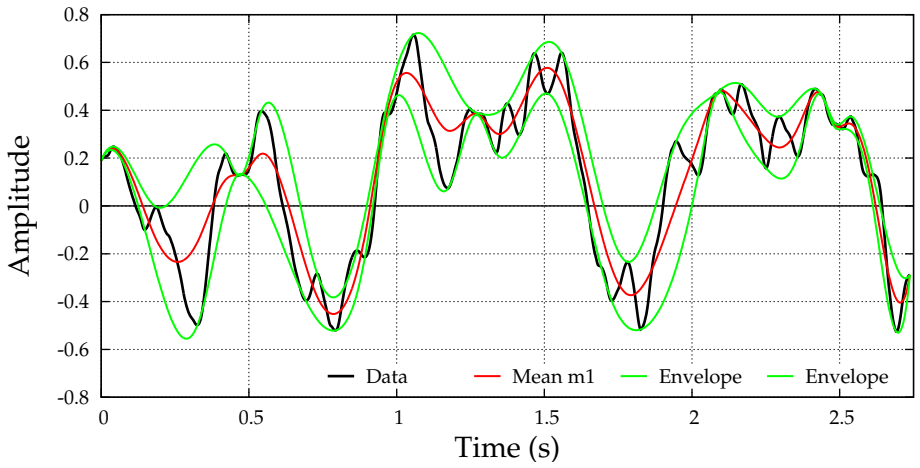

Figure 2.7: Illustration of the clamped end-point option. The first and last data points are considered as both maxima and minima. As a result, the mean curve is equal to the signal at the edges, and all the IMFs are forced to be zero at their ends.

reducing the degree of freedom of the IMFs, this technique actually creates distortion in the modes. Therefore, we will investigate in the next chapter whether the clamped end-option is adapted to minimize the propagated error due to the termination of the envelopes.

Extrema extension

The method of extrema extension was developed by Shen et al. (2005) [47], it consists of the addition of two extrema at the edges of the signal (see Figure 2.8). Considering the beginning of the signal (the procedure is exactly symmetrical for the end of the signal), the position and the amplitude of these two added extrema are calculated using the first data point, designated as (t_0, x_0), and the first two extrema, designated as (t_{e_1}, x_{e_1}) and (t_{e_2}, x_{e_2}) respectively (the nature of the extrema—minimum or maximum—has no importance). Then, the procedure to determine the extremum preceding the commencement of the signal $(t_{e_{-1}}, x_{e_{-1}})$ and the leftmost extremum $(t_{e_{-2}}, x_{e_{-2}})$ works as follows:

first,

$$\begin{cases} t_{e_{-1}} &= min(t_0, (t_{e_1} - (t_{e_2} - t_{e_1}))) \\ x_{e_{-1}} &= x_{e_2} \quad \text{if} \quad t_{e_{-1}} < t_0 \\ & \quad x_0 \quad \text{otherwise;} \end{cases} \tag{2.22}$$

second,

$$\begin{cases} t_{e_{-2}} &= t_{e_{-1}} - (t_{e_2} - t_{e_1}) \\ x_{e_{-2}} &= x_{e_1}. \end{cases} \tag{2.23}$$

This technique is actually an extension of a half-oscillation at both ends, and whose time scale and amplitude are based on the neighbouring first and last half-waves. It can be remarked that two extrapolated extrema on both sides are sufficient to calculate the cubic spline envelopes, which need at least three interpolation points. This corresponds to the number of maxima or minima of the last IMF. The main advantages of this technique are its small need in computation resources and its adaptive character. However, we can wonder whether it is sufficient to flatten the ends of the envelopes in every case. Moreover, it also assumes that the signal is locally stationary around the edges, a condition that may not always be true.

Mirror imaging extension

The mirror imaging technique is an extension of the data by reproducing the symmetry of the signal with respect to the first and last data points (see Figure 2.9). Thus, the continuity between the signal and its extension is immediate. Moreover, the nonlinearities that may exist in the signal are preserved. However, this technique imposes a strong constraint of periodicity, and any non-stationary or transient features occuring in the signal can therefore introduce some periodicity in the low-frequency IMFs. Finally, this option can also increase the computation burden if large data sets are studied.

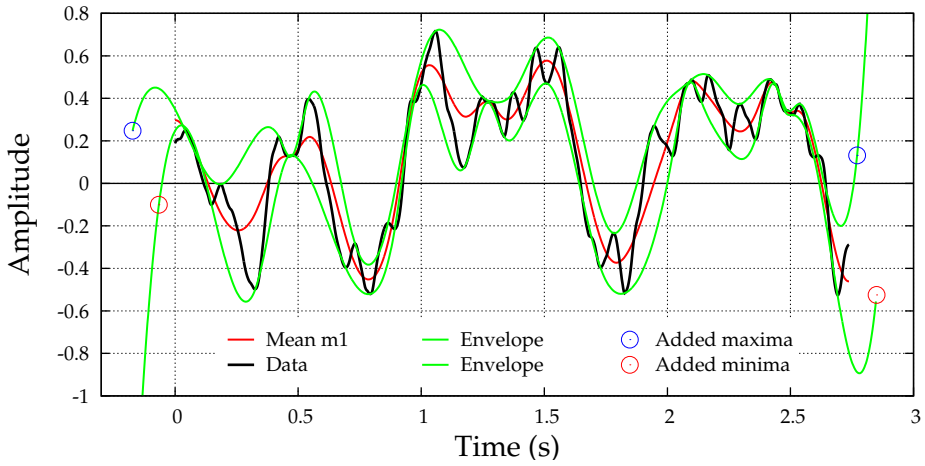

Figure 2.8: Illustration of the extrema extension technique. In this case, the two added extrema before and after the signal have been calculated with the first two and last two extrema respectively. We can observe that large swings have been created before the first interpolation point of the lower envelope and after the last interpolation point of the upper envelope. This is precisely the behaviour, which occurs when the envelopes are left unconstrained, that we must avoid within the length of the signal.

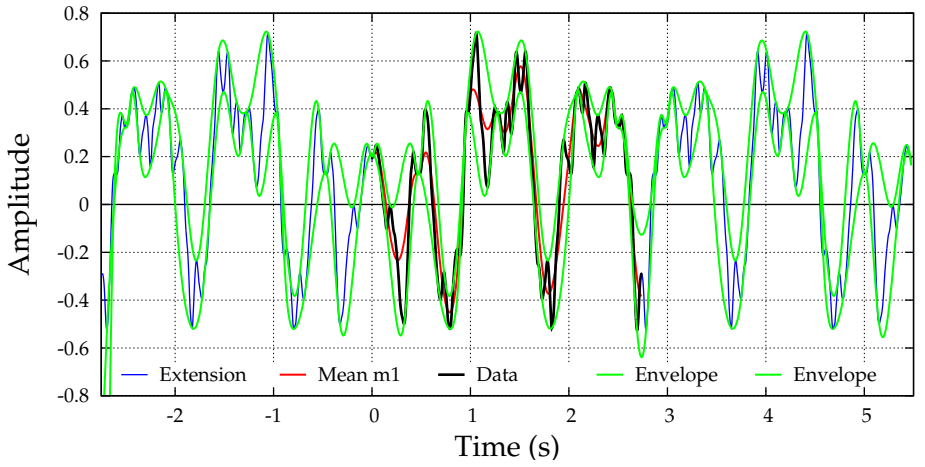

Figure 2.9: Illustration of the mirror imaging extension method.

Auto-regressive model

This last end-point option is also a signal extension technique. A damped sinusoidal curve, based on a second-order auto-regressive model, is extrapolated at the edges of the time series. The extrapolated points are calculated according to a recursive scheme based on the two preceding data points. The procedure for the extrapolation of the end of the signal (the procedure for the beginning is identical) can be described as follows:
let $X = (x(t_1), \ldots, x(t_N))$ a time series of size N, $X_{epl} = (x(t_{N+1}), \ldots, x(t_{N_{epl}}))$ the extrapolated sinusoidal curve of size N_{epl}, N_{avg} the length of averaging, κ the damping coefficient, ω_s the pulsation of the sinusoidal extension, and b_1 and b_2 two coefficients. First, the mean of the signal, μ, is shifted to zero according to the average calculated with the last N_{avg} points:

$$X_{shift} = X - \mu(N_{avg}) \quad \text{with} \quad \mu(N_{avg}) = mean(x(t_{N-N_{avg}+1}), \ldots, x(t_N)), \tag{2.24}$$

then, the two coefficients are

$$b_1 = \frac{2 - (\omega_s \Delta t)^2}{1 + \kappa \frac{\Delta t}{2}}, \quad b_2 = -\frac{1 - \kappa \frac{\Delta t}{2}}{1 + \kappa \frac{\Delta t}{2}}, \tag{2.25}$$

where $\Delta t = (t_2 - t_1)$ is the time step of the time series. Next, the extrapolated points are calculated recursively with the two preceding points,

$$x_{shift}(t_i) = b_1 \cdot x_{shift}(t_{i-1}) + b_2 \cdot x_{shift}(t_{i-2}) \quad \forall\, i \in \{(N+1), \ldots, N_{epl}\}, \tag{2.26}$$

and finally, the extrapolation sinusoidal curve is

$$X_{epl} = (x(t_{N+1}), \ldots, x(t_{N_{epl}})) = (x_{shift}(t_{N+1}), \ldots, x_{shift}(t_{N_{epl}})) + \mu(N_{avg}). \tag{2.27}$$

The pulsation ω_s (in the calculation of b_1) can be determined using the time scale defined by the nearest local extrema, as suggested by Coughlin and Tung

(2005) [11],

$$\omega_s = \frac{\pi}{\left| t_{e_l} - t_{e_{(l-1)}} \right|}, \tag{2.28}$$

where t_{e_l} and $t_{e_{(l-1)}}$ are the time instants of the last extrema and the next to last one respectively. Moreover, it has been found that the difference between the last two extrema has to be greater than four times the time step to prevent the auto-regressive model from diverging to infinity. So, the following condition is adopted in the algorithm:

$$\omega_s = \begin{cases} \frac{\pi}{\left| t_{e_l} - t_{e_{(l-1)}} \right|} & \text{if} \quad \left| t_{e_l} - t_{e_{(l-1)}} \right| \geq 4\Delta t \\ \frac{\pi}{4\Delta t} & \text{otherwise.} \end{cases} \tag{2.29}$$

It can also be remarked that the phase and the amplitude of the sinusoidal extension are automatically adjusted by the auto-regressive model.

This end-point option is illustrated in Figure 2.10. Several parameters (the length of extrapolation, the length of averaging, the size of the extrapolation and the damped coefficient) can be adjusted with this technique, and their values can depend on the signal studied. Finally, this technique is appropriate to flatten the envelopes without creating any artificial periodicity in the low frequency IMFs. However, nonlinear characteristics of the signal cannot be reproduced by this model.

2.3.4 Fourth stopping criterion

The algorithm of the fourth stopping criterion, summarized in Section 2.2.5 and initially developed by Rilling et al. (2003) [42], has been implemented using Matlab. It is described in Source Code 2.1. In the system of equations (2.19), the implementation of the first and third equations is almost straightforward, as can be seen in the source code. However, the second equation with the notion of cardinality is less obvious. So, the code can be interpreted as follows: condition 2 is verified if the proportion of data whose relative mean (ratio of the mean to the

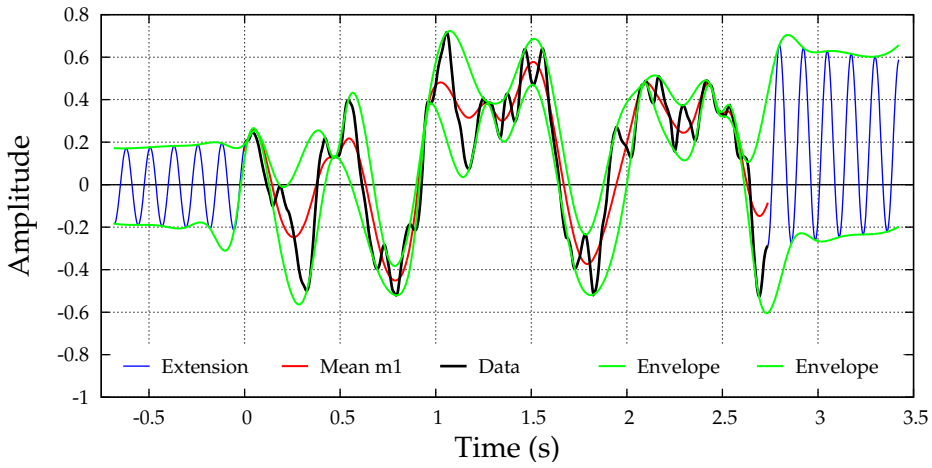

Figure 2.10: Illustration of the signal extension using an auto-regressive model. The signal has been extrapolated with a damped sinusoidal curve of length $N_{epl} = 0.25N$, a length of averaging $N_{avg} = 0.2N$ and a damping coefficient $\kappa = 0.5$.

mode amplitude) exceed the first threshold θ_1 is inferior than the tolerance α.

Source code 2.1: Matlab source code of the fourth stopping criterion

```
function stop_sifting = Stopping_criterion_4(nzc,ne,me,ma,thresholds)
% Stopping_criterion_4 returns stop_sifting == true if the current
% sifted proto-IMF satisfies the two IMF-requirements.

n_zc = nzc;                 % number of zero crossing
n_extr = ne;                % number of extrema
mean_pIMF = me;             % mean of the current proto-IMF
mode_amp = ma;              % mode amplitude
theta1 = thresholds(1); % first threshold
theta2 = thresholds(2); % second threshold
alpha = thresholds(3);  % tolerance

% Implementation of condition 1
cond_1 = (abs(n_zc - n_extr) <= 1);

% Implementation of condition 2
cond_2 = ((mean(abs(mean_pIMF) > theta1*abs(mode_amp))) < alpha);

% Implementation of condition 3
cond_3 = all(abs(mean_pIMF) < theta2*abs(mode_amp));

% Implementation of the stopping criterion
stop_sifting = (cond_1 && cond_2 && cond_3);

end
```

2.3.5 Intermittency test

Huang et al. (1999 and 2003) [26] [28] stressed the importance of using an intermittency test to prevent problems of mode mixing in the decomposition. However, they did not give a detailed description of the algorithm of this test whose principles were briefly presented in Section 2.2.6. Therefore, we have thought it would be useful to provide some explanations about its use and its implementation; in addition, the source code of the intermittency test can be found in Appendix B.3. So, the algorithm is as follows:

1. An EMD without intermittency test is performed to identify the IMFs with mode mixing.

2. The intermittent criterion $n_1(j)$ is determined for each IMF c_j (it defines the maximum half-period that can be found in c_j), zero or negative values are associated with the IMFs that do not require the test.

3. An EMD with intermittency test can be called, and the vector n_1 is added to the input parameters.

The intermittency test is automatically launched once at the beginning of the first iteration of the sifting process (more precisely, just after the search for the extrema and before the calculation of the cubic spline envelopes, see the second function of the source code in Appendix B.1) for each residue r_j which are to produce an imperfect IMF. In the residue r_j, if the distance between two successive extrema is greater than $n_1(j)$, then the upper and lower envelopes are forced to be equal to the residue in the portion of the curve between these two extrema. Therefore, the resulting mean $m_{j,int}$ is equal to r_j in the portions of the signal where the half-period is larger than $n_1(j)$, and equal to the genuine mean m_{j1} anywhere else. Finally, the intermittent IMF $c_{j,int}$—calculated by subtracting the residue from the mean—retains only the waves with half-period shorter than $n_1(j)$, and equals zero everywhere else. After that, the sifting process is immediately stopped without calling the stopping criterion function.

Four remarks regarding this algorithm:

- First, a new IMF is created for each strictly positive intermittency criterion. Indeed, imperfect IMFs must be split into as many modes as it actually contains.

- Second, the sifting process of residues which produce imperfect IMFs is stopped at the end of the first iteration (without the stopping criterion), and the first resulting proto-IMF is chosen as the IMF $c_{j,int}$ for practical reasons. In fact, we have found that if $c_{j,int}$ was sifted several times, for example until the two IMF-conditions are truly satisfied, it would result in many spurious large swings located at the portions where the curve is equal to zero. Therefore, to prevent the propagation of these strong distortions to the subsequent IMFs, we have decided to separate the intermittent IMF $c_{j,int}$ from the residue without performing a complete sifting process. Nevertheless, if the intermittent IMFs contained useful information they could still be analyzed separately with the EMD algorithm.

- Third, the choice of the intermittent criterion n_1 should be motivated by physical considerations: when confronted with plain mode mixing, and when the bandwidth of the entangled modes can be clearly separated.

- Fourth, the continuity in the intermittent IMF between the zero portions and the intermittent portions is automatically ensured by the cubic spline interpolations. That is, between the last extremum of a long wave (half-period larger than n_1) and the first two extrema of a short wave, the two envelopes separate from each other, and the upper and lower envelopes are interpolated toward their respective maximum and minimum. Therefore, a smooth transition is ensured.

In conclusion, the intermittency test separates intermittent signals from the rest of the data according to a predetermined criterion. The intermittent IMFs can be analysed separately if they have not been properly sifted. However, the

most important is eventually to prevent the mixing of modes because it does not represent any physical phenomenon.

2.3.6 Four quantitative indexes for the HHT

Having some means to assess the results given by the HHT is very important since the actual algorithm is not ideal. In fact, different parameters can be adjusted, such as those described in the previous sections, thus producing slightly different sets of IMFs for the decomposition of the same signal. Therefore, in this section are described five simple qualitative and quantitative means to assess the results of the EMD algorithm as well as the Hilbert spectrum.

Qualitative assessment

The decomposition into IMFs of a signal should first be inspected qualitatively by eye, as prescribed by Drazin (1992) [14]. Though it is a subjective and sometimes difficult approach, experience can help identify the most important features in a time-series signal. For example, trends and periodic characteristics of stationary data can be identified. Then, knowledge about the phenomenon studied can also be very useful to understand the representation of nonlinearities in the modes and eventually in the Hilbert spectrum. For example, if the original signal has some frequency or amplitude modulations, these characteristics will appear in the decomposition and will be revealed by the Hilbert transform. Finally, experience in the HHT algorithm and especially in the choice of the different control parameters, such as the stopping criterion, the end-point options and the intermittency test, can be very useful to assess the decomposition and adjust the parameters in order to improve the results.

Index of orthogonality

An index of orthogonality can assess accurately the decomposition. As discussed in Section 2.2.2, the orthogonality of the EMD is theoretically satisfied.

However, due to imperfections, the IMFs may not be orthogonal to each other in practice. Therefore, an overall index of orthogonality IO developed by Huang et al. (1998) [27] can be defined as follows

$$IO = \frac{\sum_t \sum_{j \neq k} |c_j(t) c_k(t)|}{\sum_t |X(t)|^2}. \tag{2.30}$$

The index of orthogonality should be as small as possible for a good decomposition. As an indication, Huang et al. (2003) [28] state that a decomposition is deemed correct if $IO \leq 0.1$.

Index of energy conservation

An index of energy conservation IEC was introduced by Chen et al. (2006) [8]. It can be computed as the ratio of the squared values of the IMFs to the squared values of the signal minus the residue

$$IEC = \frac{\sum_t \sum_j |c_j(t)|^2}{\sum_t |X(t) - r_n(t)|^2}. \tag{2.31}$$

The residue is not taken into account in this index because having, in some cases, a considerable energy relatively to the modes (the energy of the trend) it would have overshadowed the other modes, thus rendering the index useless. However, we can show that the index of energy conservation is actually related to the index of orthogonality. By virtue of the empirical mode decomposition we have

$$\big(X - r_n\big) = \sum_{j=1}^{n} c_j \tag{2.32}$$

Taking the square of this equation and expanding the right-hand side, we obtain

$$\big(X - r_n\big)^2 = \sum_{j=1}^{n} c_j^2 + 2 \sum_{j \neq k} \big(c_j c_k\big), \tag{2.33}$$

Finally, summing each side over time and dividing by the left-hand side we find

$$1 = IEC + 2\,IO. \tag{2.34}$$

Therefore, the index of energy conservation will not be used in this study since it is redundant with the index of orthogonality.

Index of component separation

An index of component separation, ICS, is introduced to give an accurate measure of the separation of the instantaneous bandwidth of two monocomponents. As Cohen (1995) [10] explains, a signal is a multicomponent signal if the instantaneous bandwidths of its components, defined as the ratio between the time derivative of the amplitude and the amplitude $a'(t)/a(t)$, are small compared to the difference of their instantaneous frequency. In other words, this index can be applied to a pair of successive IMFs, c_j and c_{j+1}, and taking only the oscillatory components of the Hilbert-Huang transform of a signal (from Equation (2.10)),

$$z(t) = \sum_{j=1}^{n} a_j(t) \exp\left(i \int \omega_j(t) dt\right), \tag{2.35}$$

then, two successive IMFs are separated if

$$\left|\frac{a'_j(t)}{a_j(t)}\right| \quad , \quad \left|\frac{a'_{j+1}(t)}{a_{j+1}(t)}\right| \ll |\omega_{j+1}(t) - \omega_j(t)| \quad \text{with} \quad 1 \leq j < n. \tag{2.36}$$

Therefore, the instantaneous index of component separation can be defined as the logarithm of the ratio of the right-hand side to the left-hand side of Equation (2.36) for each component

$$ICS_j(t) = \left[\log\left(\frac{|\omega_{j+1}(t) - \omega_j(t)|}{\left|\frac{a'_{j+1}(t)}{a_{j+1}(t)}\right|}\right), \log\left(\frac{|\omega_{j+1}(t) - \omega_j(t)|}{\left|\frac{a'_j(t)}{a_j(t)}\right|}\right)\right] \quad \text{with} \quad 1 \leq j < n, \tag{2.37}$$

and it must satisfy

$$ICS_j(t) > 0 \qquad \text{for all} \quad 1 \leq j < n \tag{2.38}$$

to ensure that the IMFs are well separated. If this criterion is not satisfied, it can mean that there is mode spreading over successive IMFs, in other words the algorithm has mixed some modes and this problem should be solved using the intermittency test. Otherwise, it can signify that there is over-decomposition, that is, the same monocomponent has been decomposed on two IMFs by the sifting process. This problem should be solved by relaxing the stopping criterion, which may be too strict, or by changing other control parameters. Finally, the time average of this index can be calculated for stationary signals

$$\overline{ICS_j} = \frac{1}{T}\sum_{i=1}^{N} ICS_j[t_i] \qquad \text{with} \quad 1 \leq j < n, \tag{2.39}$$

where $T = (t_N - t_1)$ is the time span of the signal. This index is very important to assess the Hilbert spectrum because it gives an evaluation of the frequency resolution.

Number of IMF

The number of IMF, designated by N_{IMF} is also a simple quantitative means to evaluate the decomposition. It is essentially useful when compared to other decompositions of the same signal. In most cases, N_{IMF} should not vary more than one IMF between different sets of IMFs. However, it should not change significantly, and a set with a very different number of IMFs than the average is considered unsatisfying.

Number of iterations

The number of iterations for each IMF $N_{ite,j}$, or the total number of iterations for a set of IMFs $N_{ite,T}$ is another simple comparative means to assess the EMD

algorithm. This criterion is greatly influenced by the stopping criterion as we will discuss in Chapter 3. Moreover, $N_{ite,j}$ can fluctuate by more than 10 or 20 iterations between different decomposition sets. This is mainly due to the stopping criterion: if the thresholds are low, which means stricter constraints, then the number of iterations tends to increase, and conversely. However, overall, the number of iterations should not vary too much for a given IMF, and a limit of 500 iterations per IMF will be set for the studies of the LOD data (see Section 3.3) and the vortex-shedding signal (see Section 3.4) to prevent problems of convergence in the sifting process.

2.3.7 Confidence limit

The confidence-limit algorithm is based on the study of Huang et al. (2003) [28] whose results were reported in Section 2.2.7. It aims at giving a quantitative view of the results given by HHT. Several sets of IMFs are calculated from the same data but with different control parameters, and the resulting sets are assumed to have equal probability. Then, the algorithm calculates the ensemble mean of the sets of IMF and the standard deviation to give a confidence limit. A preliminary test is performed before calculating the mean and the standard deviation, all the sets whose IO is not below predefined thresholds (e.g. $IO \leq 0.1$) are discarded. Moreover, Huang et al. (2003) [28] explains: "assuming that the error is normally distributed, the confidence limit is usually defined as a range of values near this mean: one standard deviation is equivalent to 68%, and two standard deviations are equivalent to a 95% confidence limit." Finally, the algorithm produces the following results: the time evolution of the mean IMFs and their standard deviation; the marginal spectra of all the cases with the mean marginal spectrum and the 68% or 95% confidence limit (CL) marginal spectra; and the mean Hilbert spectrum.

The architecture of the confidence-limit algorithm is shown in appendix B.4. Two cases, which need a different implementation, must be distinguished to

compute the ensemble mean and which need different implementation:

1. If the numbers of IMFs in each set or case, which have passed the preliminary tests, are equal, then, the ensemble mean $E(c_j)$ and the standard deviation $\sigma(c_j)$ of each IMF c_j can be computed as follows

$$E(c_j) = \overline{c_j} = \frac{1}{N_{set}} \sum_{i=1}^{N_{set}} c_{j,i} \quad \text{with} \quad 1 \leq j \leq N_{IMF}, \tag{2.40}$$

and

$$\sigma(c_j) = \sqrt{E\big((c_j - \overline{c_j})^2\big)} = \sqrt{\frac{1}{N_{set}} \sum_{i=1}^{N_{set}} (c_{j,i} - \overline{c_j})^2} \quad \text{with} \quad 1 \leq j \leq N_{IMF}, \tag{2.41}$$

where N_{set} designates the number of sets, and $c_{j,i}$ the j^{th} IMF of the i^{th} set.

2. If the numbers of IMFs differ between the cases, a straightforward computation of the ensemble mean and the standard deviation is not possible. However, a *bin method* developed by Huang et al. (2003) [28] can be used. It consists in averaging the Hilbert spectra of the different cases. A time-frequency grid with rectangular bins of width the frequency step and length the time step is defined. Then, the amplitude of all the points belonging to the same bin is averaged. The computation of the discrete-value mean Hilbert spectrum $H[t, \omega]$ is defined as follows: let $t_1, \ldots, t_j, \ldots, t_m$ m increasing time values of constant time step $\Delta t = (t_2 - t_1)$, and $\omega_1, \ldots, \omega_k, \ldots, \omega_n$ n increasing frequency values of constant frequency step $\Delta\omega = (\omega_2 - \omega_1)$, then for all $1 \leq j \leq m$ and for all $1 \leq k \leq n$

$$E(H_{j,k}) = \overline{H_{j,k}} = \overline{H[t_j, \omega_k]} = \frac{1}{N_{j,k}} \sum_{\substack{t_j \leq t_i < t_{j+1} \\ \omega_k \leq \omega_i < \omega_{k+1}}} H_i[t_i, \omega_i], \tag{2.42}$$

and for all $1 \leq j \leq m$ and for all $1 \leq k \leq n$

$$\sigma(H_{j,k}) = \sqrt{E\Big(\big(H_{j,k} - \overline{H_{j,k}}\big)^2\Big)} = \sqrt{\frac{1}{N_{j,k}} \sum_{\substack{t_j \leq t_i < t_{j+1} \\ \omega_k \leq \omega_i < \omega_{k+1}}} \Big(H_i[t_i, \omega_i] - \overline{H[t_j, \omega_k]}\Big)^2}, \tag{2.43}$$

where H_i is the amplitude of any set i and $N_{j,k}$ is the number of points $H_i[t_i, \omega_i]$ belonging to the bin defined as: $t_j \leq t_i < t_{j+1}$ and $\omega_k \leq \omega_i < \omega_{k+1}$. The definitions of the time step Δt and the frequency step $\Delta\omega$ are important: a fine grid tends to increase the squattering of the results; on the other hand, a loose grid smoothes the results but lowers the resolution.

3. Results and discussion

3.1 Procedures

First, five simple test signals, i.e. a two-component sinusoidal signal, an amplitude-modulated signal, a frequency-modulated signal, an amplitude-step signal and a frequency-shift signal, are studied with the HHT algorithm to understand the influence of the end-point option and the fourth stopping criterion on the index of orthogonality, the index of component separation, the number of iterations and the number of IMFs. Second, through a systematic analysis of length-of-day data, we investigate the optimal end-point option and the optimal thresholds and tolerance for the stopping criterion. Third, a vortex-shedding signal from hot-wire measurements in the wake of a cylinder are studied to show the importance of the control parameters and the intermittency test to obtain meaningful results.

For each study, the same formalism is used to describe the parameters chosen in the algorithms. The EMD algorithm is referred to as '`EMD`', and it has the following inputs: `EMD([t;signal],epo,thresholds)`, and a last argument `n1` (a row vector) can be added when invoking the intermittency test. Precisely, `t` designates a one-row time vector written in the Matlab formalism, i.e.

`t=t0:Dt:tN` where `t0` is the first time instant, `Dt` is the time step (no time step means `Dt=1`) and `tN` is the last point; next, `signal` is the name of the signal; `epo` is the end-point option with `epo=1` for the clamped end-point option, `epo=2` for the extrema extension option, `epo=3` for the mirror imaging option and `epo=4` for the auto-regressive extension option; and `thresholds` is a row vector with `thresholds(1)=theta1` the first threshold θ_1, `thresholds(2)=theta2` the second threshold θ_2 and `thresholds(3)=alpha` the tolerance α of the fourth stopping criterion. Finally, the intermittent criterion vector `n1` follows the same notation as used by Huang et al. (2003) [28]: it is a row vector in which each coordinate corresponds to the intermittent criterion assigned to the IMF of same index. Moreover, a zero value indicates that the intermittency test is not performed, superscripts indicate the number of consecutive IMFs having the same criterion and a negative value is used in last position if all the remaining IMFs do not need the test. For example, `n1=[5,0`^{3}`,30`^{2}`,-1]` means that an intermittency test is performed on the first IMF c_1 with $n_1(1) = 5$, then no intermittency test is called for IMFs c_2 to c_4, then a test with the same criterion $n_1(5) = n_1(6) = 30$ is performed on IMFs c_5 and c_6, and finally IMFs c_7 to c_n are sifted normally without invoking the intermittency test. A last remark: the units of `n1` are the same as the units of `t`. Similarly, the Hilbert-transform algorithm is referred to as '`HT`'. Its input arguments are: '`HT([t;signal],eo,le)`' in which `eo` designates the extension option with `eo=1` for no extension, `eo=2` for an antisymmetric mirror imaging extension and `eo=3` for an extension with a damped sinusoidal curve using an auto-regressive model (see Section 2.3.2 for details about the three extension options). `le` designates the length of extension as a proportion of the size of the signal.

3.2 Study of five simple test signals

Five simple signals are used to test the HHT algorithm (EMD and Hilbert-transform algorithms) and to study the influence of the end-point options and the stopping criterion on both the decomposition and the Hilbert spectrum. Moreover, the analysis of these test signals can help us calibrate the four quantitative criteria, IO, ICS, N_{IMF} and N_{ite}, and also give us precious information on their reliability to assess the results of the HHT. The test signals chosen were previously investigated by Huang et al. (1998) [27] and Meeson (2005) [35]. They both gave a thorough interpretation of the decomposition and the Hilbert spectrum of each signal. For example, Huang et al. (1998) [27] explained through a comparative analysis with the wavelet transform and the Fourier transform how nonlinearities, such as frequency or amplitude modulation, can be accounted for by the Hilbert spectrum. Then, Meeson (2005) [35] showed the differences between the sifting process in the EMD and filtering techniques. In particular, he found that the Hilbert spectrum provides a better representation of non-stationary phenomena such as a step of amplitude or a frequency shift. Furthermore, he discovered aliasing in the HHT which can lead to misinterpretation of signals, whose frequencies are above the peak-to-peak sampling rate, as lower-frequency components. However, the purpose of our analysis is not to interpret further these signals, but rather to study the impact of the control parameters on the HHT as well as the results found by the quantitative criteria. The five test signals are:

1. A two-component sinusoidal signal, see Figure 3.1 (a):

$$S_1(t) = \cos(2\pi t) + 0.5\cos(\pi t). \tag{3.1}$$

2. An amplitude-modulated signal, see Figure 3.1 (b):

$$S_2(t) = \Big(1 + 0.5\cos(\pi t)\Big)\cos(2\pi t). \tag{3.2}$$

3. A frequency-modulated signal, see Figure 3.1 (c):

$$S_3(t) = \cos\Big(2\pi t + 0.5\sin(2\pi t)\Big). \tag{3.3}$$

4. An amplitude-step signal, see Figure 3.1 (d):

$$\begin{cases} \sin(2\pi t) & \forall\, t \leq 0 \\ 2\sin(2\pi t) & \forall\, t \geq 0 \end{cases} \tag{3.4}$$

5. A frequency-shift signal, see Figure 3.1 (e):

$$\begin{cases} \sin(2\pi t) & \forall\, t \leq 0 \\ \sin(4\pi t) & \forall\, t \geq 0 \end{cases} \tag{3.5}$$

Each signal is studied with each end-point option and three different groups of thresholds for the stopping criterion: `thresholds=[0.01,0.1,0.01]`, a strict criterion; `thresholds=[0.05,0.5,0.05]`, the criterion suggested by Rilling et al. (2003) [42]; and `thresholds=[0.1,1,0.1]`, a loose criterion. The parameters for the fourth end-point option as well as the extension options for the Hilbert-transform algorithm are adjusted to obtain the best results. The Hilbert spectrum of a few cases are presented to illustrate interesting behaviours of the HHT, and all the results for the quantitative criteria are reported in Appendix C.

3.2.1 Two-component signal

Empirical mode decomposition: From the results presented in Table C.1 of Appendix C.1 we can notice that, overall, the EMD of the two-component signal is good independently from the choice of the end-point option. The shape of the IMFs (not shown in this report) is very similar between the last three options, slight differences can be found at the edges where the choice of the option can affect the termination of the curves. Moreover, for all the end-point tech-

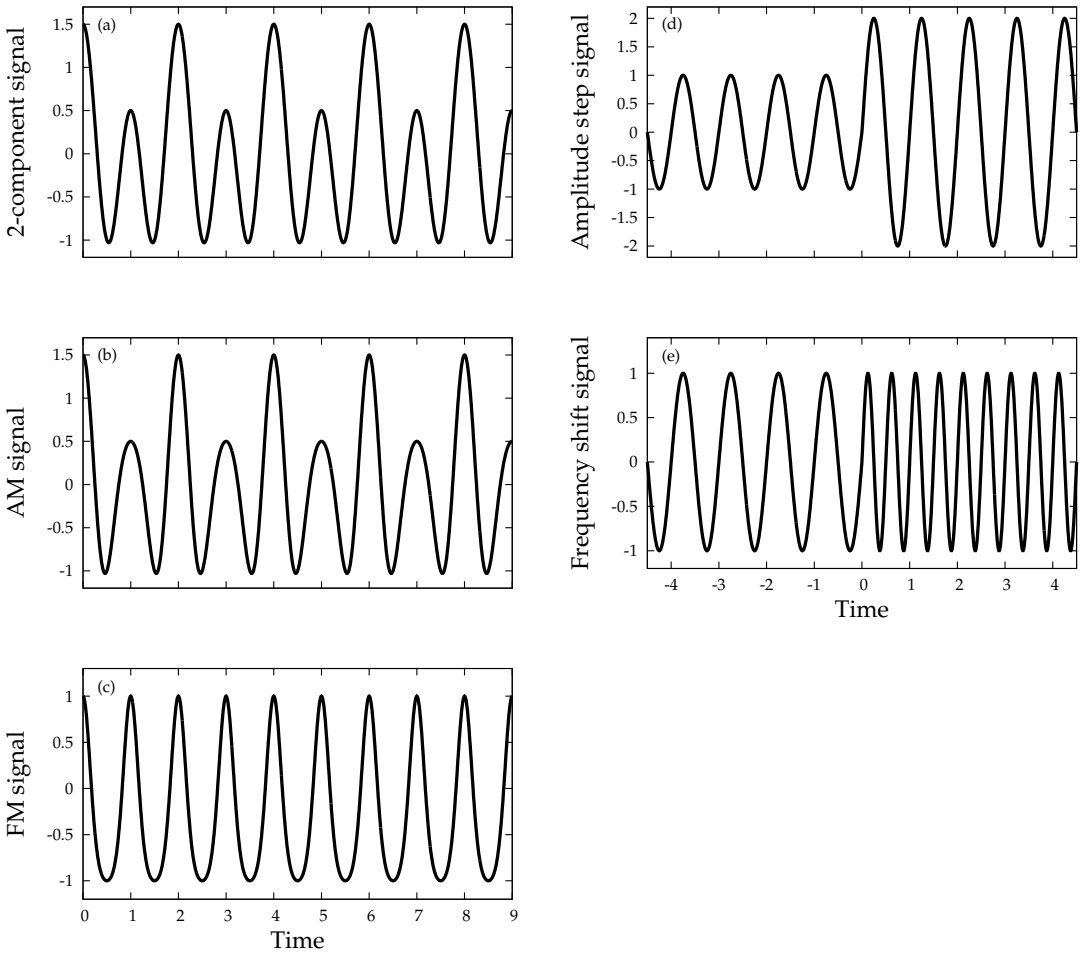

Figure 3.1: Five simple signals: (a) is a two-component sinusoidal signal, (b) is an amplitude-modulated signal, (c) is a frequency-modulated signal, (d) is an amplitude-step signal and (e) is a frequency-shift signal. The first signal is both stationary and linear, then the second and third are nonlinear signals and finally the last two are non-stationary signals.

niques, except the clamped end-point technique, the number of IMFs is either two or three. In fact, the third IMF, which actually has a very low amplitude in all cases, is generally produced with the two lowest groups of thresholds (`thresholds=[0.01,0.1,0.01]` or `thresholds=[0.05,0.5,0.05]`) because they impose more sifting iterations as can be seen with the results of $Nite, j$. Then, the index of orthogonality are very good for these options, i.e. $IO < 0.05$. Finally, the first end-point option (clamped end-point option) differs from the others since it cannot be used with the second and third conditions of the fourth stopping criterion (see Section 2.2.5) due to a problem of convergence in the sifting process. This option actually imposes a strong constraint on the ends of the IMFs, thus distorting each IMF so that the fourth stopping criterion can never be satisfied.

Hilbert spectrum: The Hilbert spectrum in the case of the first end-point option has been plotted on Figure 3.2. Because of a relaxed stopping criterion, no iterations of the sifting process were needed, so the signal was found as IMF. However, we can see from the Hilbert spectrum that the two components of the signal have not been found. Instead, it shows non-physical amplitude modulation and frequency modulation. This illustrates the importance of decomposing a signal into IMFs before applying the Hilbert transform. The results with the other options show some discrepancies caused by the occurrence of end-effect. In fact, the propagation of the error at the edges can be emphasized by the extension option. For example, the Hilbert spectrum on Figure 3.4 (with the IMFs on Figure 3.3) obtained with the second extension option (anti-symmetric mirror imaging) displays strongly distorted wavy components with much larger amplitude at the ends. This extension technique, instead of minimizing the end-effect, seems to amplify it. Moreover, this result is also supported by the index of component separation because it has negative values: $\overline{ICS_1} = (-1.72, -0.82)$. On the other hand, Figure 3.6 (with the IMFs on Figure 3.5) shows an excellent Hilbert spectrum: the frequency and the amplitude of the two components

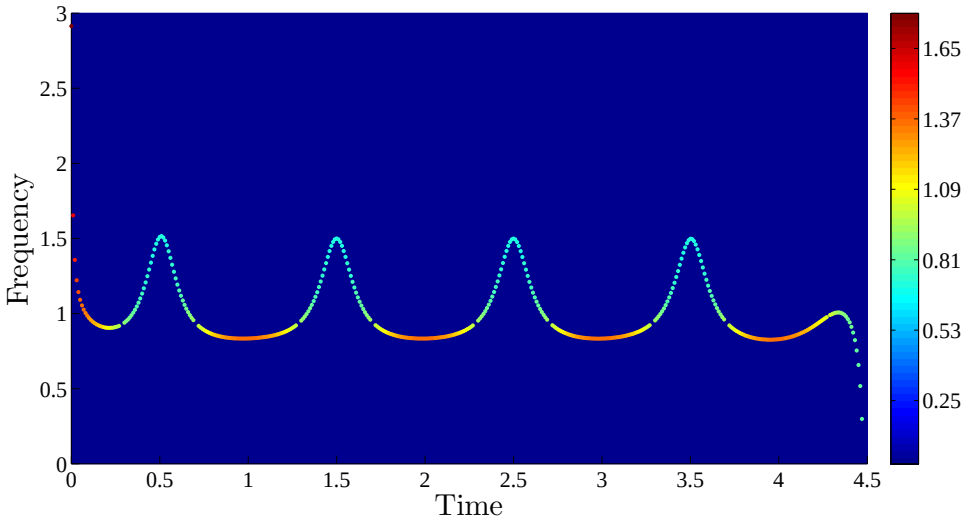

Figure 3.2: Hilbert spectrum of the two-component signal (s1) with the following parameters: `EMD([0:0.01:9;s1],1,-)` and `HT([t1;s1],1,0)`. Due to a problem of convergence with the clamped end-point option, the stopping criterion has been relaxed (only the first IMF-condition was required). Eventually the signal itself was recorded as IMF. However, the Hilbert spectrum is not correct because it shows amplitude modulation and frequency modulation in this simple signal.

are well retrieved (normalized in the figure); and the end-effect is successfully mitigated by the third extension option (auto-regressive model). Moreover, the third, and spurious, IMF has a very low amplitude and its frequency does not mix with the two main components. Finally, the indexes of component separation for this graph are very high, $ICS_1 = (2.34, 2.14)$ and $ICS_2 = (1.46, 0.79)$, thus emphasizing the good frequency separation of the IMFs.

3.2.2 Amplitude-modulated signal

Empirical mode decomposition: According to Table C.2 in Appendix C.2, similar tendencies as those observed with the previous signal can be found for the amplitude-modulated signal. First, the clamped end-point option differs once again from the other options. Second, the decomposition is still strongly influenced by the stopping criterion. Then, the index of orthogonality, the number of IMFs and the number of iterations per IMF all give a good evaluation of the results of the EMD. In this regard, the fourth end-point option is a good

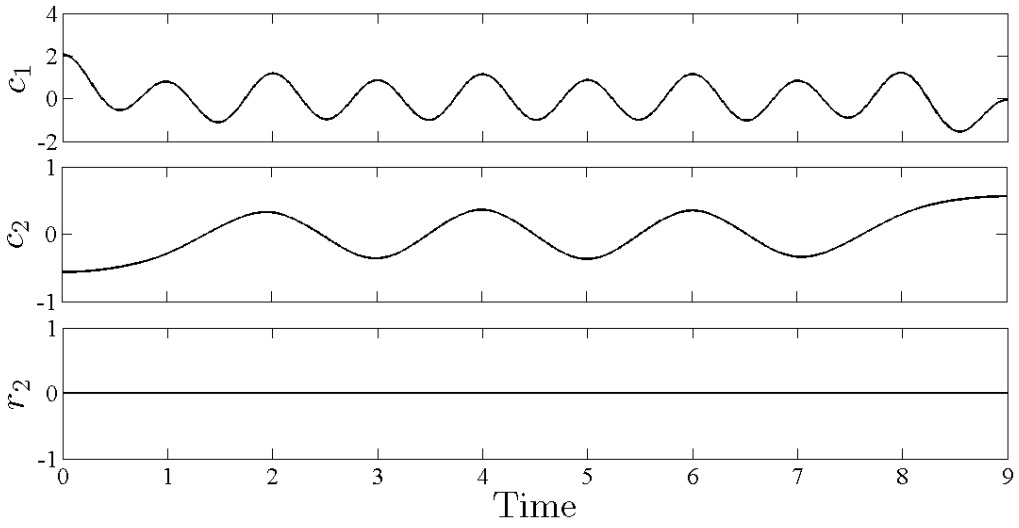

Figure 3.3: IMFs of the two-component signal (s1) with the following parameters: `EMD([0:0.01:9;s1],3,[0.1,1,0.1])`.

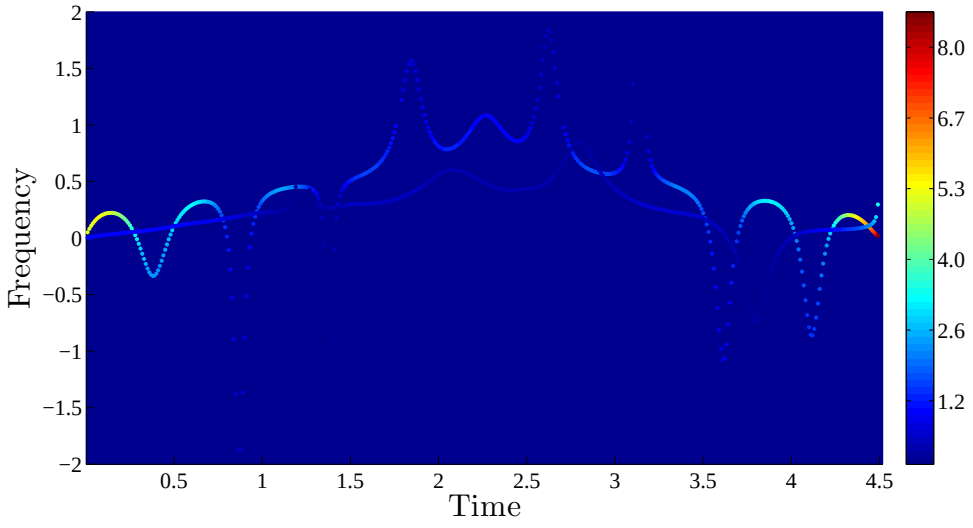

Figure 3.4: Hilbert spectrum of the two-component signal (s1) with the following parameters: `EMD([0:0.01:9;s1],3,[0.1,1,0.1])` and `HT([t1;s1],2,0.9)`. The end-effect is very strong, as shown by the spurious oscillations of the two components, thus meaning that the extension option is not adapted to attenuate it.

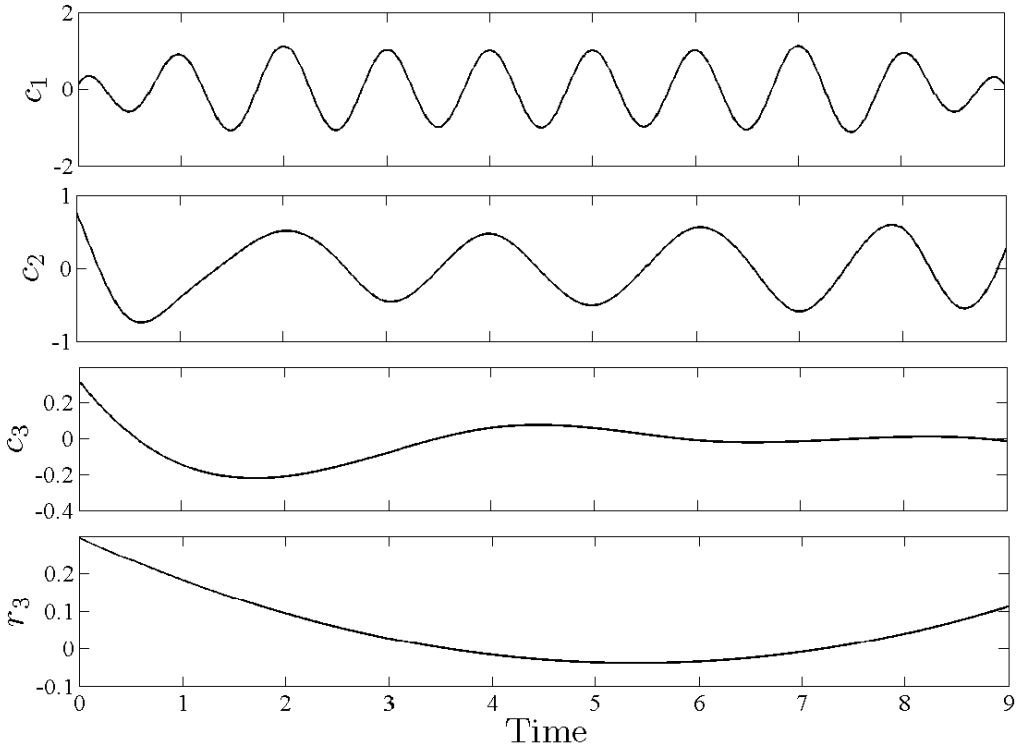

Figure 3.5: IMFs of the two-component signal (s1) with the parameters: `EMD([0:0.01:9;s1],4,[0.01,0.1,0.01])`.

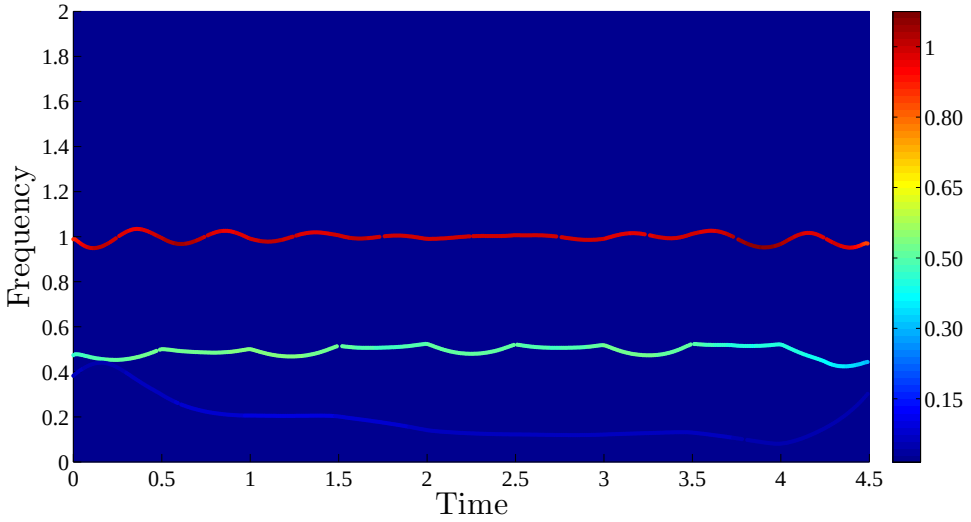

Figure 3.6: Hilbert spectrum of the two-component signal (s1) with the parameters: `EMD([0:0.01:9;s1],4,[0.01,0.1,0.01])` and `HT([t1;s1],3,1)`; moreover, the parameters for the third extension option are: $N_{avg} = 0.25N$ and $\kappa = 0.001$. The frequency resolution is excellent and the two components are very well retrieved.

illustration: low thresholds (`thresholds= [0.01,0.1,0.01]`) in the first row have induced over-sifting, as shown by the large numbers of iterations ($N_{ite,j} = (71, 66, 56, 175, 90, 27, 10)$), and over-decomposition, as shown by the large number of IMFs ($N_{IMF} = 7$), and the IO is as high as 4.7. On the other hand, the least strict constraints with this option (`thresholds=[0.1,1,0.1]`) give the best results: few iterations $N_{ite,j} = (2, 0)$, two IMFs and a good index: $IO = 0.03$. Finally, it can be remarked that, overall, the index of orthogonality is not as good as in the study of the two-component signal. This may be due to the nonlinear character of this signal. In fact, Huang et al. (1998) [27] stated that nonlinear signals are not supposed to be orthogonal in the sense of the Fourier decomposition. Nevertheless, the results show that the good decompositions of this amplitude-modulated signal have relatively low low IO.

Hilbert spectrum: The amplitude-modulated signal can be interpreted in two manners. First, the signal can be seen as a monocomponent signal with a single frequency and a modulated instantaneous amplitude. This point of view is illustrated by the Hilbert spectrum obtained with the first end-point option or clamped end-point option (see Figure 3.7). We can also notice that the end-effect has been correctly mitigated by the use of the third extension option (auto-regressive model), with length of extension equals to the signal size and a low damping coefficient $\kappa = 0.01$. Second, the signal can be decomposed in two components as shown on Figure 3.8: a first IMF with a constant amplitude $a_1(t) \approx 1$ and a modulated frequency varying from $0.73 < \omega_1 < 1.15$ and oscillating at a constant period $T_{\omega_1} = 1$; and a second IMF with a constant amplitude $a_2(t) = 0.5$ and a constant frequency $\omega_2 = 0.5$ (a third IMF has also been found by the EMD, but it has a much lower amplitude, $a_3(t) \approx 0.01$, than the first two IMFs and it is not visible on Figure 3.9). This second type of decomposition has also been found by Meeson (2005) [35] who suggests that it is not accurate because of the problem of aliasing. Then, we can observe that no extension of the IMFs has led to a localized problem of increasing or decreasing frequency (by 20% to 50%) and

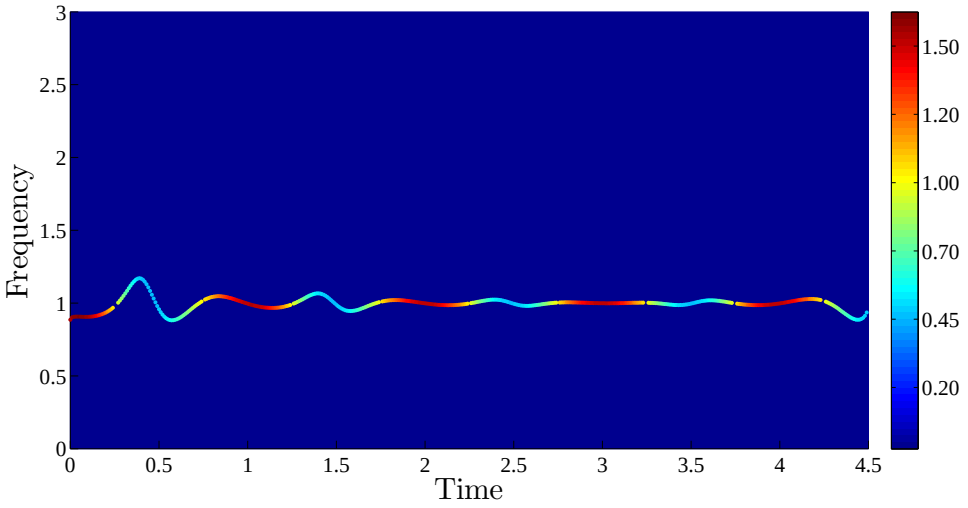

Figure 3.7: Hilbert spectrum of the amplitude-modulated signal (s2) with the clamped end-point option and the following parameters: `EMD([0:0.01:9;s2],1,-)` and `HT([t2;s2],3,1)`; moreover, the parameters for the third extension option (auto-regressive model) are: $N_{avg} = 0.25N$ and $\kappa = 0.01$. There is only one mode with a modulated frequency around $\omega_1 = 1$. Finally the end-effect has been very well handled by the third extension option.

increasing amplitude (by a factor 2) at both ends of the signal. Although there is a problem of end-effect, the index of component separation is still good for the first two IMFs: $\overline{ICS_1} = (1.43, 1.88)$, meaning that their frequency bandwidth are globally well distinguished.

3.2.3 Frequency-modulated signal

Empirical mode decomposition: When the first three end-point options are used, the frequency-modulated signal is not decomposed into modes because it already satisfies the two IMF-conditions (see Table C.3 in Appendix C.3). Only the last option finds 5 different IMFs; however, this result is not correct because the algorithm has produced spurious IMFs. In fact, the fourth option does not seem suitable with a strict stopping criterion, and the results are better when the thresholds increase.

Hilbert spectrum: All the decompositions being similar, except for the first and second cases of the fourth end-point option, the Hilbert spectra are also

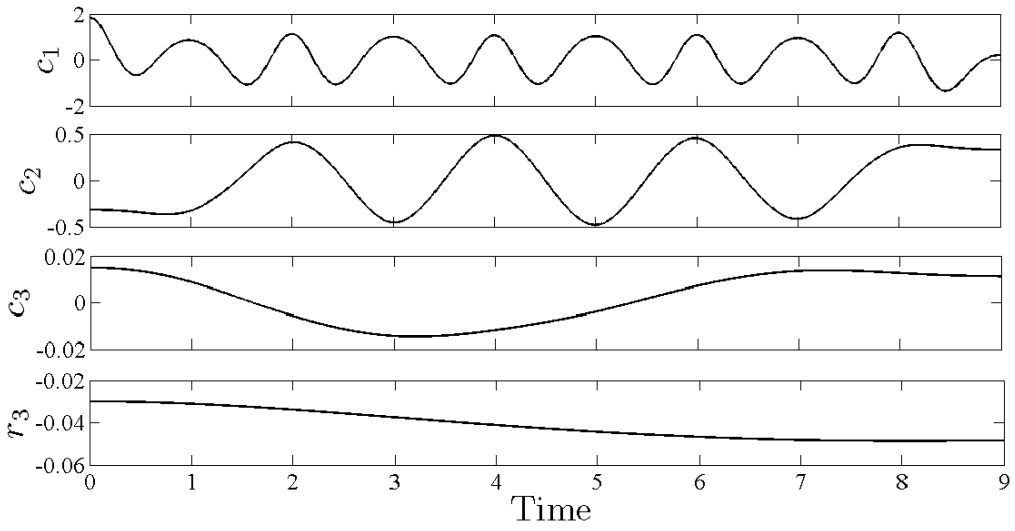

Figure 3.8: IMFs of the amplitude-modulated signal (s2) with the parameters: `EMD([0:0.01:9;s2],3,[0.05,0.5,0.05])`. There are three IMFs from this signal: a constant-amplitude frequency-modulated component, a carrier and a negligible very-low-amplitude component.

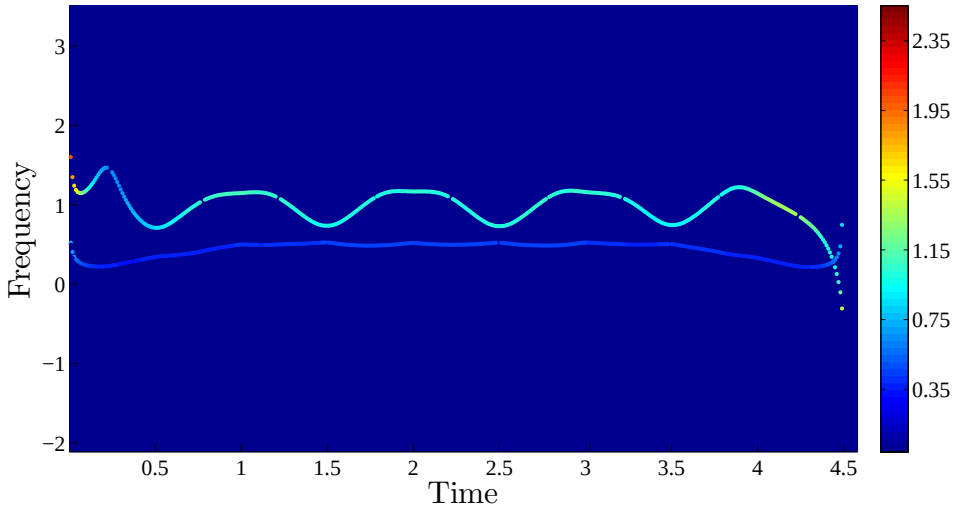

Figure 3.9: Hilbert spectrum of the amplitude-modulated signal (s2) with the parameters: `EMD([0:0.01:9;s2],3,[0.05,0.5,0.05])` and `HT([t2;s2],1,0)`. There are three IMFs from this signal: a constant-amplitude frequency-modulated component, a carrier and a negligible very-low-amplitude component. The end-effect, appearing because no extension was done, is localized at the edges and consists of increasing amplitudes (by a factor 2) and increasing or decreasing frequencies (by 20% to 50%) for the two modes.

similar. The slight differences are seen at the ends of the mode depending on the choice of the extension option. For instance, the sinusoidal extension technique is very good to alleviate the end-effect if a low damping coefficient is used, $\kappa < 0.01$. Then without extension the results are also good, as can be seen on Figure 3.10. However, the anti-symmetric mirror imaging technique does not give a good Hilbert spectrum. In fact, for this kind of signals, this extension method is not good because the extension is not symmetrical with respect to the mean of the signal. This is due to the fact that the signal is terminated by two extrema (see Figure 3.1 (c)).

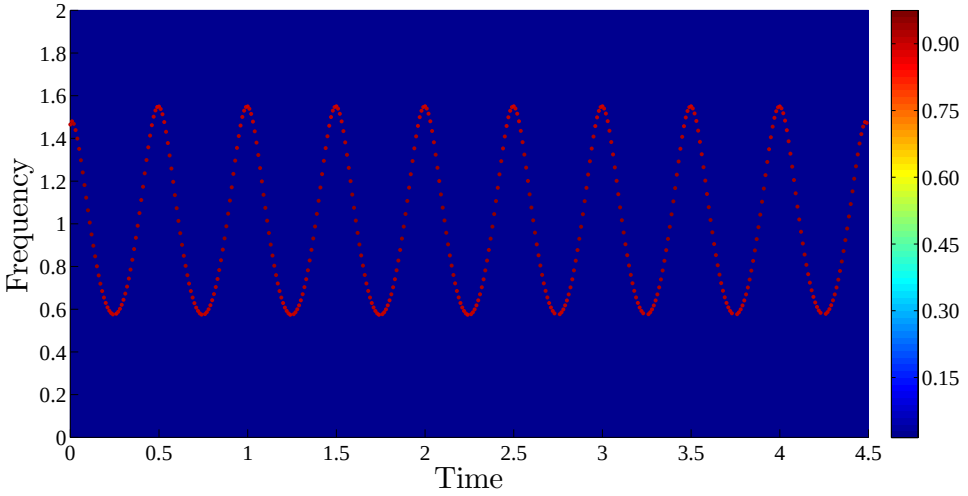

Figure 3.10: Hilbert spectrum of the frequency-modulated signal (s3) with the parameters: `EMD([0:0.01:9;s3],1,-)` and `HT([t3;s3],1,0)`. The first extension technique used here has effectively mitigated the end-effect.

3.2.4 Amplitude-step signal

Empirical mode decomposition: In this study of a non-stationary signal, we can see the power of the HHT, which is able to perform a very good analysis of the amplitude step occuring at time $t = 0$. Table C.4 in Appendix C.4 confirms the excellent results of the decomposition (displayed on Figure 3.11) with very low indexes of orthogonality $IO < 0.011$. Moreover, we can observe that strict stopping criteria increase the number of sifting iterations per IMF as well as the

number of IMFs in a few cases, whereas higher thresholds give opposite results.

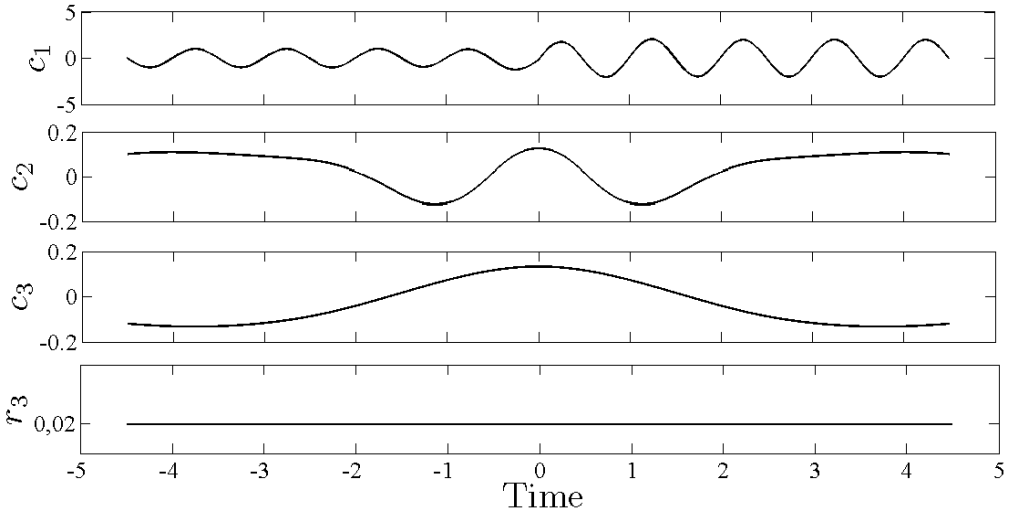

Figure 3.11: IMFs of the amplitude-step signal (s4) with the parameters: `EMD([-4.5:0.01:4.5;s4],2,[0.1,1,0.1])`. The anti-symmetric mirror imaging technique used to extend the IMFs before applying the Hilbert transform has successfully alleviated the end-effect in this case.

Hilbert spectrum: As depicted in Figure 3.12, three IMFs are produced by the EMD algorithm: the first IMF accounts for the step in amplitude, and the last two, which have very low amplitudes ($\overline{a_2}, \overline{a_3} \approx 0.17$), are created because of the brutal change in amplitude and their frequencies actually mark the time of the step. Another important remark is that, for the first time, extension option 2, which has been used to compute the Hilbert spectrum presented in Figure 3.12, effectively mitigates the end-effect at the edges of the signal. The main difference with the three previous signals is that this signal ends at its mean value (see Figure 3.1 (d)), thus allowing the anti-symmetric extensions to be symmetric with respect to the mean value. As we have already found, this point is very important in order to minimize the end-effect. Finally, Table C.4 also shows that the indexes of component separation are consistently very good between the first and second IMFs: $\overline{ICS_1} > (0.86, 0.87)$ for all the cases. However, the study of the instantaneous ICS (see Figure 3.13), which is more accurate in the case of non-stationary signals, actually shows a drop of approximately 2 points of the

first component of ICS_1 between $-1.55 < t < 1.55$. This is clearly due to the occurrence of the step of amplitude, which has induced some Gibbs phenomenon between $t \approx -1.5$ and $t \approx 1.5$.

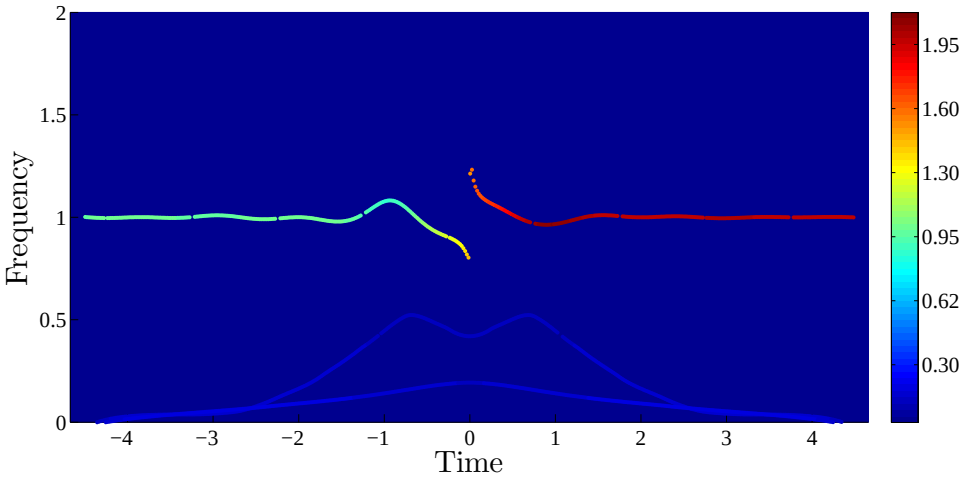

Figure 3.12: Hilbert spectrum of the amplitude-step signal (s4) with the parameters: `EMD([-4.5:0.01:4.5;s4],2,[0.1,1,0.1])` and `HT([t4;s4],2,1)`. The anti-symmetric mirror imaging technique used to extend the IMFs before applying the Hilbert transform has successfully alleviated the end-effect in this case.

3.2.5 Frequency-shift signal

Empirical mode decomposition: Likewise the frequency-modulated signal, the frequency-shift signal is analysed by the Hilbert transform without being decomposed by the sifting process. According to Table C.5 in Appendix C.5, there is only the case with the fourth end-point option and the strictest stopping criterion which differs by producing two more spurious IMFs.

Hilbert spectrum: Overall, the Hilbert spectra are similar whatever the end-point option. The only differences are due to the choice of the extension option: some end-effect is seen with the first one; otherwise the second and third one can alleviate the problem of increase in amplitude and frequency at the edges. However, from Figure 3.14 we can notice that some frequency waves still propagate toward the interior of the signal. Moreover, these waves are more important after the frequency shift than before. No explanations have been found about this

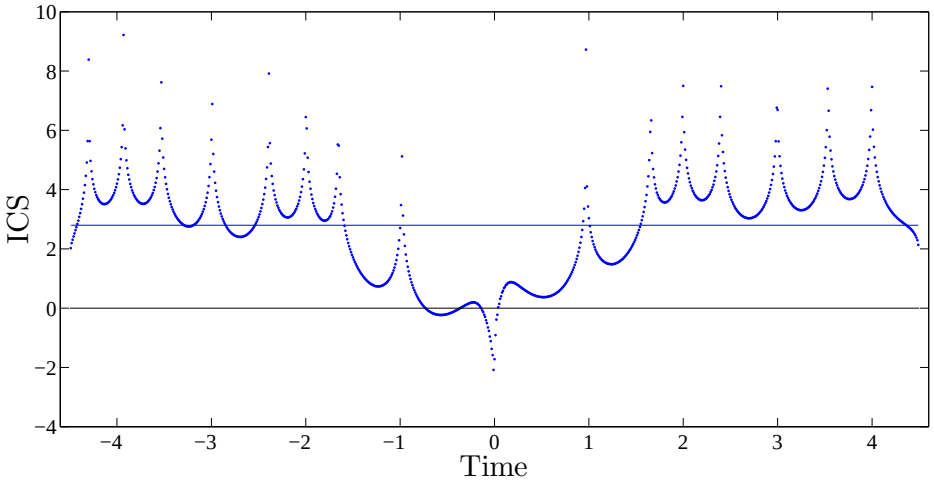

Figure 3.13: Instantaneous index of component separation $ICS_1(1)$ and its mean (blue line) for the amplitude-step signal. This index decreases sharply by approximately 2 points between $t = -1.55$ and $t = 1.55$. It is due to the step of amplitude, and it shows that the Hilbert transform has difficulties to handle brutal changes in amplitude.

none-symmetric phenomenon. Finally, comparing this case with the amplitude-step signal, it can be deduced that the Hilbert transform seems to handle frequency changes better than amplitude changes. In fact, the Gibbs phenomenon around $t = 0$ is much less important in the Hilbert spectrum of the frequency-shift signal.

3.2.6 Conclusions on the five-signal study

First, the study of these five test signals showed that the end-point options could affect the decomposition of the signal. The first option was different from the others because it had to be used with a stopping criterion that only demanded the first IMF-condition. So, it could lead to different interpretations of the signal: for the two-component signal it failed to retrieve the two components; but for the amplitude modulated signal, while all other options found two IMFs, this option resulted in a much more intuitive representation of the phenomenon. The second and third options were very similar, as it could have been expected according to their respective implementation, and no differences could be observed in the Hilbert spectrum. In fact, there were only small discrepancies in

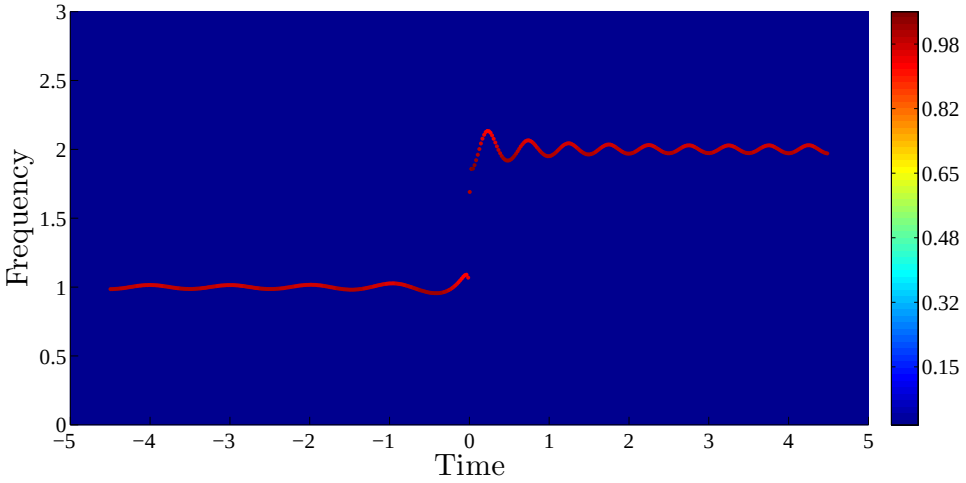

Figure 3.14: Hilbert spectrum of the frequency-shift signal (s5) obtained with the following parameters: `EMD([-4.5:0.01:4.5;s5],2,[0.01,0.1,0.01])` and `HT([t5;s5],2,1)`. The anti-symmetric mirror imaging technique used to extend the IMFs before applying the Hilbert transform has somewhat alleviated the end-effect. However, we can still observe frequency waves in the second part of the signal.

the way their respective IMFs terminated. The fourth end-point option was actually more difficult to monitor because it had several parameters such as the length of extrapolation of the sinusoidal curve, the length of averaging and the damping coefficient. Good results were found with $N_{epl} = N$, $N_{avg} = 0.3N$ and $\kappa \leq 0.01$. However, we also saw that this option was not adapted with low thresholds since it created many spurious IMFs when the stopping criterion was strict: `thresholds=[0.01,0.1,0.01]` and sometimes `thresholds=[0.05,0.5,0.05]`.

Second, the extension option which best alleviated the end-effect was the third one, the extension with a damped sinusoidal curve, when very low damping coefficients were used: $\kappa \leq 0.005$. Without extension, we could observe fallacious large increase or decrease in amplitude and frequency at the edges of the modes in the Hilbert spectrum. The second one, the anti-symmetric mirror imaging technique was found to be counter-productive when the IMFs terminated on extreme points. In other words, large frequency waves propagated from the edges toward the center of the mode. However, when the components

terminated on their mean value, so that the extensions could be also symmetrical with respect to the mean, this option was effective to mitigate the end-effect.

Third, the first three quantitative criteria—the index of orthogonality, the number of IMFs and the number of iterations per IMF—were very useful to assess the performances of the EMD algorithm. The last one, the index of component separation, was a good indicator of the frequency resolution and bandwidth separation of the modes in the Hilbert spectrum. Moreover, we found that for stationary signals its time average was sufficient. However, its instantaneous form was much more accurate to evaluate non-stationary phenomena.

Finally, the optimal combinations of the different implementation options for these five signals can be found in Table G.1 in Appendix G. These results are based on the conclusions discussed above.

3.3 Study of the length-of-day data

In the previous section we have seen the influence of the end-point option and the stopping criterion on the decomposition of simple test signals. Moreover, we have found that the index of orthogonality, the number of IMFs, the number of iterations per IMF and the index of component separation could provide valuable information about the results of the EMD and the Hilbert spectrum. In the study of length-of-day data, we use these indicators to perform a systematic investigation of the choices of the end-point option and the thresholds of the fourth stopping criterion. The aim is to establish the set of control parameters which leads to a good decomposition and to a meaningful time-frequency-amplitude representation.

The daily length-of-day data set (LOD data set) used for this study were originally constructed by Gross (2001) [20] and can be found on-line in the file 'comb2000_daily.eop'. Gross (2001) [20] gave a detailed study of the physical phenomena at the origin of the daily variations of the length of the day. Then,

Huang et al. (2003) [28] performed a thorough analysis of this data set and showed that the HHT could bring a complete understanding of the physics represented in this signal[1]. For example, influence of the tides, semi-annual cycles, annual cycles and longer-period cycles, such as the periodic El Niño event, were identified in the IMFs. They also demonstrated the importance of the intermittency test to remove aliases caused by mode mixing. Finally, they computed the mean Hilbert spectrum and calculated the confidence limit for the HHT by varying the second stopping criterion presented in Section 2.2.5. Therefore, similar to their analysis of the second stopping criterion, we use the LOD data to study the second, third and fourth end-point options, as well as the thresholds and tolerance of the fourth stopping criterion.

The study of the LOD data set proceeds as follows: for each of the second, third and fourth end-point options (epo) the signal is decomposed with the EMD method. In each case, the thresholds and tolerance of the stopping criterion vary from $(\theta_1 = 0.03, \alpha = 0.03)$ to $(\theta_1 = 0.3, \alpha = 0.3)$ with the steps $\Delta(\theta_1) = \Delta(\alpha) = 0.005$; the second threshold is maintained proportional to the first one with $\theta_2 = 10\theta_1$, as prescribed by Rilling et al. (2003) [42]. Actually, the thresholds cannot be set lower due to problems of convergence during the sifting process. Moreover, a limit of 500 iterations per IMF is set, above which the second and third conditions of the fourth stopping criterion are relaxed. In addition, this study is conducted two times: firstly without the invoking intermittency test; and secondly with the intermittency test using the criteria `n1=[4,0^3,45^2,-1]` [2]. In each case defined by a triplet $C_i = (epo_i, \theta_{1,i}, \alpha_i)$, the IO, N_{IMF}, $N_{ite,j}$ and $\overline{ICS_j}$ are computed. For the index of component separation, its time average is chosen because the data can be considered as stationary, then the arithmetical mean over all the IMFs is computed so as to give an overall evaluation of the Hilbert spectrum in a single graph, it is designated as

[1] A decomposition of the LOD data in IMFs can be found in Appendix D.1. The results obtained in this study are compared to the results obtained by Huang et al. (2003) [28].

[2] These intermittency criteria were suggested by Huang et al. (2003) [28].

$mean(\overline{ICS})$. The EMD algorithm and the Hilbert-transform algorithm presented in Appendix B are used for the computations. The first end-point option is not investigated because we found in the study of the five signals (Section 3.2.6) that this option could not be combined with the fourth stopping criterion. Finally, the extension option used for all the cases to mitigate the end-effect in the Hilbert spectrum is the third one: `HT([(1:1:14232)/365,LOD],3,1)`. Moreover, the modes are extrapolated with a damped sinusoidal curve, with the parameters: $N_{epl} = N$, $N_{avg} = 0.1N$ and $\kappa = 0.0001$. These parameters were chosen according to the conclusions reached in the five-signal study.

3.3.1 Assessing the end-point option, the stopping criterion and the intermittency test

Second end-point option, no intermittency test: As shown on Figure 3.15, four different regions can be distinguished: first, for $\theta_1 \leq 0.035$ the IO is very low at around 0.015. Second, for $0.04 \leq \theta_1 \leq 0.1$ and $\alpha \geq 0.08$ the IO is extremely low, except for $\theta_1 = 0.05$ where the IO shows a short increase. Third, the region for $0.1 \leq \theta_1 \leq 0.17$ is an intermediate area where the highest values are reached for $\alpha \leq 0.11$, and then the index of orthogonality is lower for larger α values except for the IO near $\alpha \geq 0.25$. Fourth, the rest of the space is almost flat with a low index. Overall, this index is relatively low and does not reflect serious problems in the decomposition. From the first graph of Figure 3.16 we can see that the first threshold has a significant influence on the number of IMFs. It can vary from ten IMFs for the strictest stopping criteria ($\theta_1 = 0.3, \alpha \leq 0.08$) to seven IMFs for the area where $\theta_1 \geq 0.1$. In fact, the inspection of the IMFs has revealed that seven IMFs is not enough because the lowest frequency modes are not separated (see Figure D.1 in Appendix D.1). On the other hand, 10 IMFs is also not correct because some modes have been over-decomposed due to too many sifting iterations, as shown on the second graph of Figure 3.16. Finally, with eight and nine IMFs, the differences are not so important and even if we

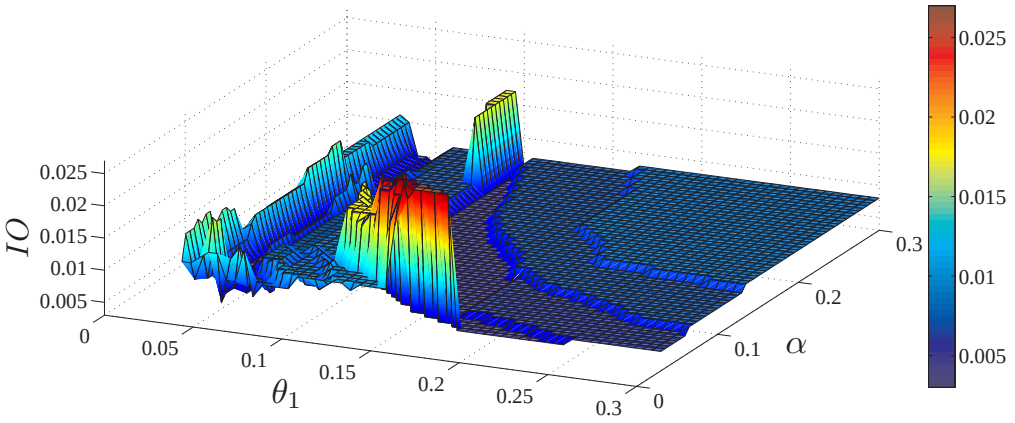

Figure 3.15: Index of orthogonality versus (θ_1, α) for the study of the LOD data with the *second end-point option* and *without intermittency test*.

attempt to correlate these two graphs with the IO results, it is still difficult to discriminate which number of IMFs gives the best decomposition. Moreover, Huang et al. (2003) [28] also found the same results: they decomposed the signal into eight IMFs with the least strict stopping criterion ($S = 2$ and $M = 100$); and, they found nine IMFs with the strictest stopping criterion ($S = 10$ and $M = 100$). Therefore, further analysis is needed to determine how many IMFs best decompose the LOD data.

Third end-point option, no intermittency test: Figure 3.17 shows that the IO is very low everywhere except in two points or small areas at $(0.03 \leq \theta_1 \leq 0.04, 0.03 \leq \alpha \leq 0.04)$ and $(\theta_1 = 0.08, \alpha = 0.03)$ where $IO = 0.2$. Figure 3.18 further reveals that the decomposition is very good almost everywhere with this end-point option. Only eight and nine IMFs have been produced, with a much larger area corresponding to eight IMFs. In addition, the number of iterations is below 200 iterations, which is good for the decomposition of the LOD data, in most cases, except for very strict thresholds $(\theta_1 = 0.03, \alpha \leq 0.11)$ and near $(\theta_1 = 0.03, \alpha = 0.08)$. Therefore, this option seems to be very efficient to decompose the LOD data except for very strict thresholds. Furthermore, these results tend to confirm that the correct number of IMFs is eight rather than nine.

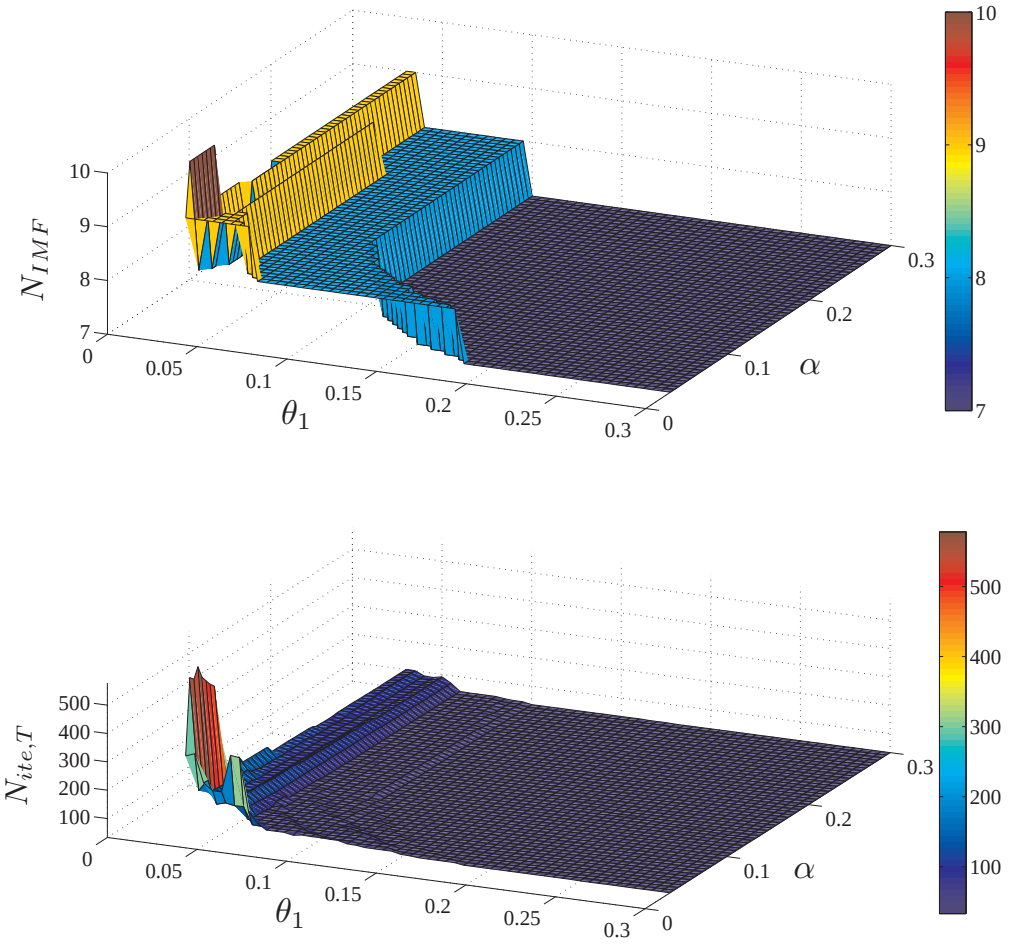

Figure 3.16: Number of IMFs (top) and total number of iterations (bottom) versus (θ_1, α) for the study of the LOD data with the *second end-point option* and *without intermittency test*.

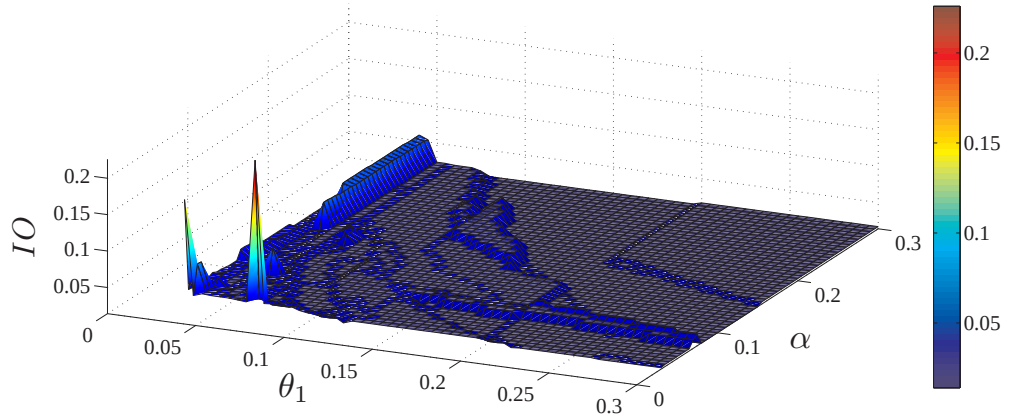

Figure 3.17: Index of orthogonality versus (θ_1, α) for the study of the LOD data with the *third end-point option* and *without intermittency test*.

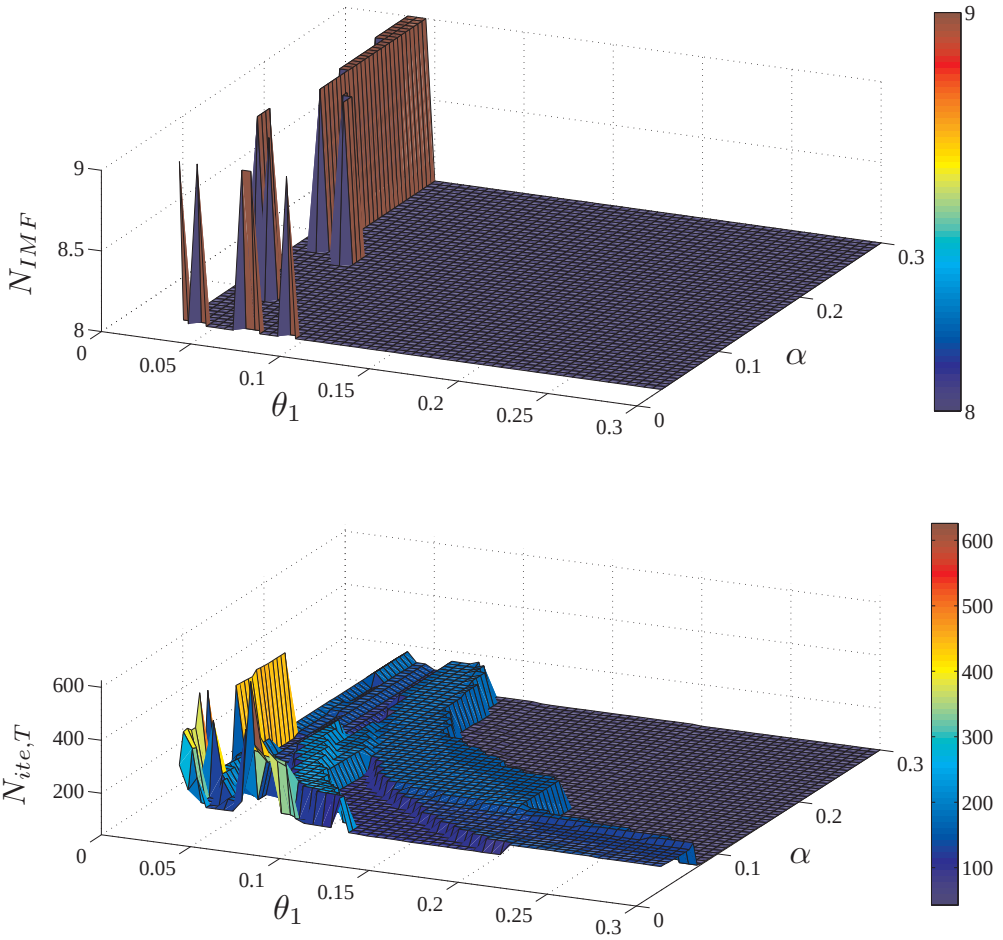

Figure 3.18: Number of IMFs (top) and total number of iterations (bottom) versus (θ_1, α) for the study of the LOD data with the *third end-point option* and *without intermittency test*.

Fourth end-point option, no intermittency test: Figure 3.19 shows excellent results for the index of orthogonality in the case of the fourth end-point option (auto-regressive model) without intermittency test. There are only two relatively high peaks at $\theta_1 = 0.02, \alpha = 0.03$ and $\theta_1 = 0.02, \alpha = 0.17$. In fact, it was possible to reach convergence in the sifting process for $\theta_1 = 0.02$ using this end-point option, except for a few points which are not represented in the graph. The

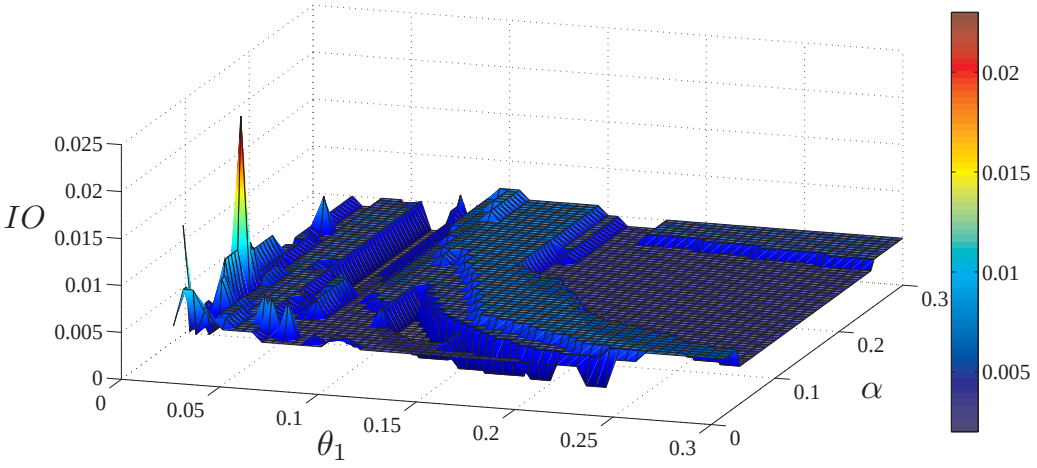

Figure 3.19: Index of orthogonality versus (θ_1, α) for the study of the LOD data with the *fourth end-point option* and *without intermittency test*.

two graphs displayed in Figure 3.20 are very much similar to the same graphs obtained in the case of the second end-point option. Nevertheless, it can be seen that the 9-IMF region covers a much smaller area and only corresponds to very strict thresholds, thus further reinforcing that a decomposition with eight IMFs is the most appropriate. At last, some cases with strict thresholds have produced ten IMFs because of over-sifting. However, from the inspection of the IMFs, it can be seen that some IMFs are actually redundant (see Figure D.2 in Appendix D.1).

Second end-point option with intermittency test: The results of the IO, for the second end-point option (extrema extension technique) *with intermittency test*, are displayed on Figure 3.21. High values are found for $\theta_1 = 0.03$ and $\alpha \geq 0.0225$ with a peak reaching 0.08 for the IO. Then, between $\theta_1 = 0.035$

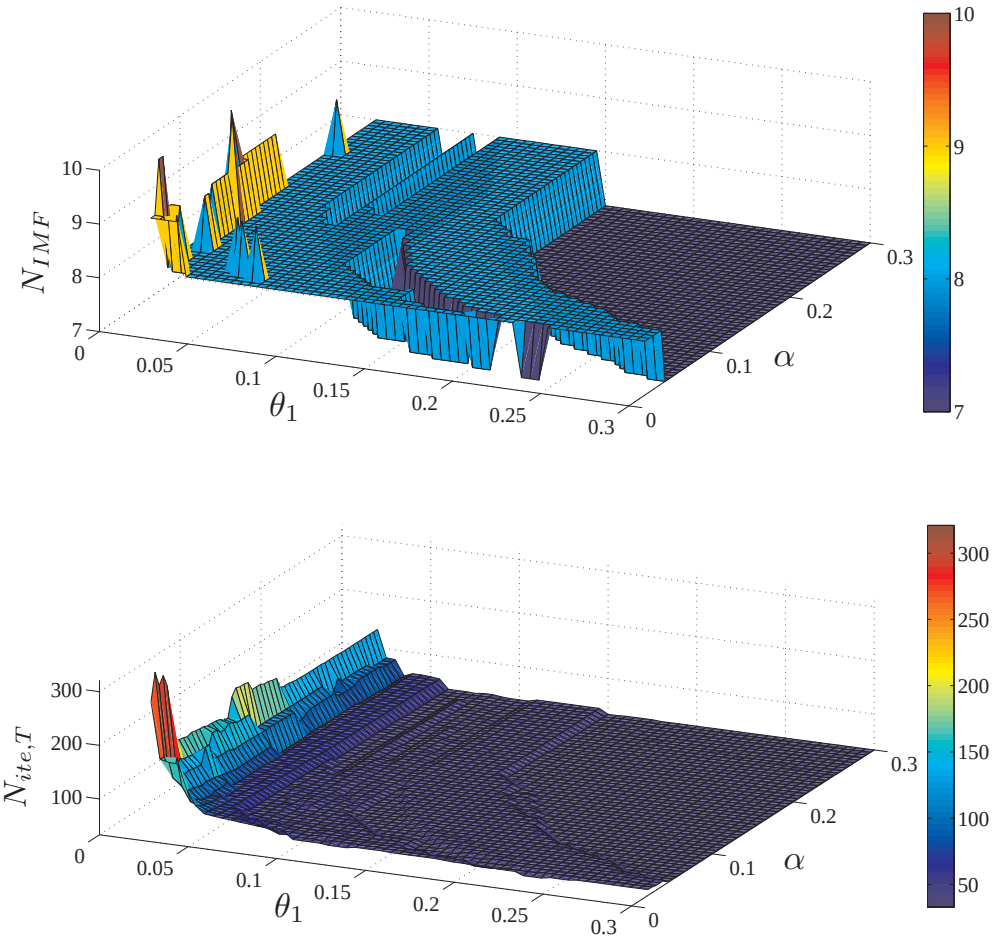

Figure 3.20: Number of IMFs (top) and total number of iterations (bottom) versus (θ_1, α) for the study of the LOD data with the *fourth end-point option* and *without intermittency test*.

and $\theta_1 = 0.09$ the index is very low. At higher thresholds, $0.09 \leq \theta_1 \leq 0.14$ and $\alpha \geq 0.2$, the IO has intermediate values. Finally, the index has very low values at very high thresholds. Moreover, the IO never exceeds the prescribed maximum value of 0.1. Finally, the observation of the IMFs reveals that some of them feature large-amplitude swings at their ends. For example, c_7 of the decomposition at the highest peak $(\theta_1 = 0.03, \alpha = 0.225)$ displays a large swing at one end on a very short distance, less than 1% of the size of the signal (see Figure D.3 in Appendix D.1). Furthermore, c_{11} and c_{12}, the last two IMFs, have also larger-amplitude waves which compensate one another at the begining of the signal. In other words, the sum of c_{11} and c_{12} is approximately zero at the begining; however, the sum of their energy is not zero but twice the energy of each of them. This is a common problem encountered at strict stopping criteria. Next,

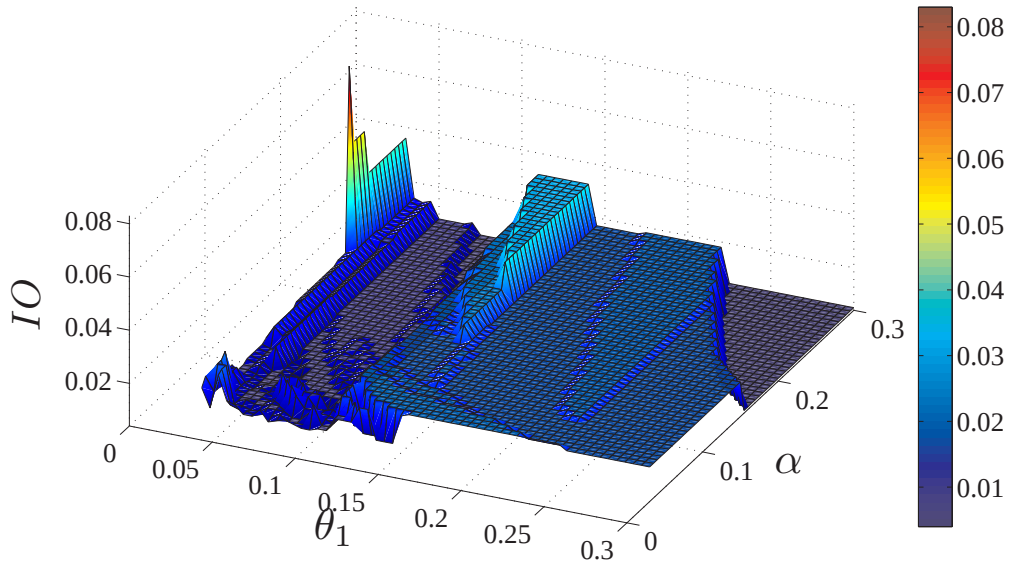

Figure 3.21: Index of orthogonality versus (θ_1, α) for the study of the LOD data with the *second end-point option* and *with intermittency test*.

Figure 3.22 shows that in most cases, eleven IMFs are found, corresponding to the eight IMFs found without intermittency test plus three intermittent IMFs. Ten IMFs are created for very high thresholds, $(\theta_1 \geq 0.23, \alpha \geq 0.17)$. On the other hand, twelve IMFs can be produced if the stopping criterion is very strict, $\theta_1 \leq 0.045$. High numbers of IMFs very well coincide with high total numbers

of iterations, $N_{ite,T} \geq 500$. Actually, further explorations of the number of iterations per IMF reveal that when $N_{ite,T} \geq 500$, for one IMF, generally c_7 which is after the third intermittency test, the sifting process could not converge and the stopping criterion had to be relaxed after the 500^{th} iteration. Normally, approximately 100 iterations are needed to find all the IMFs when the intermittency test is invoked.

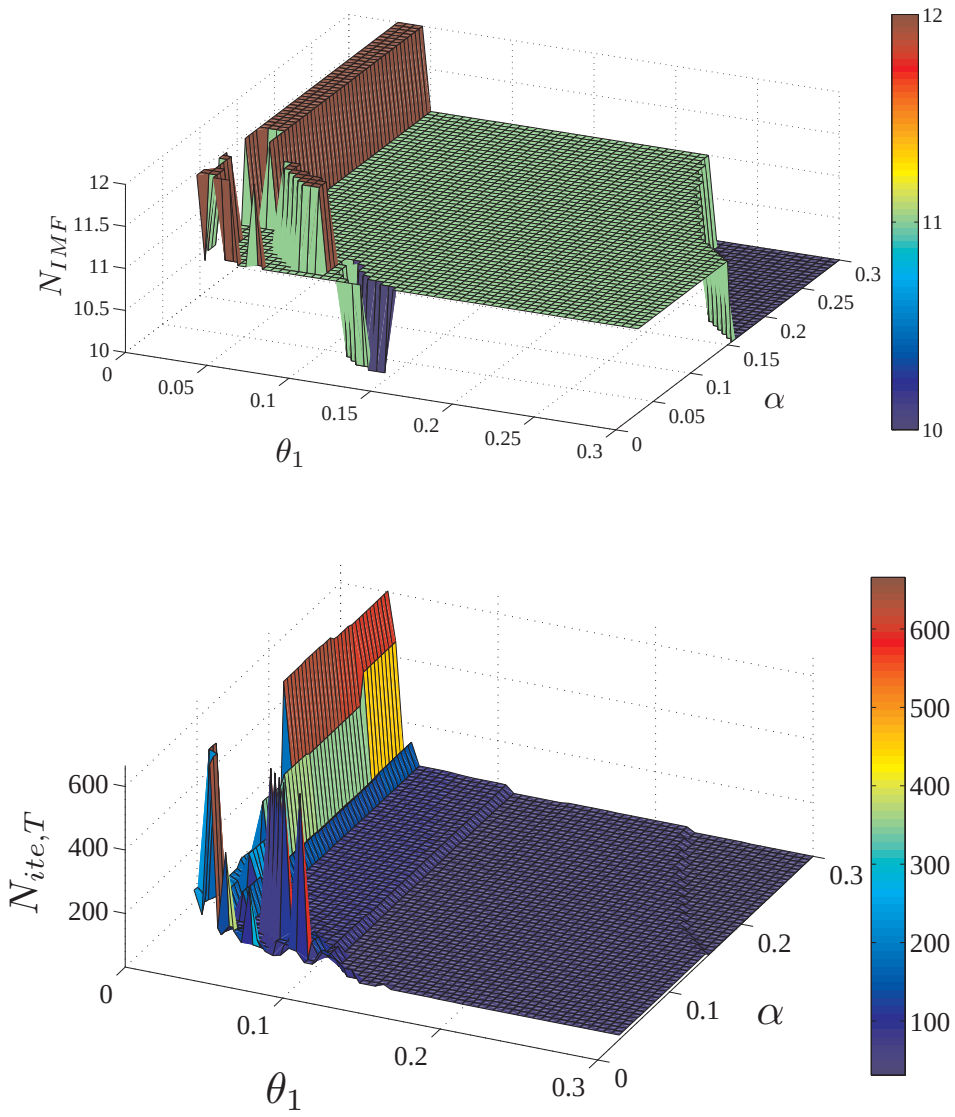

Figure 3.22: Number of IMFs (top) and total number of iterations (bottom) versus (θ_1, α) for the study of the LOD data with the *second end-point option* and *with intermittency test*.

Third end-point option with intermittency test: In three regions the index of orthogonality reaches high values, $IO \geq 0.08$, as shown on Figure 3.23. First, near ($\theta_1 = 0.06, \alpha = 0.035$); second, for $0.05 \leq \theta_1 \leq 0.055$ and $\alpha \geq 0.095$; and third, in a vast area with $0.135 \leq \theta_1 \leq 0.19$ and $\alpha \geq 0.08$. In other regions, this index is lower than 0.04. Compared to the results without intermittency test, the index of orthogonality is higher and the high values cover larger areas. Next, similar to the results found with the second end-point option, Figure 3.24

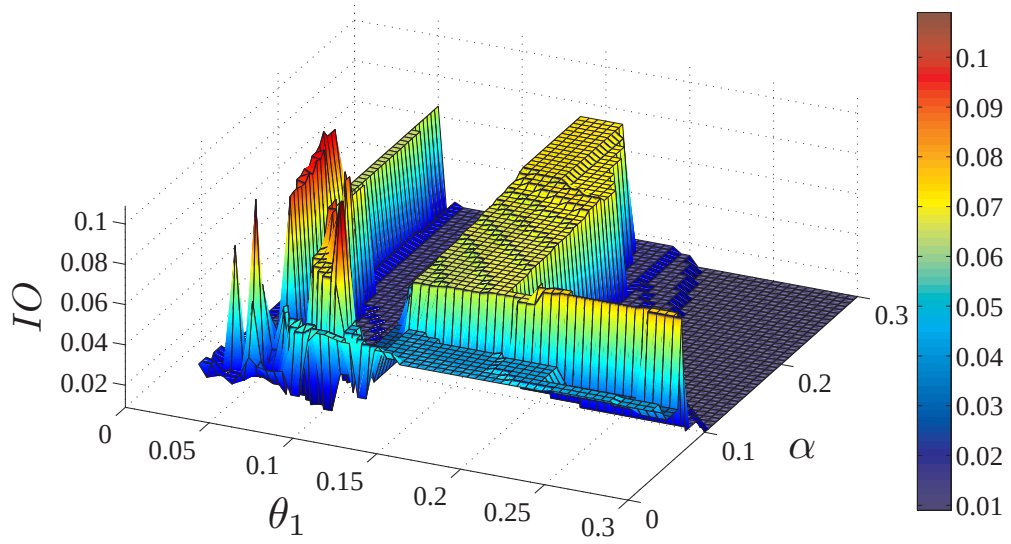

Figure 3.23: Index of orthogonality versus (θ_1, α) for the study of the LOD data with the *third end-point option* and *with intermittency test*.

shows that the number of IMFs is either ten, eleven or twelve, varying from the least strict stopping criteria to the strictest ones. Moreover, the total number of iterations shows that the maximum limit of 500 iterations is reached for low thresholds, $\theta_1 \approx 0.055$ and $\alpha \geq 0.17$.

Fourth end-point option with intermittency test: The results of the IO displayed on Figure 3.25 are excellent in the case of the fourth end-point option (auto-regressive model) with intermittency test. The IO is below 0.01 everywhere except at one peak where it is approximately 0.18. According to this graph, the fourth end-point option seems to be much better adapted to decom-

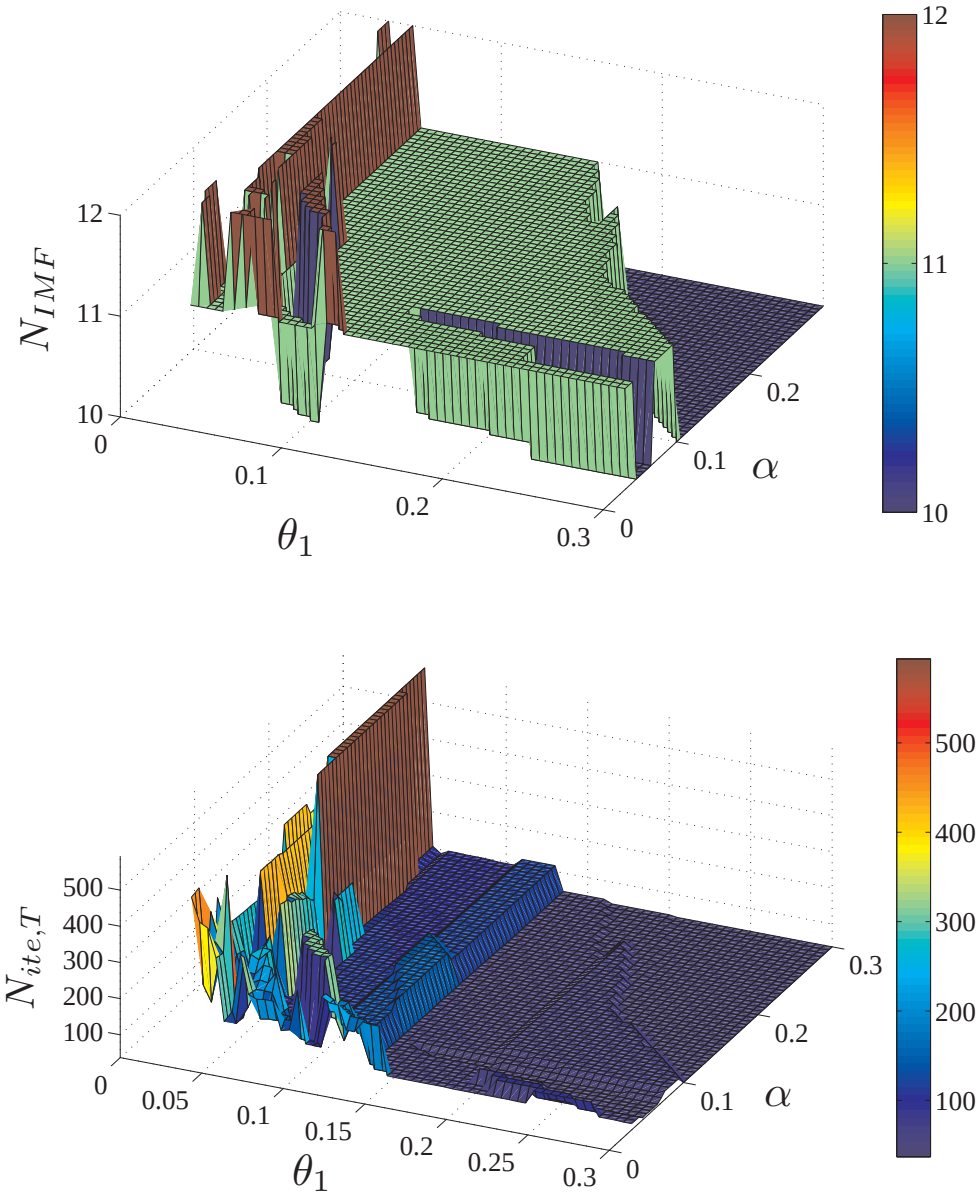

Figure 3.24: Number of IMFs (top) and total number of iterations (bottom) versus (θ_1, α) for the study of the LOD data with the *third end-point option* and *with intermittency test*.

pose the LOD data, with the intermittency test, than other end-point options (see Figure D.4 in Appendix D.1). Next, as can be seen on Figure 3.26, the re-

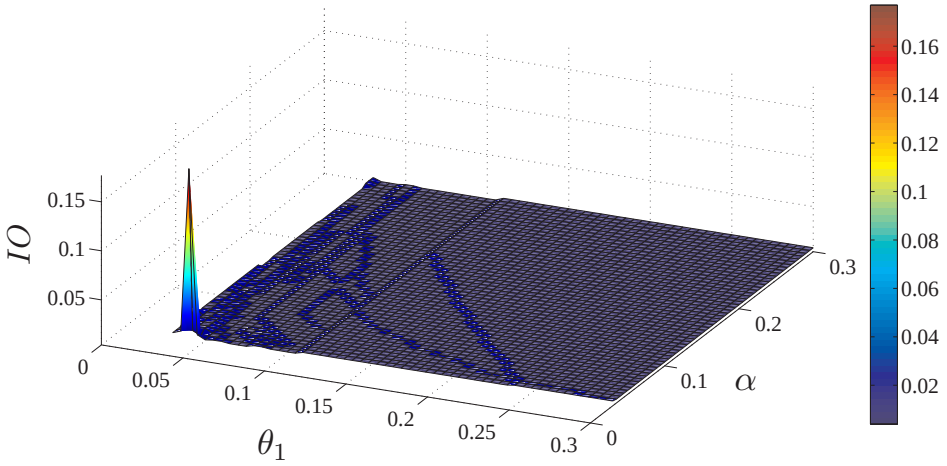

Figure 3.25: Index of orthogonality versus (θ_1, α) for the study of the LOD data with the *fourth end-point option* and *with intermittency test*.

sults for the number of IMFs and the number of iterations confirm the excellent results found with the index of orthogonality. Clearly, the LOD data should be decomposed in ten or eleven IMFs (eight IMFs, two or three intermittent IMFs and one residue) and in approximately 100 iterations. Only very low thresholds, $\theta_1 = 0.035$, can have a problem of convergence and actually over-decompose the data.

Index of component separation, no intermittency test: Figure 3.27 shows the results of the mean of the average index of component separation for the second (top, first graph), third (middle, second graph) and fourth (bottom, third graph) end-point options without intermittency test. First, we can notice that, overall, the index is low in the three graphs, $mean(\overline{ICS}) \leq 0.7$, except for a few points in the second graphs $(epo = 3, \theta_1 = 0.03, 0.1 \leq \alpha \leq 0.2)$. Second, it is clear that for the second and fourth end-point options, the $mean(\overline{ICS})$ is better when θ_1 is low, regardless of the value of the tolerance. The limit is found at approximately $\theta_1 = 0.1$. This suggests that a strict stopping criterion is more likely to separate the modes in the frequency space. Finally, we can also observe that the index is

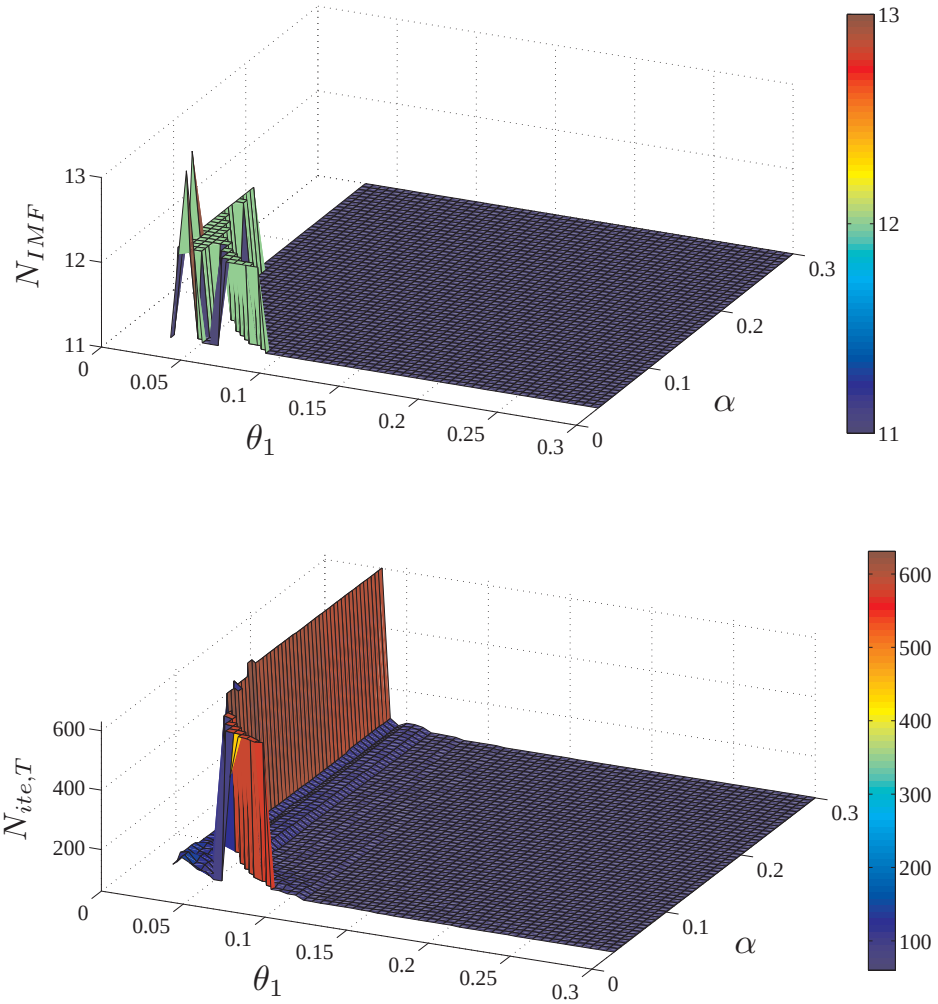

Figure 3.26: Number of IMF (top) and total number of iterations (bottom) versus (θ_1, α) for the study of the LOD data with the *fourth end-point option* and *with intermittency test*.

consistently higher in the case of the fourth end-point option than in the other cases, by approximately 0.15 compared to the first graph, and 0.4 to 0.6 compared to the second graph. In fact, a comparison of the IMFs produced by each end-point option shows that the fourth one can better alleviate the end-effect than the other two options, which might explain why its index of component separation is consistently higher.

Index of component separation with intermittency test: Figure 3.28 presents the results of the mean of the average index of component separation for the second (top), third (middle) and fourth (bottom) end-point options with intermittency test. In these results, the fifth component of the $\overline{ICS}$ has not been taken into account in the computation of the mean value because, corresponding to two intermittent IMFs ($c_{5,int}$ and $c_{6,int}$), this index was almost always infinite for all the cases: $\overline{ICS_5(C)} = (-\infty, -\infty)$. This is due to the fact that the intermittent IMFs $c_{5,int}$ and $c_{6,int}$ are almost always zero over the time span. Firstly, we can observe that the results on Figure 3.28 are unequivocally much better than the results without intermittency test. Actually, in this case, the $mean(\overline{ICS})$ is almost always higher than the highest value found in the previous graphs. Therefore, it demonstrates the importance of the intermittency test to prevent mode mixing in the IMFs and to separate the bandwidths of the modes in the frequency spectrum. Secondly, we can notice that the three graphs are similar in values and variations: for low thresholds, $\theta_1 \leq 0.12$, the index is approximately $mean(\overline{ICS}) = 1$; then, for higher thresholds, $\theta_1 \geq 0.12$, the index is approximately $mean(\overline{ICS}) = 1.5$. Finally, it can be seen that the transition between the two regions is very abrupt.

3.3.2 Remarks and discussion

The results given by the index of orthogonality, the number of IMFs, the total number of iterations and the index of component separation for the LOD data

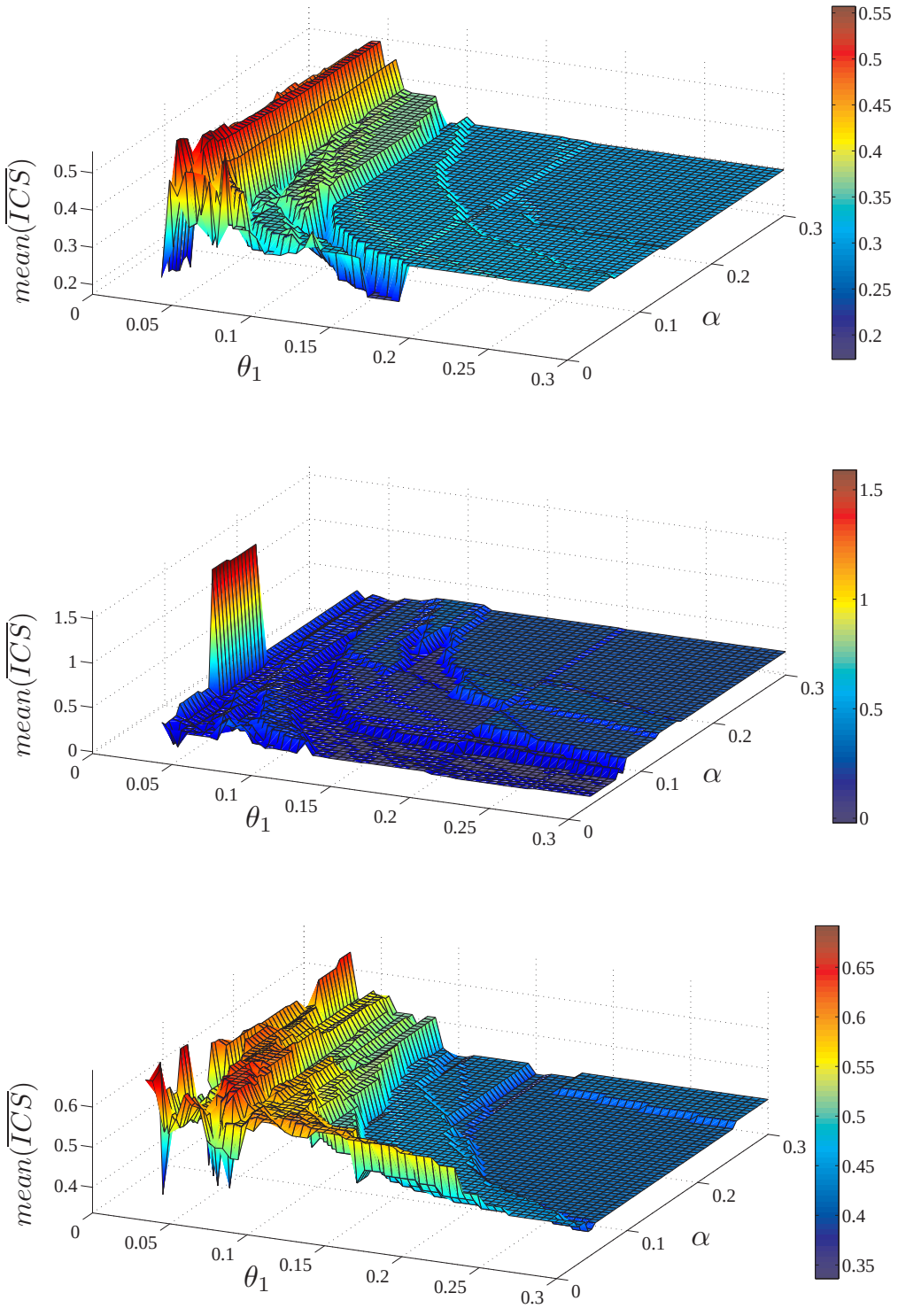

Figure 3.27: Results of the mean of the average index of component separation, $mean(\overline{ICS})$, versus (θ_1, α) for the study of the LOD data with each end-point option and *without intermittency test*: top, second end-point option; middle, third end-point option; bottom, fourth end-point option.

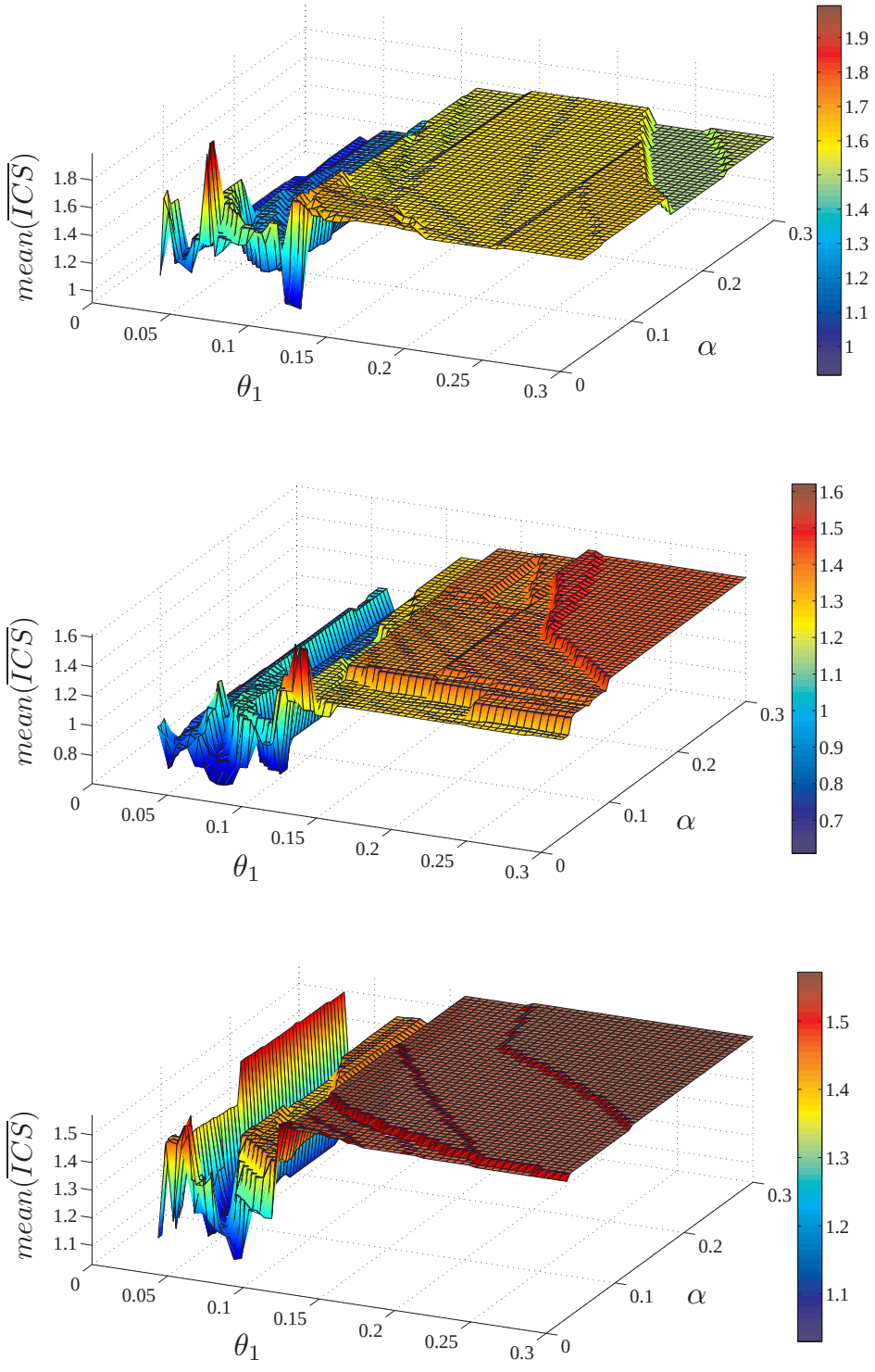

Figure 3.28: Results of the mean of the average index of component separation, $mean(\overline{ICS})$, versus (θ_1, α) for the study of the LOD data with each end-point option and *with intermittency test*: top, second end-point option; middle, third end-point option; bottom, fourth end-point option.

analysed with the second, third and fourth end-point options and with various thresholds for the fourth stopping criterion are extremely rich. We have found that the data set is better decomposed in eight IMFs and one residue, or eight IMFs, three intermittent IMFs and one residue when the intermittency test is invoked, and that the total number of iterations should not exceed 80 or 100 iterations. The index of orthogonality is very good for all the end-point options without intermittency test except for low thresholds ($\theta_1 \leq 0.04$). With the intermittency test, the best regions were at either intermediate thresholds $0.04 \leq \theta_1 \leq 0.12$ or at very high thresholds $\theta_1 \geq 0.22$. However, the problem encountered at very high thresholds is an incomplete decomposition of the signal, only seven IMFs are found. Then, we have found that though the intermittency test tends to increase the IO in some regions (low thresholds or high thresholds), it gives considerably better results for the index of component separation. Therefore, the intermittency test is efficient to reduce mode mixing. Finally, comparing the three end-point options, it can be concluded that the last one, the auto-regressive model, gives the best results and is efficient to alleviate end-effect, except for very low thresholds. Therefore, this option is recommended with the following options: a length of extension equals to the signal size ($N_{epn} = N$), a length of averaging approximately equals to one-tenth of the signal size ($N_{avg} = 0.1N$) and a very low damping coefficient $\kappa \approx 10^{-4}$.

The main limitation regarding these indexes is that they cannot give a very precise assessment of the decomposition or the Hilbert spectrum because they average the results over the time span and over all the IMFs. For example, it cannot show which IMF is actually affected by end-effect or the ones that are affected by mode mixing. In fact, a somewhat finer analysis could be performed by computing the number of iterations per IMF, the index of orthogonality between two IMFs and the instantaneous index of component separation per IMF. However, due to the great number of cases involved in this study, it was not possible to analyse in detail each of these indicators.

Then, we can wonder whether the results of the indexes in the space (θ_1, α) are actually accurate and whether or not they can depend on the fineness of the mesh, $\Delta(\theta_1) = \Delta(\alpha) = 0.005$. First, we can remark that the results are exact at least at the point where they are calculated. Second, the continuity of the indexes is a delicate question since it very much depends on each step of the sifting process. In our understanding, the functions IO and $\overline{ICS}$ should be continuous only in subsets of the space. In fact, discontinuities clearly exist. For example, at a change of number of IMFs, the discontinuity is very important. Then, the occurrence of large swings at the ends of the IMFs is another example of disruptions in the space (θ_1, α). Their appearance and their disappearance are always sudden and depend on the calculation of the cubic spline interpolations. In other words, a slight variation in the stopping criterion can provoke a completely different interpolation of the extrema. Third, according to incomplete computations obtained with a much finer grid, $\Delta(\theta_1) = \Delta(\alpha) = 0.001$, after several days, the results for the second end-point option were identical, thus asserting that the finesse of the mesh is not very important. In fact, the borders delimiting each continuous region of similar values is a much more important information.

Another important point is that the actual value of the indicators is very useful through comparisons of the results obtained from the same signal. However, with different time-series the order of magnitude of the indexes may change, thus making difficult the comparisons between various signals. Finally, these indicators should never be analysed separately from the IMFs and the Hilbert spectrum. It is only the inspection of all these elements that can give the most comprehensive and the clearest evaluation of the results given by the HHT.

3.3.3 Mean marginal spectrum, confidence limit and deviation

The mean marginal spectrum of the LOD data and its confidence limit are computed in two different cases: with and without intermittency test. It means

that the mean spectrum $\overline{h}$, and its standard deviation σ_h, are computed with the results from all the different decompositions with the three end-point options and the 3025 various stopping criteria (as an illustration, the results of eleven individual cases, the mean marginal spectrum and its 95% confidence limit can be found in Figure D.5 Appendix D.2). Then, the sum over frequency of the squared deviation between each case—designated by $sd(C_i)$ where C_i is a combination of an end-point option and a stopping criterion—and the mean marginal spectrum is calculated according to the method presented by Huang et al. (2003) [28],

$$sd(C_i) = \sum_{\omega} \left(h(\omega, C_i) - \overline{h(\omega)} \right)^2. \tag{3.6}$$

All the graphs presented on Figures 3.29 (*without intermittency test*) and 3.30 (with intermittency test) have the same cumulative squared deviation scale (z-axis, $0 \leq sd(C) \leq 5.5\ 10^{-4}$) to allow direct comparisons between them, but not the same color scale. Figure 3.29 displays the results of $sd(C)$ for the second (top, first graph), third (middle, second graph) and fourth (bottom, third graph) end-point option in the space defined by (θ_1, α) and without intermittency test. As can be seen on the first graph, we can distinguish three different zones: very low thresholds, $(\theta_1 \leq 0.06, \alpha \leq 0.17)$, where the squared deviation is low, $sd(C) \approx 10^{-5}$; a second zone with low and intermediate thresholds, $0.065 \leq \theta_1 \leq 0.17$ where $sd(C)$ has intermediate values or high values, $sd(C) \geq 10^{-4}$ for $\alpha \leq 0.06$; in the rest of the space, the squared deviation is decreasing toward very low values as θ_1 and α increases. It is very interesting to compare these results to the results of the IO on Figure 3.15 with the same parameters. The same peak is found in the second region, and the rest of the space almost does not change. Therefore, it shows that the index of orthogonality gives a good evaluation of the marginal spectrum. Second, the squared deviation with the third end-point option appears to be mostly lower than with the second option. Nevertheless, $sd(C)$ is very high over a small area for $(\theta_1 = 0.03, 0.1 \leq \alpha \leq 0.16)$ with $sd(C) \geq 2\ 10^{-4}$. These high values exactly

correspond to the cases with the highest number of iterations, as presented on Figure 3.18. So, it means that problems of convergence in the sifting process can have an impact on the marginal spectrum. In the third graph, the squared deviation is globally lower than in the first two graphs, with values equal to or lower than 10^{-5}. These slightly better results for the fourth end-point option were also reported when comparing the indexes of component separation on Figure 3.27.

Next, Figure 3.30 depicts the results of the squared deviation for the second (top, first graph), third (middle, second graph) and fourth (bottom, third graph) end-point option with the intermittency test. Once again, the results shown in the first graph are almost identical to the results of the IO for the second end-point option with intermittency test (see Figure 3.21). Apart from the high peak near $(\theta_1 = 0.12, \alpha = 0.03)$, the variations of $sd(C)$ exactly follow the variations of the IO, thus asserting that this indicator is useful to assess the marginal spectrum. Second, the results of the squared deviation for the third end-point option can also be compared to the results of the IO (see Figure 3.23). However, though the comparison is good for low thresholds $\theta_1 \leq 0.12$, it is not anymore for higher thresholds. In fact, the results of the marginal spectrum are more similar to the $mean(\overline{ICS})$ (see Figure 3.28), which is much better for high thresholds than for low thresholds. The last graph shows that the results obtained with the fourth end-point option are clearly the best, similarly to the results of the index of component separation (see Figure 3.28). In this case, the squared deviation is approximately one order of magnitude lower than with the first two end-point options. Therefore, it is strongly recommended to use this option in the HHT algorithm.

3.3.4 Optimal sifting parameters

The four indicators, the index of orthogonality, the number of IMFs, the total number of iterations and the index of component separation, have not only proved to be useful to assess the results of the decomposition and the Hilbert

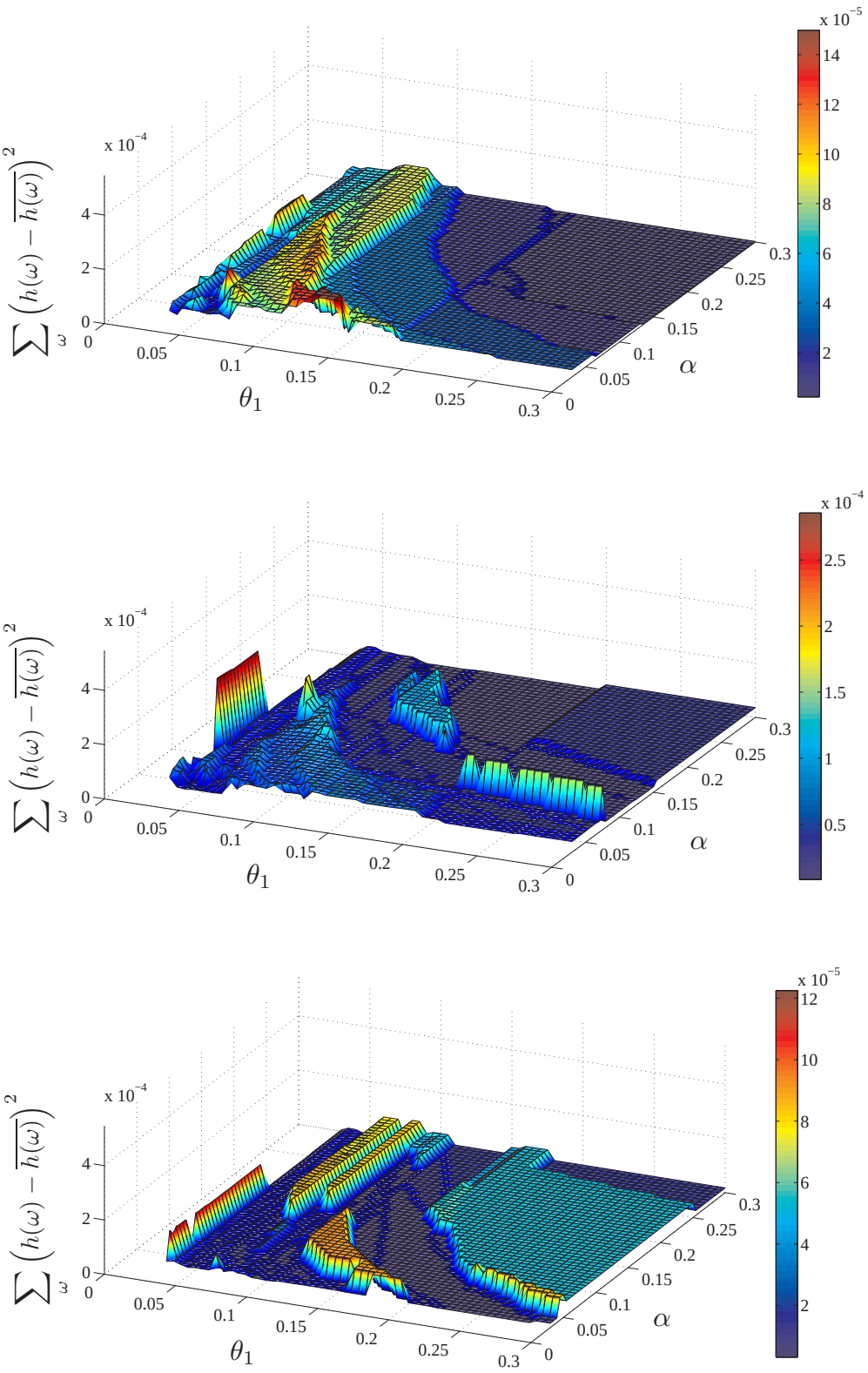

Figure 3.29: Cumulative squared deviation between the mean marginal spectrum and marginal spectra of the LOD data according to the end-point option and *without intermittency test*: top, second end-point option; middle, third end-point option; bottom, fourth end-point option.

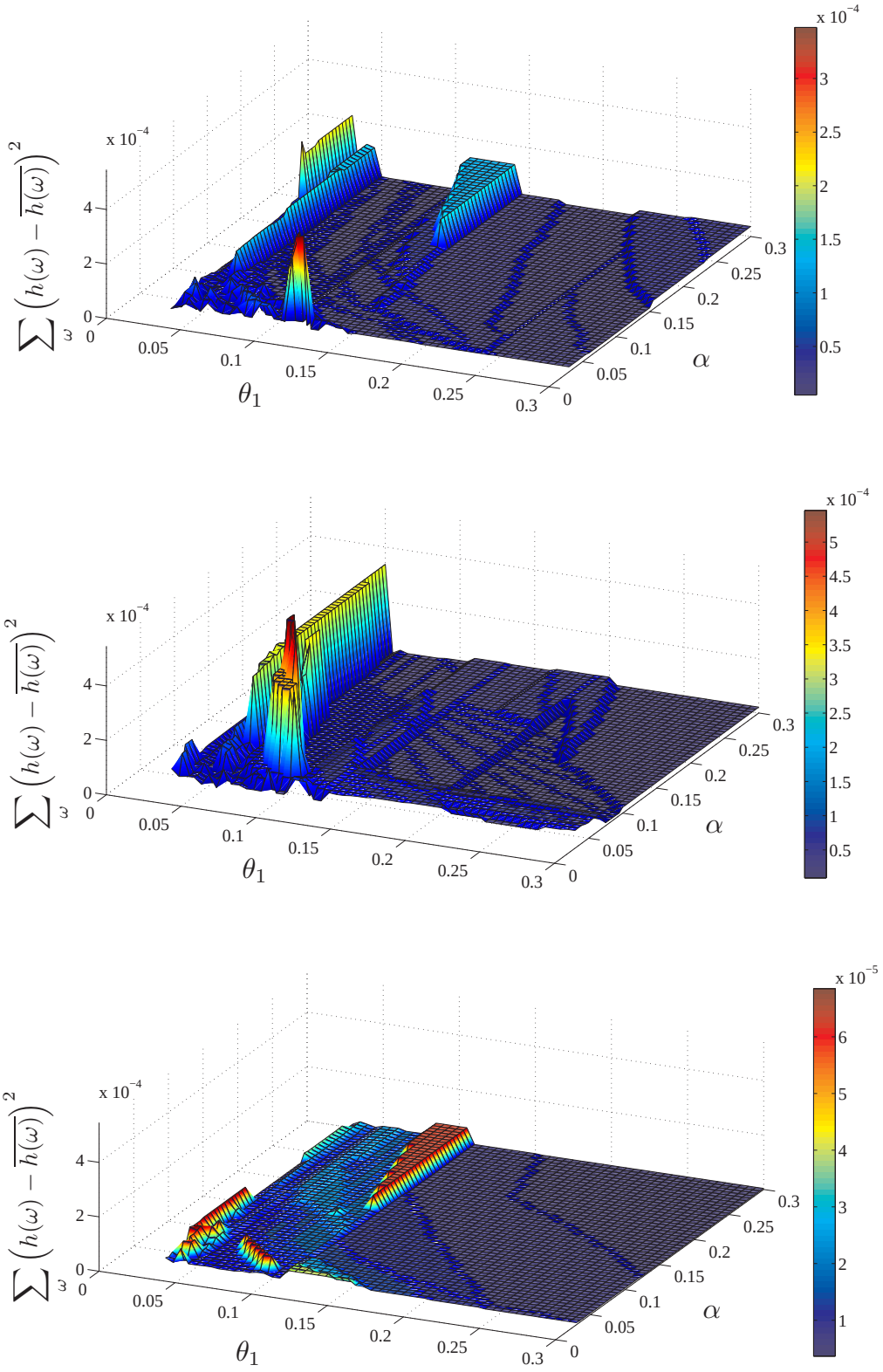

Figure 3.30: Cumulative squared deviation between the mean marginal spectrum and marginal spectra of the LOD data according to the end-point option and *with intermittency test*: top, second end-point option; middle, third end-point option; bottom, fourth end-point option.

spectrum, but they are also well correlated to the squared deviation of the marginal spectrum to the mean marginal spectrum. Therefore, these indexes are recommended to be used together in the HHT algorithm in order to find the control parameters that can perform the best decomposition. From this study of the LOD data, it can be concluded that the auto-regressive model, which is used to extend both the residues in the sifting process and the modes before the application of the Hilbert transform, and not too low thresholds in the fourth stopping criterion, $\theta_1 \geq 0.12$, yield the best results. Finally, even though the efficiency of this set of parameters has not been demonstrated for every type of signals, these parameters are recommended as the starting point of any investigation with the Hilbert-Huang transform.

3.4 Study of vortex-shedding data

A signal measured with a hot-wire sensor located in the wake of a circular cylinder is presently considered. The streamwise component of the flow velocity (see Figure 3.31) has been recorded for a Reynolds number $Re = 105$, which corresponds to the stable 2-dimensional vortex shedding range ($50 \leq Re \leq 150$) described by Roshko (1954) [43]. The aim is to decompose this noisy signal with the EMD algorithm in order to retrieve the vortex-shedding component. Then, applying the Hilbert transform to this mode, the results of the instantaneous vortex-shedding frequency are compared with the results given by the Fourier transform. This phenomenon has been extensively studied since the experiments of Strouhal in 1878, and a thorough introduction can be found in Williamson (1988, 1996a and 1996b) [51] [52] [53].

From Roshko (1954) [43], the theoretical vortex-shedding frequency $F_{s,T}$ at $Re = 105$ can be estimated (with 1% precision) as follows:

$$F_{s,T} = \frac{\nu}{d^2}(0.212Re - 4.5) = 0.07 \text{ Hz} \tag{3.7}$$

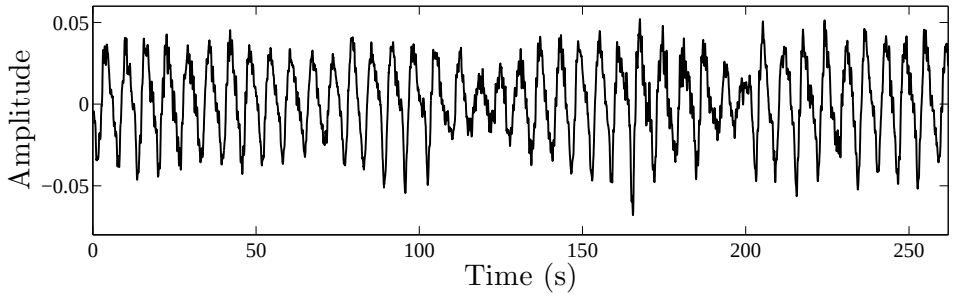

Figure 3.31: Hot-wire measurements in the wake of a circular cylinder at $Re = 105$. The signal has been measured at the centerline of the vortex streets, thus showing the influence of the two vortex rows.

where $\nu = 1.003 \times 10^{-6}$ denotes the kinematic viscosity at 20°C and $d = 15.9$ mm denotes the cylinder diameter. Since the probe is located at the centerline of the vortex streets the measurements are influenced by both rows. Therefore, it is expected to find twice the vortex shedding frequency: $2F_{s,T} = 0.14$ Hz. Before performing the decomposition with the algorithm, a low-pass filter has been applied to the signal in order to reduce the size of the data from 262144 points to 4946 points. In fact, problems of time computation or even convergence in the sifting process were encountered with data sets larger than 40000 points. Hence, the sampling frequency is decreased from 1000 Hz to 18.9 Hz, which is deemed largely sufficient to retrieve the vortex-shedding frequency. Moreover, only noise and turbulent signals can be observed at high frequencies in the signal. Actually, the sampling frequency must be carefully chosen because the success of the decomposition greatly depends on the accuracy with which the extrema are located. As we have seen, the identification of the extrema is a primordial step in the sifting process because the cubic spline interpolations and thus the computation of the mean of the signal directly depend upon it. Rilling et al. (2003) [42] also emphasized using "a fair amount of *oversampling*" for the discrete-time signal, so that extrema be correctly identified.

3.4.1 Optimal parameters for the decomposition of the vortex-shedding signal

A study of the end-point options and the fourth stopping criterion similar to the one performed with the LOD data has been conducted. The results of the four quantitative criteria and the cumulative squared deviation of the marginal spectra from the mean spectrum can be found in Appendix E. Overall, the same tendencies observed with the LOD data are found with this signal, though the actual values of the indicators can differ. For example, with the second and third end-point options the highest values of the IO are found for the strictest thresholds, $\theta_1 \leq 0.04$. Low values of the first threshold also correspond to the largest numbers of IMFs and iterations. The fourth end-point option leads to different results, a few very high peaks for the IO, N_{IMF} and $N_{ite,T}$ can be observed at very low thresholds, $(\theta_1 \leq 0.04, \alpha \leq 0.1)$. Overall, the results of the $mean(\overline{ICS})$ are somewhat better for low and intermediate thresholds. Finally, it can be remarked that the influence of the tolerance on the different indicators is much less important in this study than in the study of the LOD data. From these results, no end-point option clearly appears to be better than others. However, it can be seen that intermediate thresholds ($\theta_1 \approx 0.1, \alpha \approx 0.1$) show relatively better performances.

Table 3.1 presents the results of the quantitative criteria and the squared deviation for the case ($\theta_1 = 0.095, \alpha = 0.125$) with the three different end-point options. At this point, the results produced using the fourth end-point option are the best: the IO and the $sd(C)$ have the lowest values, and the $mean(\overline{ICS})$ is the highest. It can be further noticed that this end-point option gives one more IMF and needs slightly more iterations, in particular for the first IMF. However, a difference of one IMF is not very important. Therefore, the fourth end-point option and these thresholds are chosen to perform the decomposition of the data.

Table 3.1: Results of the index of overall orthogonality IO, the number of IMF N_{IMF}, the number of iteration per IMF $N_{ite,j}$, the mean index of component separation $mean(\overline{ICS})$ and the squared deviation between the marginal spectrum of this case ($\theta_1 = 0.095, \alpha = 0.125$) and the mean marginal spectrum for the vortex-shedding signal (without intermittency test).

End-point option	IO	N_{IMF}	$N_{ite,j}$ $(c_1, c_2, \ldots, c_n)$	$mean(\overline{ICS})$	$sd(C)$
2	0.095	8	$(59, 20, 13, 2, 35, 6, 3, 3)$	0.765	0.611
3	0.207	8	$(59, 20, 17, 3, 8, 6, 2, 3)$	0.254	0.162
4	0.088	9	$(71, 16, 21, 3, 6, 4, 4, 7, 1)$	1.03	0.117

3.4.2 Decomposition of the vortex-shedding signal

Figure 3.32 displays the nine IMFs and the residue obtained from the decomposition of the vortex-shedding signal with the fourth end-point option and ($\theta_1 = 0.095, \alpha = 0.125$). Overall, the decomposition has no major problems in the vicinity of the edges, contrary to the decompositions found with the same stopping criterion but with the other two end-point options—this is another element in favor of using the auto-regressive model. Next, inspecting closer the modes, the first three IMFs seem to contain only noise, whereas the fourth one, which has the largest amplitude, is the signal that we are interested in. However, we can notice some intermittent disturbances in c_4, three between $t = 50$ s and $t = 100$ s and two others before $t = 150$ s and $t = 250$ s. These intermittent signals are actually responsible for the spread of the vortex-shedding signal on c_4 and c_5. As a result, we can identify in c_5 oscillations of similar period as in c_4 and which are located precisely at the timing of the disturbances. Then, it is difficult to know whether c_6 is physically meaningful or whether it also results from the mode mixing. Finally, the last three IMFs have an amplitude approximately one order of magnitude lower than c_4, thus suggesting they might account for low-frequency disturbances occurring during the experiment.

The mode mixing can be removed by invoking an intermittency test before the fourth IMF. An intermittency criterion of $n_1 = 2.7$ s is applied in the sifting process, this criterion has been chosen slightly smaller than the theoretical half-period of the vortex-shedding signal, which is $1/4T_s = 3.55$ s. As can be seen on

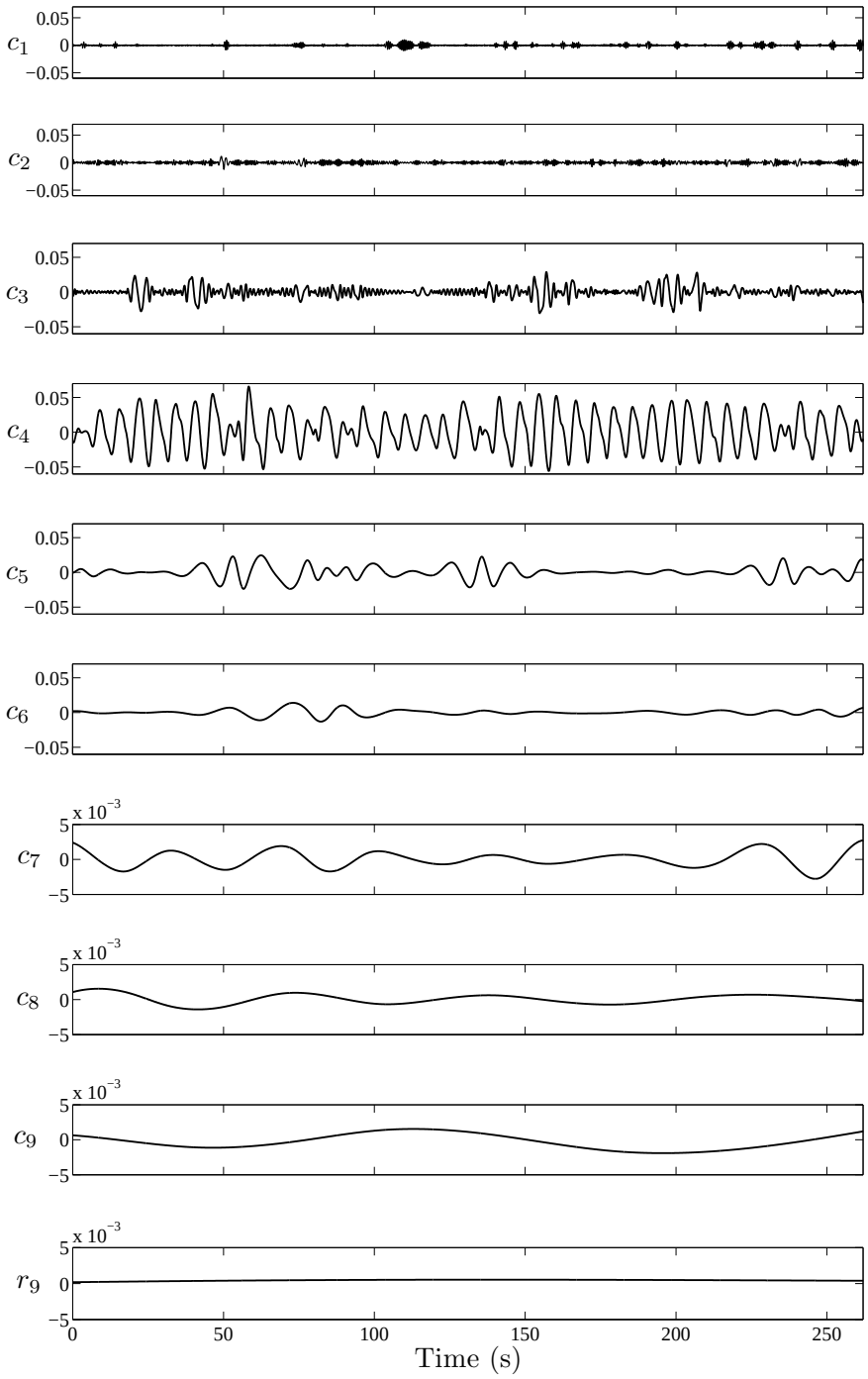

Figure 3.32: The IMF components of the vortex-shedding data (*no intermittency test*) for the case: `EMD([1:0.053:262; V-S 105],4,[0.095,0.95,0.125])` with $N_{epn} = N$, $N_{avg} = 0.2N$ and $\kappa = 10^{-3}$. The fourth IMF, which contains most of the vortex-shedding signal, shows some mode mixing occurring between $t = 50$ s and $t = 100$ s and before $t = 150$ s and $t = 250$ s. Consequently, the oscillations of the signal at these timing are also found in the next IMF, c_5.

Figure 3.33, the intermittent signals (in c_4) have been separated from the vortex-shedding signal (c_5). Moreover, the signal is not anymore spread on two modes but it is fully isolated in a single IMF. Finally, the quantitative criteria for this decomposition have been greatly improved: $IO = 0.073$ and $mean(\overline{ICS}) = 25$. In particular, the index of component separation has considerably increased, thus demonstrating the efficiency of the intermittency test to remove mode mixing. The number of IMF has not changed, which may mean that at least one IMF was created by mode mixing in the decomposition when the intermittency test was not invoked.

3.4.3 Identification of intra-wave frequency modulation

Figure 3.34 (top) depicts the marginal spectrum of the vortex-shedding signal. First, we can see the main broad peak centered on 0.165 Hz and spanning from 0.135 Hz to 0.195 Hz, this is the contribution of IMF c_5. Second, the first three IMFs clearly appear as noise of low amplitude in the high frequency range (the frequency scale has been truncated). Third, four secondary steep peaks can be observed in the low frequencies: at 0.06 Hz, 0.025 Hz, 0.02 Hz and 0.01 Hz. They correspond to the last IMFs and their physical origin has not been completely determined, the first secondary peak, which is near the actual shedding frequency, may be due to an unbalanced influence between the two vortex rows. For instance, the probe could be located not exactly in the middle of the two rows, thus detecting a different level of energy between the two vortex streets. These secondary peaks may also represent low frequency disturbances in the experiments, as we have already suggested. Fourth, the very high spike at 0.004 Hz is a fallacious peak, probably created during the computation of the Hilbert transform, and whose period actually corresponds to the time span. On the other hand, the application of the Fourier transform to the signal (without filtering) gives the Fourier spectrum displayed on Figure 3.34 (bottom). The main steep peak at 0.16 Hz corresponds to the vortex-shedding frequency. Then, there

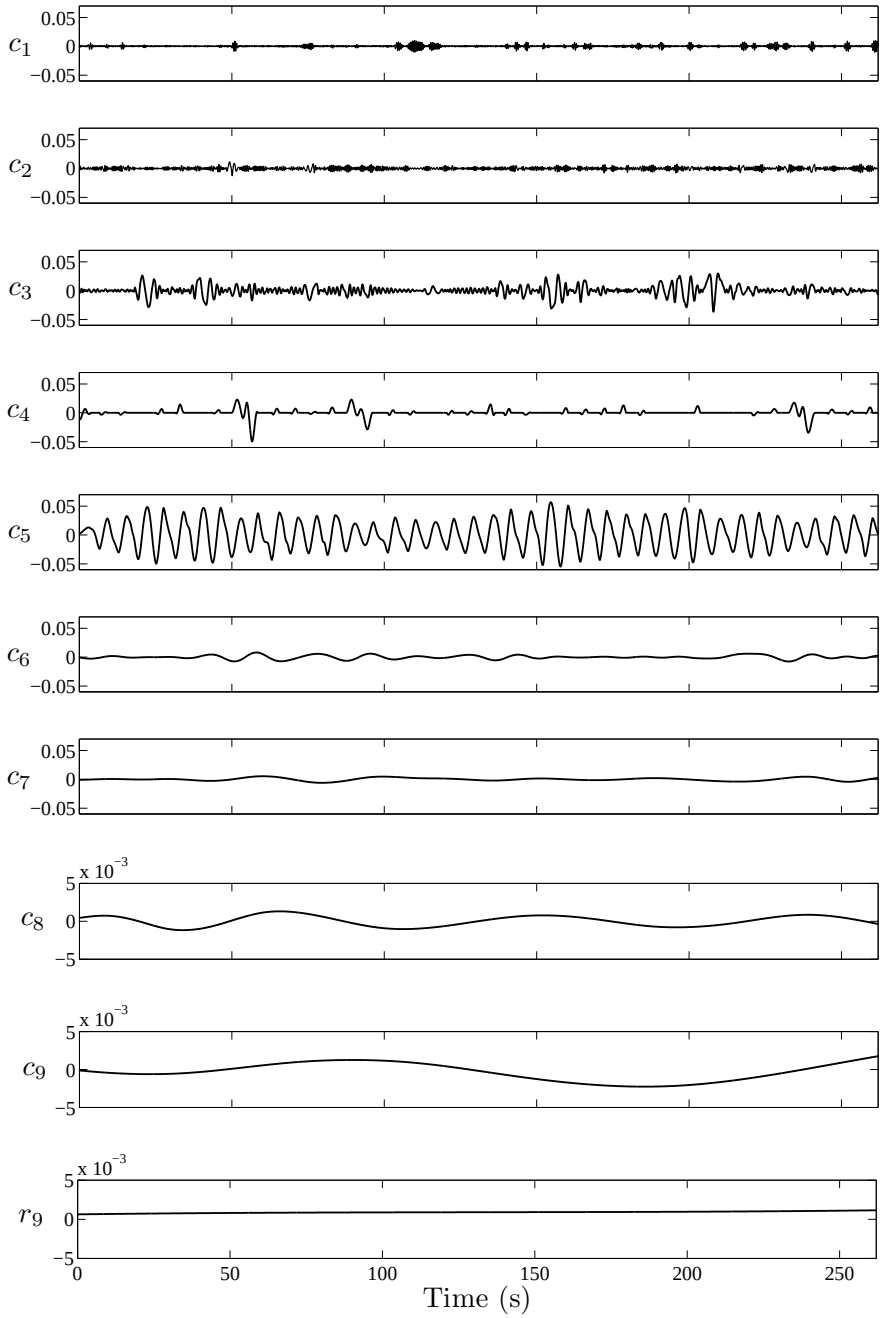

Figure 3.33: The IMF components of the vortex-shedding data using the fourth end-point option and for the case: `EMD([1:0.053:262; V-S 105], 4,[0.095,0.95,0.125],[0,0,0,2.7,-1])` with $N_{epn} = N$, $N_{avg} = 0.1N$ and $\kappa = 10^{-3}$ and *with intermittency test*. The mode mixing has been successfully prevented by the intermittency test. The intermittent signals are found in c_4 and the vortex-shedding signal is completely recovered in c_5.

are no other lower frequency peaks but only the harmonics of the main peak at 0.32 Hz, 0.48 Hz and 0.73 Hz. However, these harmonics are mathematically meaningful, but their physical interpretation has not yet been fully understood. In conclusion, the vortex-shedding frequency recovered by the HHT,

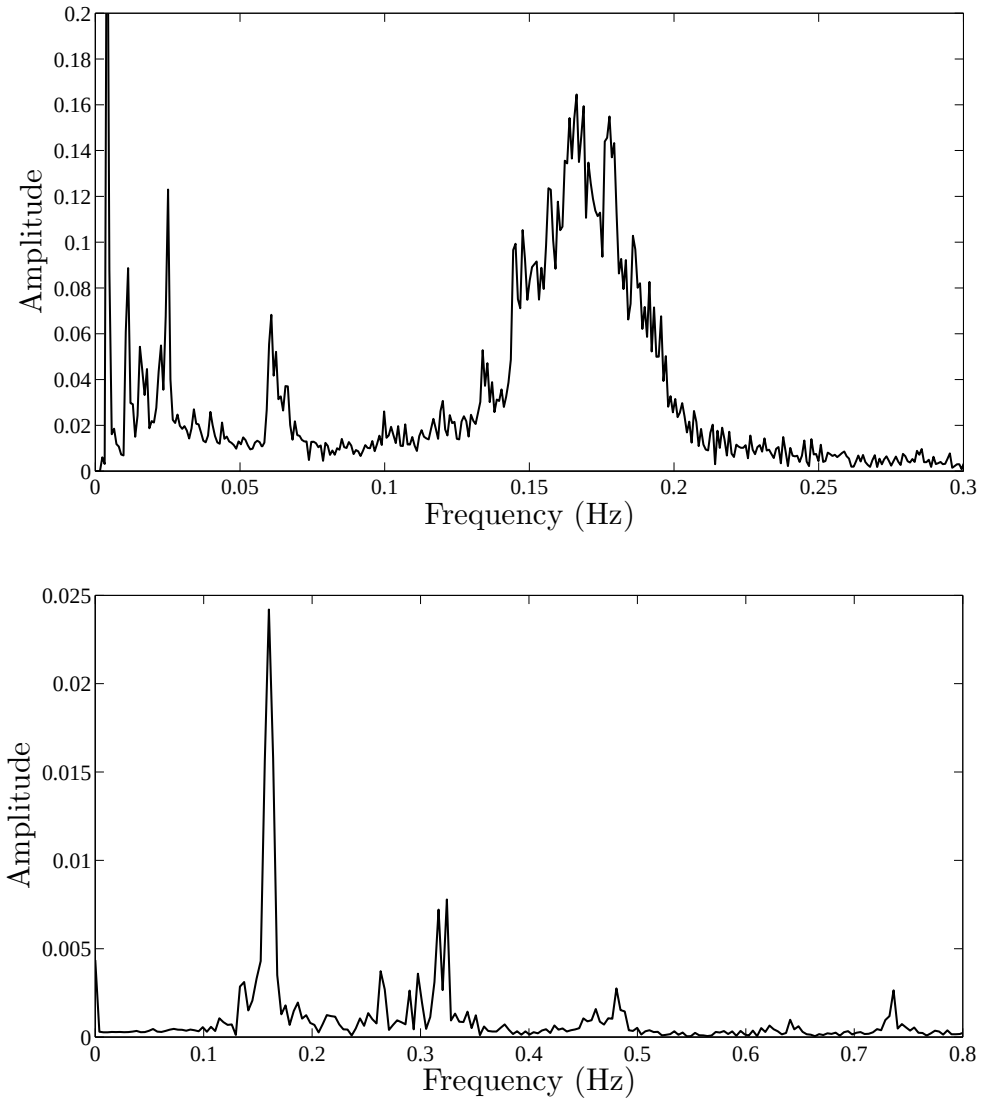

Figure 3.34: Marginal spectrum (top) and Fourier spectrum (bottom) of the vortex-shedding signal. The two spectra have their main peaks centered on the shedding frequency at approximately 0.16 Hz. However, the marginal spectrum shows a much wider peak than the Fourier spectrum.

$2F_{s,HHT} = 0.165$ Hz, and the Fourier transform, $2F_{s,FT} = 0.165$ Hz, show an excellent agreement. Nevertheless, the theoretical estimation, $2F_{s,T} = 0.14$ Hz, is slightly lower than these two results. This difference may be due to experimental approximations.

The comparison between the marginal spectrum and the Fourier spectrum actually brings much more information than solely the value of the shedding frequency. In fact, the great difference between the width of the main peaks in the two spectra is characteristic of a shortcoming of the Fourier transform: it cannot handle nonlinear phenomena. In other words, the broad peak found in the marginal spectrum reveals that the vortex-shedding frequency is not constant over time but that it varies from approximately 0.135 Hz to 0.195 Hz. Actually, the result given by the Fourier spectrum is the average of the true instantaneous vortex-shedding frequency. As an example showing the inability of the Fourier transform to correctly analyse a nonlinear signal, the Fourier spectrum and the marginal spectrum of the frequency modulated signal studied in Section 3.2.3 have been plotted on Figure F.1. Furthermore, as can be seen on the Hilbert spectrum plotted on Figure 3.35, the oscillations of the instantaneous frequency of c_5 clearly seem to be periodical. Further inspections can reveal that the frequency oscillates at approximately twice the shedding frequency. Applying both the Hilbert-Huang transform and the Fourier transform to the instantaneous frequency signal of c_5 (ω_5), the marginal spectrum and the Fourier spectrum presented on Figure 3.36 clearly reveal a main peak at $2F_s$. On the other hand, the secondary peaks that are found in the marginal spectrum below 0.08 Hz are not taken into account since they may be the result of too many manipulations. Therefore, ω_5 oscillates at a frequency of $2F_s$ about its mean $\overline{\omega_5} = 0.163$ Hz and with a standard deviation of $\sigma(\omega_5) = 0.043$ Hz.

A second vortex-shedding signal at $Re = 145$ has also been studied using the same method and invoking the intermittency test to prevent mode mixing. The results of the marginal spectrum and the Hilbert spectrum can be found in Appendix E.2. Similarly, the same phenomenon of frequency modulation of the vortex-shedding component (the third IMF) can be observed, and ω_3 is also found to oscillate at approximately $2F_s$ about its mean $\overline{\omega_3} = 0.232$ Hz and with a standard deviation of $\sigma(\omega_3) = 0.063$ Hz (at this Reynolds number, the theoretical

shedding frequency given by Equation (3.7) is $2F_{s,T} = 0.21$ Hz and the shedding frequency found by the Fourier transform is $2F_{s,FT} = 0.235$ Hz).

In conclusion, a nonlinear phenomenon of intra-wave frequency modulation varying at the frequency $2F_s$ has been identified in the vortex-shedding signal for Reynolds number belonging to the stable regime.

3.4.4 Discussion and interpretation of the phenomenon of intra-wave frequency modulation

In this study of vortex-shedding signals, we have shown how the results given by the quantitative indexes can help choose the best control parameters in the HHT algorithm in order to yield a successful decomposition. Then, the intermittency test has been invoked to prevent mode mixing in the resulting IMFs. This step was very important to isolate the vortex-shedding signal. Then, after noting a significant difference between the marginal spectrum and the Fourier spectrum, the Hilbert spectrum has revealed a phenomenon of intra-wave frequency modulation in the vortex-shedding signal. Moreover, the instantaneous frequency has been found to oscillate at a frequency equal to two times the vortex-shedding frequency, and its variations range between approximately $+25\%$ and -25% with respect to the mean frequency at both $Re = 105$ and $Re = 145$, the mean frequency being equal to $2F_s$.

The periodical variations of the vortex-shedding frequency physically correspond to the underlying mechanisms of the vortex-shedding phenomenon. Let us consider the line of vortices at a distance sufficiently far downstream from the cylinder so that the vortices are fully developed. The vortex cycle starts with the appearance of a vortex and ends with the appearance of the following. This cycle has a fixed period of $T_s = 1/F_s$. However, within a cycle the energy measured by a probe is not constant: a rapid increase of the energy is first recorded for a short time interval which corresponds to the peak of vorticity convected by the flow; second, the energy slowly dissipates for a longer time interval, during

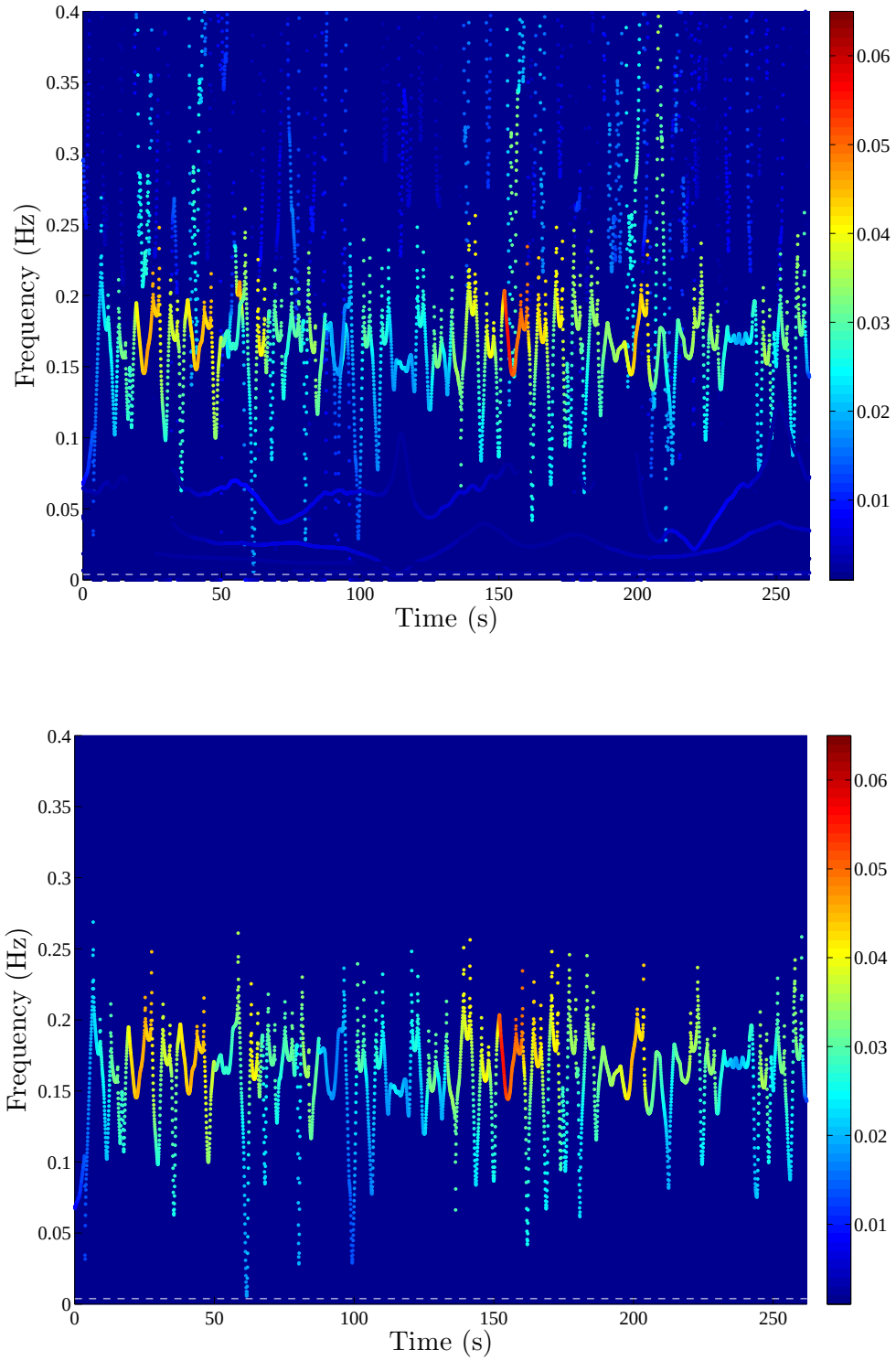

Figure 3.35: Hilbert spectrum of the vortex-shedding signal: all the IMF components (top), IMF c_5 (bottom). The evolution of the instantaneous frequency of c_5 (with the largest amplitude) seems to oscillate at a fixed period. The large swings that can be observed, especially in the high frequencies, are due to the computation of the Hilbert transform. The white dashed line at 0.004 Hz is the minimum frequency that can be recovered from this signal. However, the frequency evolution of c_5 is well above this limit, contrary to lower-frequency IMFs.

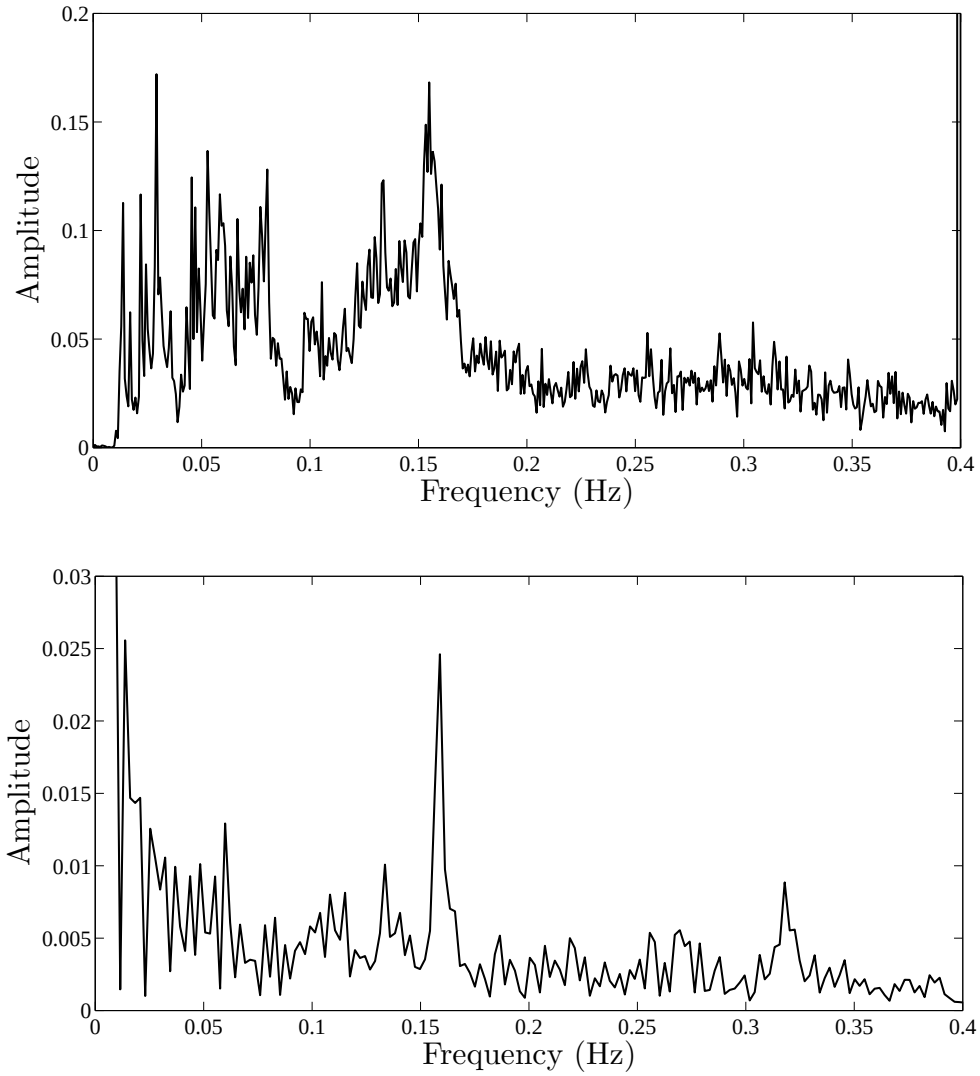

Figure 3.36: Marginal spectrum (top) and Fourier spectrum (bottom) of the instantaneous frequency of c_5. The main peaks found at 0.16 Hz in the two spectra clearly show that ω_5 oscillates at a frequency of $2F_s$.

which the flow tends to return to a quiescent state. Therefore, this is precisely these instantaneous fluctuations of the energy which are put forward by the Hilbert-Huang transform.

This phenomenon does not seem to have been reported before. Only one study from Hu et al. (2002) [22] reported having observed—in similar experiments with T-shaped cylinder, but at $Re = 1.51 \times 10^4$—that "the instantaneous vortex-shedding frequency does not appear as a constant although the variations are rather small". However, they thought that these variations were the consequences of either perturbations in the flow or numerical errors in the computation of the Hilbert transform. From the analysis of their experimental results, we believe that they actually noticed the same phenomenon of intra-wave frequency modulation as found in this study, thus strengthening our findings. Except in this study, the Hilbert-Huang transform has never been applied before to the phenomenon of vortex shedding. In fact, the Fourier transform, which cannot handle nonlinearities, has always prevailed. Therefore, we believe this is the reason why this phenomenon has never been discovered.

In conclusion, this analysis of an experimental nonlinear signal with the HHT has revealed an unexpected finding in the physics of vortex shedding. However, the phenomenon can be further explored by investigating more signals in other ranges of the Reynolds number. The magnitude of the variations of the instantaneous frequency with respect to the average shedding frequency can also be analysed to know whether or not it is invariant with the Reynolds number. Furthermore, measurements from only one vortex row should be considered in order to see whether low frequency modes are still present in the decomposition. Finally, this study is another demonstration of the power of the HHT compared to Fourier-based techniques, and we strongly recommend its use in the domain of fluid mechanics, where mostly nonlinear and transient mechanisms are encountered.

4. Conclusion

In the current work, a new and powerful signal-processing technique, the Hilbert-Huang transform has been studied. The novelty of this method is that it can yield the instantaneous frequency and the instantaneous amplitude of any non-linear or non-stationary data, contrary to other existing techniques. However, the algorithm of the empirical mode decomposition, which decomposes the signal into monocomponents or IMFs, has several flaws, namely the end-effect, the stopping criterion and the mode mixing phenomenon, which we addressed in great detail. Four end-point options, the clamped end-point option, the mirror imaging extension technique, the extrema extension method and the damped sinusoidal extrapolation based on a second-order auto-regressive model, were described and tested with a stopping criterion based on the two IMF conditions. To evaluate the performance of the HHT algorithm with different combinations of control parameters, four quantitative indexes were introduced: the index of orthogonality, the number of IMF, the number of iteration and the index of component separation. From the systematic study of five test signals and the length-of-day data, we showed that these indicators were very useful to determine the control parameters that can yield a successful decomposition, a task which is usually troublesome. In most cases, the fourth end-point option

gave the best results. Then, the thresholds of the stopping criterion should not generally be set to very low values to avoid problems of convergence or over-decomposition of the modes. On the other hand, very high thresholds can fail to fully decompose the signal. In other words, the intermediate range of thresholds: $(0.04 \leq \theta_1 \leq 0.12, \theta_2 = 10\theta_1, 0.05 \leq \alpha \leq 0.2)$, is recommended for this stopping criterion. A summary of the optimal combination of implementation options for each signal studied can be found in Table G.1. Finally, after providing the source code of each algorithm, we demonstrated their efficiency in the study of a vortex-shedding signal. A decomposition free of mode mixing was achieved with the intermittency test and the Hilbert spectrum revealed an unexpected result: a nonlinear phenomenon of quasi-periodic intra-wave frequency modulation accounting for the intrinsic physics of vortex shedding. For Reynolds number in the stable regime, the instantaneous frequency was found to vary $\pm 25\%$ with respect to the theoretical shedding frequency. This result demonstrates the power of the HHT compared to Fourier based techniques, which cannot handle nonlinear phenomena. We strongly recommend the use of this method as it yields physically meaningful results and brings a rich interpretation of the underlying mechanisms of any signal when there is sufficient physical knowledge about the data.

Bibliography

[1] Nii O. Attoh-Okine. Perspectives on the Theory and Practices of the Hilbert-Huang Transform. In Norden E. Huang and Nii O. Attoh-Okine, editors, *The Hilbert-Huang Transform in Engineering*, pages 281–305. Taylor & Francis Group, LLC, Boca Raton, The United States of America, 2005.

[2] Arthur E. Barnes. The Calculation of Instantaneous Frequency and Instantaneous Bandwidth. *Geophysics*, 57:1520–1524, 1992.

[3] Julius S. Bendat and Allan G. Piersol. *Random Data Analysis and Measurement Procedures*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., New York, The United States of America, third edition, 2004.

[4] Boualem Boashash. Estimating and Interpreting the Instantaneous Frequency of a Signal—Part 2: Algorithms and Applications. *Proceedings of the IEEE*, 80:539–568, 1992a.

[5] Boualem Boashash, editor. *Time-Frequency Signal - Analysis Methods and Applications*. Longman Cheshire, Melbourne, Australia, 1992b.

[6] Jean-Michel Bony. *Intégration et analyse hilbertienne*. Les Éditions de l'École Polytechnique, Palaiseau, France, 2003.

[7] Peter J. Brockwell and Richard A. Davis. *Introduction to Time Series and Forecasting*. Springer-Verlag, Inc., New York, The United States of America, 1996.

[8] Qiuhui Chen, Norden E. Huang, Sherman Riemenschneider, and Yuesheng Xu. A B-Spline Approach for Empirical Mode Decompositions. *Advances in Computational Mathematics*, 24:171–195, 2006.

[9] Junsheng Cheng, Dejie Yu, and Yu Yang. Application of Support Vector Regression Machines to the Processing of End Effects of Hilbert-Huang Transform. *Mechanical Systems and Signal Processing*, 21:1197–1211, 2007.

[10] Leon Cohen. *Time-Frequency Analysis*. Prentice Hall PTR, Upper Saddle River, The United States of America, 1995.

[11] Katie Coughlin and Ka K. Tung. Empirical Mode Decomposition and Climate Variability. In Norden E. Huang and Samuel S. P. Shen, editors, *Hilbert-Huang Transform and Its Applications*, pages 149–165. World Scientific Publishing Co. Pte. Ltd., Singapore, 2005.

[12] Christophe Damerval, Sylvain Meignen, and Valérie Perrier. A Fast Algorithm for Bidimensional EMD. *IEEE Signal Processing Letters*, 12:701–704, 2005.

[13] Ingrid Daubechies. *Ten Lectures on Wavelets*. CBMS-NFS Regional Conference Series in Applied Mathematics. Society For Industrial & Applied Mathematics, The United States of America, 1992.

[14] P. G. Drazin. *Nonlinear Systems*. Cambridge University Press, Cambridge, The United Kingdom, 1992.

[15] Dean G. Duffy. The Application of Hilbert-Huang Transforms to Meteorological Datasets. In Norden E. Huang and Samuel S. P. Shen, editors, *Hilbert-Huang Transform and Its Applications*, pages 129–147. World Scientific Publishing Co. Pte. Ltd., Singapore, 2005.

[16] Patrick Flandrin, Paulo Gonçalvès, and Gabriel Rilling. EMD Equivalent Filter Banks, from Interpretation to Applications. In Norden E. Huang and Samuel S. P. Shen, editors, *Hilbert-Huang Transform and Its Applications*, pages 57–74. World Scientific Publishing Co. Pte. Ltd., Singapore, 2005.

[17] Yunchao Gao, Guangtao Ge, Zhengyan Sheng, and Enfang Sang. Analysis and Solution to the Mode Mixing Phenomenon in EMD. *2008 Congress on Image and Signal Processing*, 5:223–227, 2008.

[18] Dennis Gàbor. Theory of Communication. *Proceedings of the Institution of Electrical Engineers, London*, 93:429–457, 1946.

[19] Robert M. Gray and Lee D. Davisson. *Ergodic and Information Theory*. Dowden, Hutchinson & Ross, Inc., The United States of America, 1977.

[20] Richard S. Gross. Combinations of Earth Orientation Measurements: SPACE2000, COMB2000, and POLE2000. JPL Publication 01-2. Technical report, Jet Propulsion Laboratory, Pasadena, The United States of America, 2001.

[21] Stefan L. Hahn. *Hilbert Transforms in Signal Processing*. Artech House, Boston, The United States of America, 1995.

[22] Chih-Chung Hu, Jiun-Jih Miau, and Jung-Hua Chou. Instantaneous Vortex-Shedding Behaviour in Periodically Varying Flow. *Proceedings of the Royal Society London, Serie A*, 458:911–932, 2002.

[23] Norden E. Huang. Introduction to Hilbert-Huang Transform and Some Recent Developments. In Norden E. Huang and Nii O. Attoh-Okine, editors, *The Hilbert-Huang Transform in Engineering*, pages 1–23. Taylor & Francis Group, LLC, Boca Raton, The United States of America, 2005a.

[24] Norden E. Huang. Introduction to the Hilbert-Huang Transform and Its Related Mathematical Problems. In Norden E. Huang and Samuel S. P. Shen,

editors, *Hilbert-Huang Transform and Its Applications*, pages 1–26. World Scientific Publishing Co. Pte. Ltd., Singapore, 2005b.

[25] Norden E. Huang, Kang Huang, and Wei-Ling Chiang. HHT-Based Bridge Structural Health Monitoring Method. In Norden E. Huang and Samuel S. P. Shen, editors, *Hilbert-Huang Transform and Its Applications*, pages 263–287. World Scientific Publishing Co. Pte. Ltd., Singapore, 2005.

[26] Norden E. Huang, Zheng Shen, and Steven R. Long. A New View on Nonlinear Water Waves: the Hilbert Spectrum. *Annual Review of Fluid Mechanics*, 31:417–57, 1999.

[27] Norden E. Huang, Zheng Shen, Steven R. Long, Man-Li C. Wu, Hsing H. Shih, Quanan Zheng, Nai-Chyuan Yen, Chi C. Tung, and Henry H. Liu. The Empirical Mode Decomposition and the Hilbert Spectrum for Nonlinear and Non-Stationary Time Series Analysis. *Proceedings of the Royal Society London, Serie A*, 454:903–995, 1998.

[28] Norden E. Huang, Man-Li C. Wu, Steven R. Long, Samuel S. P. Shen, Wendong Qu, Per Gloersen, and Kuang L. Fan. A Confidence Limit for the Empirical Mode Decomposition and Hilbert Spectral Analysis. *Proceedings of the Royal Society London, Serie A*, 459:2317–2345, 2003.

[29] Norden E. Huang, Man-Li C. Wu, Wendong Qu, Steven R. Long, and Samuel S. P. Shen. Applications of Hilbert-Huang Transform to Non-Stationary Financial Time Series Analysis. *Applied Stochastic Models in Business and Industry*, 19:245–268, 2003.

[30] Wei Huang, Zheng Shen, Norden E. Huang, and Yuan C. Fung. Engineering Analysis of Biological Variables: An Example of Blood Pressure Over 1 Day. *Proceedings of the National Academy of Sciences of the United States of America*, 95:4816–4821, 1998.

[31] Wei Huang, Zheng Shen, Norden E. Huang, and Yuan C. Fung. Nonlinear Indicial Response of Complex Nonstationary Oscillations as Pulmonary Hypertension Responding to Step Hypoxia. *Proceedings of the National Academy of Sciences of the United States of America*, 96:1834–1839, 1999.

[32] Paul A. Hwang, David W. Wang, and James M. Kaihatu. A Comparison of the Energy Flux Computation of Shoaling Waves Using Hilbert and Wavelet Spectral Analysis Techniques. In Norden E. Huang and Nii O. Attoh-Okine, editors, *The Hilbert-Huang Transform in Engineering*, pages 83–95. Taylor & Francis Group, LLC, Boca Raton, The United States of America, 2005.

[33] Steven R. Long. Applications of HHT in Image Analysis. In Norden E. Huang and Samuel S. P. Shen, editors, *Hilbert-Huang Transform and Its Applications*, pages 289–305. World Scientific Publishing Co. Pte. Ltd., Singapore, 2005.

[34] Wolfgang F. G. Mecklenbräuker and Franz Hlawatsch, editors. *The Wigner Distribution - Theory and Applications in Signal Processing*. Elsevier Science B.V., Amsterdam, The Netherlands, 1997.

[35] Reginald N. Meeson, Jr. HHT Sifting and Filtering. In Norden E. Huang and Samuel S. P. Shen, editors, *Hilbert-Huang Transform and Its Applications*, pages 75–105. World Scientific Publishing Co. Pte. Ltd., Singapore, 2005.

[36] Jean-Claude Nunes, Steve Guyot, and Éric Deléchelle. Texture Analysis Based on Local Analysis of the Bidimensional Empirical Mode Decomposition. *Machine Vision and Applications*, 16:177–188, 2005.

[37] Stephen C. Phillips, Robert J. Gledhill, and Jonathan W. Essex. Application of the Hilbert-Huang Transform to the Analysis of Molecular Dynamics Simulations. *Journal of Physical Chemistry A*, 107:4869–4876, 2003.

[38] Jorge E. Pinzón, Molly E. Brown, and Compton J. Tucker. EMD Correction of Orbital Drift Artifacts in Satellite Data Stream. In Norden E. Huang and Samuel S. P. Shen, editors, *Hilbert-Huang Transform and Its Applications*, pages 167–186. World Scientific Publishing Co. Pte. Ltd., Singapore, 2005.

[39] Michel Plancherel. Sur les formules de réciprocité du type de Fourier. *Journal of the London Mathematical Society, Serie 1*, 8:220–226, 1933.

[40] Lakshman Prasad and Sundararaja S. Iyengar. *Wavelet Analysis with Applications to Image Processing*. CRC Press LLC, Boca Raton, The United States of America, 1997.

[41] Sherman Riemenschneider, Bao Liu, Yuesheng Xu, and Norden E. Huang. B-Spline Based Empirical Mode Decomposition. In Norden E. Huang and Samuel S. P. Shen, editors, *Hilbert-Huang Transform and Its Applications*, pages 27–55. World Scientific Publishing Co. Pte. Ltd., Singapore, 2005.

[42] Gabriel Rilling, Patrick Flandrin, and Paulo Gonçalvès. On Empirical Mode Decomposition and Its Algorithms. In *IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing. NSIP-03*, Grado, Italie, 2003.

[43] Anatol Roshko. On the Development of Turbulent Wakes from Vortex Streets. Report 1191, National Advisory Committee for Aeronautics, Washington D. C. The United States of America, 1954.

[44] Liming W. Salvino, Darryll J. Pines, Michael Todd, and Jonathan M. Nichols. EMD and Instantaneous Phase Detection of Structural Damage. In Norden E. Huang and Samuel S. P. Shen, editors, *Hilbert-Huang Transform and Its Applications*, pages 227–262. World Scientific Publishing Co. Pte. Ltd., Singapore, 2005.

[45] Christos Saragiotis. crossings.m. World Wide Web, 2007. Matlab file.

[46] Torsten Schlurmann and Marcus Dätig. Carrier and Riding Wave Structure of Rogue Waves. In Norden E. Huang and Nii O. Attoh-Okine, editors,

The Hilbert-Huang Transform in Engineering, pages 25–57. Taylor & Francis Group, LLC, Boca Raton, The United States of America, 2005.

[47] Samuel S. P. Shen, Tingting Shu, Norden E. Huang, Zhaohua Wu, Gerald R. North, Thomas R. Karl, and David R. Easterling. HHT Analysis of the Nonlinear and Non-Stationary Annual Cycle of Daily Surface Air Temperature Data. In Norden E. Huang and Samuel S. P. Shen, editors, *Hilbert-Huang Transform and Its Applications*, pages 187–209. World Scientific Publishing Co. Pte. Ltd., Singapore, 2005.

[48] Merrill I. Skolnik. *Introduction to Radar Systems*. McGraw Hill, Boston, The United States of America, 2001.

[49] Eric W. Weisstein. *Gibbs Phenomenon*. From *Mathworld*–A Wolfram Web Resource, `http://mathworld.wolfram.com/GibbsPhenomenon.html`.

[50] Adrian P. Wiley, Robert J. Gledhill, Stephen C. Phillips, Martin T. Swain, Colin M. Edge, and Jonathan W. Essex. The Analysis of Molecular Dynamics Simulations by the Hilbert-Huang Transform. In Norden E. Huang and Nii O. Attoh-Okine, editors, *The Hilbert-Huang Transform in Engineering*, pages 245–265. Taylor & Francis Group, LLC, Boca Raton, The United States of America, 2005.

[51] Charles H. K. Williamson. Defining a Universal and Continuous Strouhal-Reynolds Number Relationship for the Laminar Vortex Shedding of a Circular Cylinder. *Physics of Fluids*, 31:2742–2744, 1988.

[52] Charles H. K. Williamson. Three-Dimensional Vortex Dynamics in Bluff Body Wakes. *Experimental Thermal and Fluid Science*, 12:150–168, 1996a.

[53] Charles H. K. Williamson. Vortex Dynamics in the Cylinder Wake. *Annual Review of Fluid Mechanics*, 28:477–539, 1996b.

[54] Xiao-Hai Yan, Young-Heon Jo, Brian Dzwonkowski, and Lide Jiang. Applications of Hilbert-Huang Transform to Ocean-Atmosphere Remote Sensing

Research. In Norden E. Huang and Nii O. Attoh-Okine, editors, *The Hilbert-Huang Transform in Engineering*, pages 59–81. Taylor & Francis Group, LLC, Boca Raton, The United States of America, 2005.

[55] Athanasios Zeris and Panayotis Prinos. Coherent Structures Analysis in Turbulent Open Channel Flow Using Hilbert-Huang and Wavelets Transforms. In Norden E. Huang and Nii O. Attoh-Okine, editors, *The Hilbert-Huang Transform in Engineering*, pages 141–157. Taylor & Francis Group, LLC, Boca Raton, The United States of America, 2005.

Appendices

A. Mathematical formulae

A.1 Definition of stationarity

Formally, a real-valued time series $X(t)$ is strictly stationary (see Brockwell and Davis 1996 [7]) if its cumulative distribution function F at time $t_1, \ldots, t_k$ satisfies

$$F[x(t_1), \ldots, x(t_k)] = F[x(t_1 + \tau), \ldots, x(t_k + \tau)], \tag{A.1}$$

for all k, for all τ and for all $t_1, \ldots, t_k$.

In signal processing, a weak form of stationarity is traditionally employed since the strict definition is usually extremely difficult to obtain. It is known as weak-sense stationarity, wide-sense stationarity, second-order stationarity, or covariance stationarity. So, a real-valued time series, X(t), is stationary in the wide sense, if, for all t and for all τ,

$$E[X(t)] = \mu, \tag{A.2}$$

$$Var[X(t)] = \sigma^2 < \infty, \tag{A.3}$$

$$Cov[X(t_1), X(t_2)] = Cov[X(t_1 + \tau), X(t_2 + \tau)] = Cov(t_1 - t_2), \tag{A.4}$$

in which $E[\cdot]$ and $Var[\cdot]$ define the ensemble average and the variance of the

quantity respectively, μ is the mean value or first moment, σ is the standard deviation and $Cov[\cdot]$ is the autocorrelation function. The weak-sense stationarity can, therefore, be interpreted as follows:

- First, Equation (A.2) means that the ensemble mean is constant over time, i.e. there is no trend.
- Second, Equation (A.3) means that the variance is a finite value and is also constant over time.
- Third, Equation (A.4) means that the autocorrelation function only depends on the difference between t_1 and t_2.

A.2 Hilbert transform and analytic signal

The Hilbert transform (see Hahn 1995 [21] or Bendat and Piersol 2000 [3]) of any real valued function $x(t)$ of L^p class [1] extending over $\mathbf{R}$ [2] is a real valued function $y(t)$ defined by

$$y(t) = \mathcal{H}[x(t)] = \frac{1}{\pi} PV \int_{-\infty}^{\infty} \frac{x(\tau)}{(t-\tau)} d\tau, \tag{A.5}$$

where PV (alternative notations to PV are P or CPV) indicates the Cauchy principal value of the singular integral. This improper integral can also be defined by the limit

$$y(t) = \lim_{\substack{\varepsilon \to 0 \\ A \to \infty}} \frac{1}{\pi} \left\{ \int_{-A}^{t-\varepsilon} \frac{x(\tau)}{(t-\tau)} d\tau + \int_{t+\varepsilon}^{A} \frac{x(\tau)}{(t-\tau)} d\tau \right\}. \tag{A.6}$$

Thus, $y(t)$ can be defined as the convolution product of $x(t)$ and $1/(\pi t)$

$$y(t) = x(t) * \left(\frac{1}{\pi t} \right). \tag{A.7}$$

[1] The L^p class denotes the space of p-power integrable functions in the sense of Lebesgue integration (see Bony (2003) [6]).

[2] $\mathbf{R}$ denotes the set of all real numbers.

Then, the analytic signal of $x(t)$ is the complex signal $z(t)$ whose imaginary part is the Hilbert transform of $x(t)$, therefore we have

$$z(t) = \mathcal{A}[z(t)] = x(t) + iy(t) = a(t)e^{i\theta(t)}, \tag{A.8}$$

where i denotes the imaginary unit, a is the amplitude function and θ is the phase function. Finally, the instantaneous frequency ω is the derivative of the phase function:

$$\omega(t) = \frac{d\theta(t)}{dt}. \tag{A.9}$$

B. HHT algorithm

B.1 EMD algorithm and sifting process

The first function of Source Code B.1 is the main loop, which calls the sifting process, computes Equation (2.4) and also tests whether the residue is the last one. Then comes the sifting process, which calls several sub-functions such as the intermittency test, the calculation of the envelopes and finally tests whether the sifted proto-IMF is an IMF by calling the fourth stopping criterion function presented in Source Code 2.1. Next follows all the sub-functions. The only function which has not been written in this source code is the function `crossings` that appears in the function `extrema`, the Matlab file 'crossings' has been implemented by Saragiotis (2007) [45].

Source code B.1: Matlab source code of the EMD algorithm

```
function [set_IMF,residu] = EMDint(t,signal,epo,thresholds,n1)

% EMD is the algorithm of the Empirical Mode Decomposition
% with possibility of invoking the intermittency test.
% It returns the set of IMFs and the last residue.

% t: time vector, to determine the time step,
%    if t = () the program will assign dt = 1.
% signal: time series (1-row vector)
% epo: end-point option: 1 -> clamped end-points
%                        2 -> extrema extension
%                        3 -> mirror imaging
%                        4 -> auto-regressive model
```

```
% thresholds = (theta1,theta2,alpha): fourth stopping criterion
% n1: intermittency criterion for each IMF (row vector of
% length(n1) = nb of IMF, and based on a time step = 1)
% or set n1 = () not to invoke the intermittency test

% Initialisation
residu = signal;
set_IMF = [];
stop_main_loop = 0;       % Stopping criterion for the last IMF
i = 1;                    % Number of IMF
if isempty(t)
    dt = 1;
else
    dt = (t(2) - t(1));
end
% Non-dimensionalisation of the intermittent criterion vector
n1 = n1/dt;

% Main Loop
while ~stop_main_loop
    % Call the sifting process
    IMF = sifting_process(residu,epo,thresholds,n1,i);
    set_IMF = cat(1,set_IMF,IMF); % Store the IMF
    residu = residu - IMF;        % New residue
    % Call the last_residue function
    stop_main_loop = last_residu(residu);
    i = i + 1;
end
end

%-----------------------------------------------------------------
function IMF = sifting_process(residu,epo,thresholds,n1,i)
stop_sifting = 0;             % Stopping criterion for the sifting
pIMF = residu;                % Initialisation of the proto-IMF

% Sifting loop
while ~stop_sifting
    % Calculate the number of zero-crossings
    nzc = length(crossings(pIMF,[],0,'dis'));
    % Call the function to index the extrema
    [maxima_idx, minima_idx] = extrema(pIMF);
    if (isempty(maxima_idx) || isempty(minima_idx))
        IMF = pIMFsave;
        return
    end
    % Option to call the intermittency test
    if ((i <= length(n1)) && (n1(i) > 0))
        IMF = intermittency_test(pIMF,epo,n1(i),maxima_idx,minima_idx);
        return
    end
    % Call the function to calculate the mean and the mode amplitude
    [me, ma] = envelope(pIMF,epo,maxima_idx,minima_idx);
    ne = (numel(maxima_idx) + numel(minima_idx)); % Number of extrema
    % Call the fourth stopping criterion fonction
    if ~(epo==1)
        stop_sifting = Stopping_criterion_4(nzc,ne,me,ma,thresholds);
    else
        stop_sifting = (abs(nzc - ne) <= 1);
    end
    if stop_sifting
        IMF = pIMF;
        return
    else
        pIMFsave = pIMF;
        pIMF = pIMF - me;  % New pIMF
    end
end
```

```
81  end
82
83  %---------------------------------------------------------------
84  function stop_main_loop = last_residu(residu)
85  [maxima_idx, minima_idx] = extrema(residu);
86  ne_residu = (numel(maxima_idx) + numel(minima_idx));
87  % Crude estimation of the amplitude
88  amp_residu = abs(max(residu) - min(residu));
89  % Test of last residue
90  stop_main_loop = ((ne_residu < 3) || (amp_residu < 1e-10));
91  end
92
93  %---------------------------------------------------------------
94  function [maxima_idx, minima_idx] = extrema(x)
95  sign_deriv = sign(diff(x)); % Sign of the derivative of the signal x
96  extr_idx = crossings(sign_deriv,[],0,'dis');    % Indexes of extrema
97  if isempty(extr_idx)
98      maxima_idx = [];
99      minima_idx = [];
100     return
101 end
102 % Find the indexes of the maxima and minima
103 if (sign_deriv(extr_idx(1)) < 0)
104     maxima_idx = extr_idx(1:2:end);
105     minima_idx = extr_idx(2:2:end);
106 elseif (sign_deriv(extr_idx(1)) == 0)
107     before_0 = extr_idx(1)-1;
108     while (sign_deriv(before_0) == 0)
109         before_0 = before_0-1;
110     end
111     if (sign_deriv(before_0) < 0)
112         minima_idx = extr_idx(1:2:end);
113         maxima_idx = extr_idx(2:2:end);
114     else
115         maxima_idx = extr_idx(1:2:end);
116         minima_idx = extr_idx(2:2:end);
117     end
118 else
119     minima_idx = extr_idx(1:2:end);
120     maxima_idx = extr_idx(2:2:end);
121 end
122 end
123
124 %---------------------------------------------------------------
125 function [mean_amp,mode_amp] = envelope(pIMF,epo,maxima_idx,minima_idx)
126 % Find the position and value of all the extrema (signal+extensions)
127 % according to the end-point option.
128 % Find also the length of the signal+extension (xx) and the
129 % indexes of the  first and last points of the signal (xstart and xend)
130 [all_max,all_min,xx,xstart,xend] = end_point(pIMF,maxima_idx,minima_idx,epo);
131 upper_envelope = interp1(all_max(1,:),all_max(2,:),xx,'spline');
132 lower_envelope = interp1(all_min(1,:),all_min(2,:),xx,'spline');
133 mean_amp = (upper_envelope(xstart:xend) + lower_envelope(xstart:xend))/2;
134 mode_amp = (upper_envelope(xstart:xend) - lower_envelope(xstart:xend))/2;
135 end
136
137 %---------------------------------------------------------------
138 function [all_max,all_min,xx,xstart,xend]=end_point(pIMF,max_idx,min_idx,epo)
139 x = [1:length(pIMF);pIMF];        % Initialisation
140 n = length(x);
141 if (epo == 1)                     % First end-point option
142     % The endpoints are both local maxima and minima
143     all_max = [x(:,1) x(:,max_idx) x(:,end)];
144     all_min = [x(:,1) x(:,min_idx) x(:,end)];
145     xx = x(1,:);
146     xstart = 1;
147     xend = n;
```

```
148 elseif (epo == 2)                % Second end-point option
149     % Method of extrapolation of the first and last two extrema
150     extr_idx = sort([max_idx min_idx]);
151     tp1 = min(1,(2*x(1,extr_idx(1))-x(1,extr_idx(2))));
152     tp2 = tp1 - (x(1,extr_idx(2))- x(1,extr_idx(1)));
153     if tp1 < x(1,1)
154         xp1 = x(2,extr_idx(2));
155     else
156         xp1 = x(2,1);
157     end
158     xp2 = x(2,extr_idx(1));
159     tpad = max(x(1,end),(2*x(1,extr_idx(end))-x(1,extr_idx(end-1))));
160     tpd = tpad + (x(1,extr_idx(end)) - x(1,extr_idx(end-1)));
161     if tpad > x(1,end)
162         xpad = x(2,extr_idx(end-1));
163     else
164         xpad = x(2,end);
165     end
166     xpd = x(2,extr_idx(end));
167     if  xp1 < xp2
168         all_max = [[tp2 ; xp2] x(:,max_idx)];
169         all_min = [[tp1 ; xp1] x(:,min_idx)];
170     else
171         all_max = [[tp1 ; xp1] x(:,max_idx)];
172         all_min = [[tp2 ; xp2] x(:,min_idx)];
173     end
174     if  xpad < xpd
175         all_max = [all_max [tpd ; xpd]];
176         all_min = [all_min [tpad ; xpad]];
177     else
178         all_max = [all_max [tpad ; xpad]];
179         all_min = [all_min [tpd ; xpd]];
180     end
181     xx = tp2:0;
182     xstart = length(xx)+1;
183     xx = [xx x(1,:)];
184     xend = length(xx);
185     xx = [xx ((x(1,end)+1):tpd)];
186 elseif (epo == 3)                % Third end-point option
187     % Mirror imaging of the signals beside the edges
188     xspan = (n - 1);
189     xx = [x(1,1:(end-1))-xspan, x(1,1:(end-1)), x(1,:)+xspan];
190     ymirror = [fliplr(x(2,2:end)), x(2,:), fliplr(x(2,1:(end-1)))];
191     [mamr, mimr] = extrema(ymirror);
192     % We must remove the fallacious local extrema at the two endpoints of the
193     % initial signal
194     if (numel(find(n==mimr)) == 1) && (numel(find(2*n-1==mimr)) == 1)
195         mimr = mimr([1:(find(n==mimr)-1),...
196             (find(n==mimr)+1):(find((2*n-1)==mimr)-1),...
197             (find((2*n-1)==mimr)+1):end]);
198     elseif (numel(find(n==mamr)) == 1) && (numel(find(2*n-1==mimr)) == 1)
199         mamr = mamr([1:(find(n==mamr)-1) (find(n==mamr)+1):end]);
200         mimr = mimr([1:(find((2*n-1)==mimr)-1) (find((2*n-1)==mimr)+1):end]);
201     elseif (numel(find(n==mimr)) == 1) && (numel(find(2*n-1==mamr)) == 1)
202         mamr = mamr([1:(find((2*n-1)==mamr)-1) (find((2*n-1)==mamr)+1):end]);
203         mimr = mimr([1:(find(n==mimr)-1) (find(n==mimr)+1):end]);
204     elseif (numel(find(n==mamr)) == 1) && (numel(find(2*n-1==mamr)) == 1)
205         mamr = mamr([1:(find(n==mamr)-1),...
206             (find(n==mamr)+1):(find((2*n-1)==mamr)-1),...
207             (find((2*n-1)==mamr)+1):end]);
208     end
209     all_max = [xx(mamr); ymirror(mamr)];
210     all_min = [xx(mimr); ymirror(mimr)];
211     xstart = n;
212     xend = (2*xstart - 1);
213 elseif (epo == 4)                % Fourth end-point option
214     % Extrapolation of the curves with a damped sinusoidal curve using an
```

```
215     % auto-regressive model.
216     % The following three parameters can be adjusted
217     lex = 1;                    % Length of extrapolation
218     lav = 1/10;                   % Length of average
219     kappa = 0.001;                 % Damping coefficient
220
221     avb = mean(x(2,(1:floor(n*lav))));
222     ave = mean(x(2,(end-floor(n*lav)):end));
223     xb = zeros(1,floor(n*(lex+1))); xe = zeros(1,floor(n*(lex+1)));
224     xb(1,(floor(n*lex)+1):end) = x(2,:)-avb; xe(1,1:n) = x(2,:)-ave;
225     % The time scale is based on the first and last two extrema
226     Tb = 2*abs(x(1,max_idx(1))-x(1,min_idx(1)));
227     Te = 2*abs(x(1,max_idx(end))-x(1,min_idx(end)));
228     % Condition of minimum period
229     if (Tb < 4)
230         Tb = 4;
231     end
232     if (Te < 4)
233         Te = 4;
234     end
235     omegab = 2*pi/(Tb);          % Pulsation of the sine wave at the beginning
236     omegae = 2*pi/(Te);          % Pulsation of the sine wave at the end
237
238     b1b = (2 - omegab^2)/(1 + kappa/2);
239     b1e = (2 - omegae^2)/(1 + kappa/2);
240     b2 = -(1 - kappa/2)/(1 + kappa/2);
241
242     % Iteration process to calculate the damped sinusoidal extensions
243     for ii=1:(floor(n*lex))
244         pointb = b1b*xb(floor(n*lex)-ii+2)+b2*xb(floor(n*lex)-ii+3);
245         pointe = b1e*xe(n+ii-1)+b2*xe(n+ii-2);
246         xb(floor(n*lex)+1-ii) = pointb;
247         xe(n+ii) = pointe;
248     end
249     xext = [(xb(1:floor(n*lex))+avb), x(2,:), (xe(n+1:end)+ave)];
250     xx = [fliplr(0:-1:(1-n*lex)), x(1,:), (n+1):(n+n*lex)];
251     [maex, miex] = extrema(xext);
252     all_max = [xx(maex); xext(maex)];
253     all_min = [xx(miex); xext(miex)];
254     xstart = (1 + floor(n*lex));
255     xend = floor(n*(lex+1));
256 end
257 end
```

B.2 Hilbert-transform algorithm

Source code B.2 determines the instantaneous frequency and the instantaneous amplitude of a signal or IMFs using the Hilbert transform. Besides the time-series and its time vector (essentially to define the time-step), the inputs are the extension option (1, no extension; 2, extension with an anti-symmetric mirror imaging; 3, extension with a damped sinusoidal curve using an auto-regressive model) and the length of extension, which is a proportion of the size of the data.

Source code B.2: Matlab source code of the Hilbert-transform algorithm

```
function [Amplitude, Frequency] = Hilbert_transform(t,signal,e_o,l_e)
% Hilbert transform of a signal or set of IMF (set_IMF)
% signal is a multiple-row vector
% t: time vector
% Four extension-options possible:
% e_o = 1 -> No extension
%       2 -> Anti-symmetric mirror imaging
%       3 -> Extension with a damped sinusoidal curve (AR model)
% l_e: length of extension (proportion of the signal size)
%      with 0 < l_e < n

if isempty(t), dt = 1;
else dt = t(2)-t(1);          % Initialisation of the time step
end
m = size(signal,1);

for k = 1:m
    % Call the extension function
    [x, xstart, xend] = extension(signal(k,:),e_o,l_e);
    if (k == 1)
        Amplitude = zeros(m,length(x));
        Frequency = zeros(m,length(x));
    end
    % Computation of the analytic signal with the Hilbert transform
    Ana_sig = hilbert(x);
    Amplitude(k,:) = abs(Ana_sig);
    Frequency0 = 1/(4*pi*dt)*...
        unwrap(atan2((real(Ana_sig(1:end-2)).*...
        imag(Ana_sig(3:end))-real(Ana_sig(3:end)).*...
        imag(Ana_sig(1:end-2))),(real(Ana_sig(1:end-2)).*...
        real(Ana_sig(3:end))+imag(Ana_sig(3:end)).*...
        imag(Ana_sig(1:end-2)))));
    Frequency(k,:) = [Frequency0(1) Frequency0 Frequency(end)];
end
Amplitude = Amplitude(:,xstart:xend);
Frequency = Frequency(:,xstart:xend);
end

%-------------------------------------------------------------------
function [signal_ext, xstart, xend] = extension(x, e_o, l_e)
n = length(x);

if (e_o == 1)                 % 1st extension option
    % No extension
    signal_ext = x;
    xstart = 1;
    xend = n;
elseif (e_o == 2)             % 2nd extension option
    % Anti-symmetric mirror imaging
    if (l_e == 1)
        l_e = (1 - 1/n);
    end
    if (l_e == 0)
        l_e = 1/n;
    end
    signal_ext = [(2*x(1)-fliplr(x(2:floor(n*l_e)))), x,...
        (2*x(1)-fliplr(x((n-floor(n*l_e)):(end-1))))];
    xstart = floor(n*l_e);
    xend = (floor(n*l_e) + n - 1);
elseif (e_o == 3)             % 3rd end-point option
    % Extrapolation of the curves with a damped sinusoidal
    % curve using an auto-regressive model.
    % The following parameters can be adjusted
    lav = 1/10;                % Length of average
```

```
65      kappa = 0.001;                  % Damping coefficient
66
67      extr_idx = crossings(sign(diff(x)),[],0,'dis'); % Index the extrema
68      if (length(extr_idx) < 2)
69          warning('Need at least two extrema for the AR model extension');
70          % Extension with a constant
71          signal_ext = [x(1)*ones(1,floor(n*l_e)), x,...
72              x(end)*ones(1,floor(n*l_e))];
73          xstart = 1 + floor(n*l_e);
74          xend = floor(n*(l_e+1));
75          return
76      end
77      avb = mean(x(1:floor(n*lav)));
78      ave = mean(x((end-floor(n*lav)):end));
79      xb = zeros(1,floor(n*(l_e+1))); xe = zeros(1,floor(n*(l_e+1)));
80      xb((floor(n*l_e)+1):end) = x - avb; xe(1:n) = x - ave;
81      % The time scale is based on the first and last two extrema
82      Tb = 2*abs(extr_idx(2) - extr_idx(1));
83      Te = 2*abs(extr_idx(end) - extr_idx(end-1));
84      if (Tb < 4)
85          Tb = 4;
86      end
87      if (Te < 4)
88          Te = 4;
89      end
90      omegab = 2*pi/Tb;       % Pulsation of the sine wave at the beginning
91      omegae = 2*pi/Te;       % Pulsation of the sine wave at the end
92
93      b1b = (2 - omegab^2)/(1 + kappa/2);
94      b1e = (2 - omegae^2)/(1 + kappa/2);
95      b2 = -(1 - kappa/2)/(1 + kappa/2);
96
97      % Iteration process to calculate the damped sinusoidal extensions
98      for ii=1:(floor(n*l_e))
99          pointb = b1b*xb(floor(n*l_e)-ii+2)+b2*xb(floor(n*l_e)-ii+3);
100         pointe = b1e*xe(n+ii-1)+b2*xe(n+ii-2);
101         xb(floor(n*l_e)+1-ii) = pointb;
102         xe(n+ii) = pointe;
103     end
104     signal_ext = [(xb(1:floor(n*l_e))+avb), x, (xe(n+1:end)+ave)];
105     xstart = 1 + floor(n*l_e);
106     xend = floor(n*(l_e+1));
107 end
108 end
```

B.3 Intermittency test

Source code B.3 is the intermittency test that can be called in the sifting process (see the second function in Source Code B.1). Its inputs are: a proto-IMF, which is actually a residue or the signal itself because it is the first iteration of the sifting process; the end-point option; the intermittency criterion for the current residue; and the indexes of the extrema of this residue. Its unique output is the resulting intermittent IMF.

Source code B.3: Matlab source code of the intermittency test

```
function IMF = intermittency_test(pIMF,epo,n1,max_idx,min_idx)

% Call the end_point function to get the indexes of all the extrema
[lmax,lmin,xx,xstart,xend] = end_point(pIMF,max_idx,min_idx,epo);

t_extrema = sort([lmax(1,:), lmin(1,:)]);
% Identification of the portions of pIMF
% in which waves have a half-period > n1
portion_sup_n1 = (diff(t_extrema)> n1);
portion_idx = find(portion_sup_n1==1);

% Identification of the extremas for the upper and lower envelopes
if isempty(portion_idx)
    lmaxint = lmax;
    lminint = lmin;
else
    if (lmax(1) < lmin(1))
        double_max_idx = ceil((portion_idx+1)/2);
        double_min_idx = floor((portion_idx+1)/2);
    else
        double_min_idx = ceil((portion_idx+1)/2);
        double_max_idx = floor((portion_idx+1)/2);
    end
    while (double_min_idx(end) > length(lmin))
        double_min_idx(end) = [];
    end
    lmaxint = [lmax, lmin(:,double_min_idx)];
    [t_lmaxint, lmaxint_sort_idx] = sort(lmaxint(1,:));
    lmaxint = lmaxint(:,lmaxint_sort_idx);

    while (double_max_idx(end) > length(lmax))
        double_max_idx(end) = [];
    end
    lminint = [lmin, lmax(:,double_max_idx)];
    [t_lminint, lminint_sort_idx] = sort(lminint(1,:));
    lminint = lminint(:,lminint_sort_idx);

    [b1, m1] = unique(lmaxint(1,:), 'first');
    [b2, m2] = unique(lminint(1,:), 'first');
    lmaxint = lmaxint(:,m1);
    lminint = lminint(:,m2);
end

% Calculation of the mean of the pIMF
lower_envelope = interp1(lminint(1,:),lminint(2,:),xx,'spline');
upper_envelope = interp1(lmaxint(1,:),lmaxint(2,:),xx,'spline');
me = (upper_envelope + lower_envelope)/2;
xnew = zeros(1,length(xx));
xnew(xstart:xend)=pIMF;

% The mean is forced to be equal to the pIMF and the long-wave portions
for i = 1:length(portion_sup_n1)
    if (portion_sup_n1(i) == 1)
        xx1 = find(xx == t_extrema(i));
        xx2 = find(xx == t_extrema(i+1));
        me(xx1:xx2) = xnew(xx1:xx2);
    end
end
me = me(xstart:xend);

% Calculation of the IMF
IMF = pIMF - me;
end
```

B.4 Confidence-limit algorithm

The overall architecture of the confidence-limit algorithm is described below in the pseudo code B.4. Besides the signal, the inputs are the control parameters for each case. The outputs are: the time evolution of the mean IMFs and their standard deviation; the marginal spectra of all the cases with the mean marginal spectrum and the 68% or 95% confidence limit marginal spectra; and the mean Hilbert spectrum.

Pseudo code B.4: Architecture of the confidence limit algorithm

```
function [Mean_IMFs,Std_IMFs] = Confidence_limit(t,signal,parameters)

% set_IMFs is a 3D array, with IMFs as row vectors
% and different sets varying with the last dimension

N_set = size(parameters);        % Number of sets

for i=1:N_set

 % Call the EMD algorithm
 [set_IMFs(:,:,i),IO(i)] = EMDint(t,signal,parameters(i));

 % Preliminray tests
 if (IO(i) > 0.1)
  set_IMFs(:,:,i) = [];
 end
end

N_set = size(set_IMFs); % Reinitialisation of the number of sets

% Equal number of IMFs in each set?
eq_nb_IMFs =1;
for i=1:N_set

 for j=1:N_set
  if size(set_IMFs(:,:,i),2) ~= size(set_IMFs(:,:,j),2)
   eq_nb_IMFs =0;
  end
 end

 % Instantaneous frequency and amplitude
 [IF_IMF(:,:,i),IA_IMF(:,:,i)] = Hilbert_transform(set_IMFs(:,:,i));
 % Marginal spectrum
 [M_F(i,:),M_A(i,:)] = Marg_spectrum(IF_IMF(:,:,i),IA_IMF(:,:,i));
end

% Mean and 95% CL marginal spectrum
Mean_M_A = 1/N_set*sum(M_F,1);
Std_M_A = std(M_A,1);
% Plot Marginal spectrum with its mean and 95% CL
plot(M_F,M_A)
hold on
plot(M_F,Mean_M_A)
plot(M_F,(Mean_M_A + 2*Std_M_A))
plot(M_F,(Mean_M_A - 2*Std_M_A))
hold off
```

```
47
48 % Calculation of the mean_IMFs, Std_IMFs, and Hilbert spectrum
49 if eq_nb_IMFs            % 1: equal numbers of IMFs
50
51  Mean_IMFs = 1/N_set*sum(set_IMFs,3);
52  Std_IMFs = std(set_IMFs,3);
53
54  % Mean Hilbert spectrum
55  Mean_IF_IMF = 1/N_set*sum(IF_IMF,3);
56  Mean_IA_IMF = 1/N_set*sum(IA_IMF,3);
57  % Plot mean Hilbert spectrum
58
59  plot3(t,Mean_IF_IMF,Mean_IA_IMF)
60
61 else     % 2: different numbers of IMFs
62
63  % Bin method to calculate the mean Hilbert spectrum
64  [Mean_IF_IMF, Mean_IA_IMF] = Bin_method(IF_IMF, IA_IMF);
65
66  % Plot mean Hilbert spectrum
67  plot3(t,Mean_IF_IMF,Mean_IA_IMF)
68 end
69 end
```

C. Results for the five test signals

C.1 Two-component signal

Table C.1 shows the results of the quantitative criteria for the study of the two-component signal with the HHT algorithm. In the second row, second column: 'n. a.' means 'not applicable'. In other words, the fourth stopping criterion cannot be used in its current form, and the second and third conditions are dropped. Furthermore, '$-$' means that the index cannot be computed because there is only one IMF found by the EMD algorithm.

Table C.1: Results of the index of overall orthogonality IO, the number of IMFs N_{IMF}, the number of iterations per IMF $N_{ite,j}$, and the index of component separation per IMF $\overline{ICS_j}$ for the two-component signal.

End-point option	4^{th} Stopping criterion $(\theta_1, \theta_2, \alpha)$	IO	N_{IMF}	$N_{ite,j}$ $(c_1, c_2, \ldots, c_n)$	$\overline{ICS_j}$ $[c_j - c_{j+1}]$ $[c_{n-1} - c_n]$
1	n. a.	–	1	(0)	$(-, -)$
2	$(0.01, 0.1, 0.01)$	0.014	3	$(134, 3, 15)$	$(2.85, 2.10)$ $(1.56, 1.24)$
	$(0.05, 0.5, 0.05)$	0.007	2	$(5, 0)$	$(-0.48, -0.16)$
	$(0.1, 1, 0.1)$	0.016	3	$(2, 1, 4)$	$(0.64, 0.97)$ $(0.24, 0.55)$
3	$(0.01, 0.1, 0.01)$	0.012	3	$(16, 3, 2)$	$(2.12, 1.75)$ $(0.83, 1.62)$
	$(0.05, 0.5, 0.05)$	0.021	3	$(6, 1, 4)$	$(0.56, 1)$ $(-0.25, 0.9)$
	$(0.1, 1, 0.1)$	0.062	2	$(2, 0)$	$(-1.72, -0.82)$
4	$(0.01, 0.1, 0.01)$	0.007	3	$(9, 9, 33)$	$(2.43, 2.14)$ $(1.46, 0.79)$
	$(0.05, 0.5, 0.05)$	0.029	3	$(3, 1, 19)$	$(1.50, 1.90)$ $(1.42, 0.61)$
	$(0.1, 1, 0.1)$	0.039	2	$(2, 0)$	$(-0.77, -0.38)$

C.2 Amplitude-modulated signal

Table C.2 shows the results of the quantitative criteria for the study of the amplitude-modulated signal with the HHT algorithm.

Table C.2: Results of the index of overall orthogonality IO, the number of IMFs N_{IMF}, the number of iterations per IMF $N_{ite,j}$, and the index of component separation per IMF $\overline{ICS_j}$ for the amplitude-modulated signal.

End-point option	4^{th} Stopping criterion $(\theta_1, \theta_2, \alpha)$	IO	N_{IMF}	$N_{ite,j}$ $(c_1, c_2, \ldots, c_n)$	$\overline{ICS_j}$ $[c_j - c_{j+1}]$ $[c_{n-1} - c_n]$
1	n. a.	$-$	1	(0)	$(-, -)$
2	$(0.01, 0.1, 0.01)$	0.105	3	$(6, 18, 15)$	$(1.92, 2.21)$
					$(1.26, -0.08)$
	$(0.05, 0.5, 0.05)$	0.074	3	$(3, 1, 14)$	$(-0.51, -0.29)$
					$(-1.15, 0.12)$
	$(0.1, 1, 0.1)$	0.046	2	$(2, 0)$	$(0.85, 0.74)$
3	$(0.01, 0.1, 0.01)$	0.107	3	$(17, 4, 6)$	$(1.40, 1.26)$
					$(0.20, -0.22)$
	$(0.05, 0.5, 0.05)$	0.098	3	$(5, 1, 6)$	$(1.43, 1.88)$
					$(1.26, 1.51)$
	$(0.1, 1, 0.1)$	0.115	3	$(2, 1, 4)$	$(-1.20, -0.04)$
					$(-0.68, 0.27)$
4	$(0.01, 0.1, 0.01)$	4.686	7	$(71, 66, 56, 175, \ldots$	$(1.00, -0.19)$
				$\ldots 90, 27, 10)$	$(-1.28, -0.25)$
					$(-0.25, -0.70)$
					$(-1.69, 0.35)$
					$(0.42, -0.91)$
					$(-2.09, -1.64)$
	$(0.05, 0.5, 0.05)$	0.053	3	$(3, 1, 9)$	$(1.38, 2.27)$
					$(1.46, 0.81)$
	$(0.1, 1, 0.1)$	0.027	2	$(2, 0)$	$(0.84, 1.36)$

C.3 Frequency-modulated signal

Table C.3 shows the results of the quantitative criteria for the study of the frequency-modulated signal with the HHT algorithm.

Table C.3: Results of the index of overall orthogonality IO, the number of IMFs N_{IMF}, the number of iterations per IMF $N_{ite,j}$, the and index of component separation per IMF $\overline{ICS_j}$ for the frequency-modulated signal.

End-point option	4^{th} Stopping criterion $(\theta_1, \theta_2, \alpha)$	IO	N_{IMF}	$N_{ite,j}$ $(c_1, c_2, \ldots, c_n)$	$\overline{ICS_j}$ $[c_j - c_{j+1}]$ $[c_{n-1} - c_n]$
1	n. a.	–	1	(0)	$(-, -)$
2	$(0.01, 0.1, 0.01)$	–	1	(0)	$(-, -)$
	$(0.05, 0.5, 0.05)$	–	1	(0)	$(-, -)$
	$(0.1, 1, 0.1)$	–	1	(0)	$(-, -)$
3	$(0.01, 0.1, 0.01)$	–	1	(0)	$(-, -)$
	$(0.05, 0.5, 0.05)$	–	1	(0)	$(-, -)$
	$(0.1, 1, 0.1)$	–	1	(0)	$(-, -)$
4	$(0.01, 0.1, 0.01)$	0.524	5	$(31, 126, 46, \ldots$ $\ldots 36, 12)$	$(0.84, 1.34)$ $(-0.84, -0.72)$ $(-0.28, 0.90)$ $(1.09, 0.77)$
	$(0.05, 0.5, 0.05)$	0.018	2	$(1, 18)$	$(1.64, 3.67)$
	$(0.1, 1, 0.1)$	–	1	(0)	$(-, -)$

C.4 Amplitude-step signal

Table C.4 shows the results of the quantitative criteria for the study of the amplitude-step signal with the HHT algorithm.

Table C.4: Results of the index of overall orthogonality IO, the number of IMFs N_{IMF}, the number of iterations per IMF $N_{ite,j}$, and the index of component separation per IMF $\overline{ICS_j}$ for the amplitude-step signal.

End-point option	4^{th} Stopping criterion $(\theta_1, \theta_2, \alpha)$	IO	N_{IMF}	$N_{ite,j}$ $(c_1, c_2, \ldots, c_n)$	$\overline{ICS_j}$ $[c_j - c_{j+1}]$ $[c_{n-1} - c_n]$
1	n. a.	$-$	1	0	$(-, -)$
2	$(0.01, 0.1, 0.01)$	0.003	3	$(5, 95, 6)$	$(1.56, 1.22)$ $(-0.17, -0.01)$
	$(0.05, 0.5, 0.05)$	0.010	3	$(1, 5, 1)$	$(2.84, 2.62)$ $(-0.31, -0.40)$
	$(0.1, 1, 0.1)$	0.009	3	$(1, 4, 1)$	$(3.34, 2.39)$ $(-0.71, -0.85)$
3	$(0.01, 0.1, 0.01)$	0.007	4	$(5, 138, 27, 3)$	$(2.07, 1.01)$ $(-0.53, 0.14)$ $(0.69, 0.54)$
	$(0.05, 0.5, 0.05)$	0.009	3	$(1, 4, 1)$	$(2.59, 2.54)$ $(-0.50, -0.16)$
	$(0.1, 1, 0.1)$	0.008	3	$(1, 3, 1)$	$(1.83, 2.53)$ $(-0.47, 2.42)$
4	$(0.01, 0.1, 0.01)$	0.011	4	$(5, 26, 18, 19)$	$(0.86, 0.87)$ $(-0.60, -0.47)$ $(-0.48, -0.47)$
	$(0.05, 0.5, 0.05)$	0.005	3	$(1, 15, 4)$	$(1.49, 1.09)$ $(-0.33, -0.78)$
	$(0.1, 1, 0.1)$	0.004	3	$(1, 9, 3)$	$(2.48, 0.87)$ $(-0.31, 0.15)$

C.5 Frequency-shift signal

Table C.5 shows the results of the quantitative criteria for the study of the frequency-shift signal with the HHT algorithm.

Table C.5: Results of the index of overall orthogonality IO, the number of IMFs N_{IMF}, the number of iterations per IMF $N_{ite,j}$, and the index of component separation per IMF $\overline{ICS_j}$ for the frequency-shift signal.

End-point option	4^{th} Stopping criterion $(\theta_1, \theta_2, \alpha)$	IO	N_{IMF}	$N_{ite,j}$ $(c_1, c_2, \ldots, c_n)$	$\overline{ICS_j}$ $[c_j - c_{j+1}]$ $[c_{n-1} - c_n]$
1	n. a.	–	1	(0)	$(-, -)$
2	$(0.01, 0.1, 0.01)$	–	1	(0)	$(-, -)$
	$(0.05, 0.5, 0.05)$	–	1	(0)	$(-, -)$
	$(0.1, 1, 0.1)$	–	1	(0)	$(-, -)$
3	$(0.01, 0.1, 0.01)$	–	1	(0)	$(-, -)$
	$(0.05, 0.5, 0.05)$	–	1	(0)	$(-, -)$
	$(0.1, 1, 0.1)$	–	1	(0)	$(-, -)$
4	$(0.01, 0.1, 0.01)$	0.015	3	$(3, 26, 33)$	$(1.46, 1.83)$ $(-2.16, -1.95)$
	$(0.05, 0.5, 0.05)$	–	1	(0)	$(-, -)$
	$(0.1, 1, 0.1)$	–	1	(0)	$(-, -)$

D. Length-of-day results

D.1 IMF components

The decomposition of the LOD data with the EMD algorithm is presented on Figure D.1 and D.2 without intermittency test and Figure D.3 and D.4 with intermittency test. These results have been compared to the results obtained by Huang et al. (2003) [28]. Overall, the same IMFs are found between the two studies although different stopping criteria and end-point options have been used. Only a few discrepancies have been observed near the edges of some IMFs. This is clearly due to the use of different ways to handle the end-effect problem. Then, in both cases the intermittency test improves the decomposition by removing the mode mixing. Finally, it can be remarked that a strict stopping criterion corresponds to low thresholds in this study (e.g. $\theta_1 < 0.05, \theta_2 = 10\theta_1, \alpha < 0.05$), and to a high S-number in Huang et al. (2003) [28]'s study (e.g. $S \geq 10$). In both studies, a strict stopping criterion tends to increase the number of IMFs and over-decompose the signal.

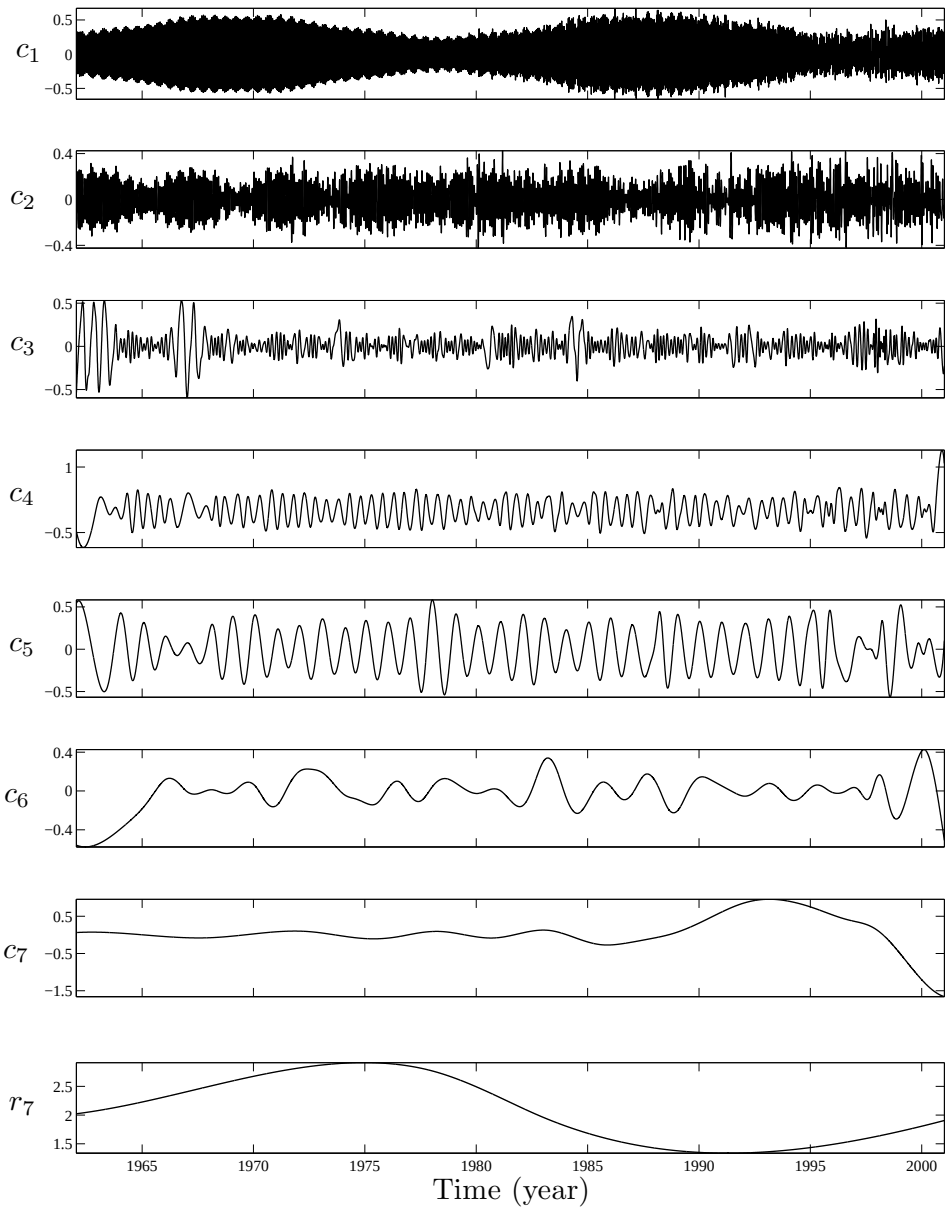

Figure D.1: The IMF components (multiplied by a factor 1000) of the case: `EMD([1962:2001; LOD data],2,[0.2,2,0.15])` (*no intermittency test*). Only seven IMFs have been found because the stopping criterion is not strict enough. In fact, the eighth IMF is mixed with the seventh.

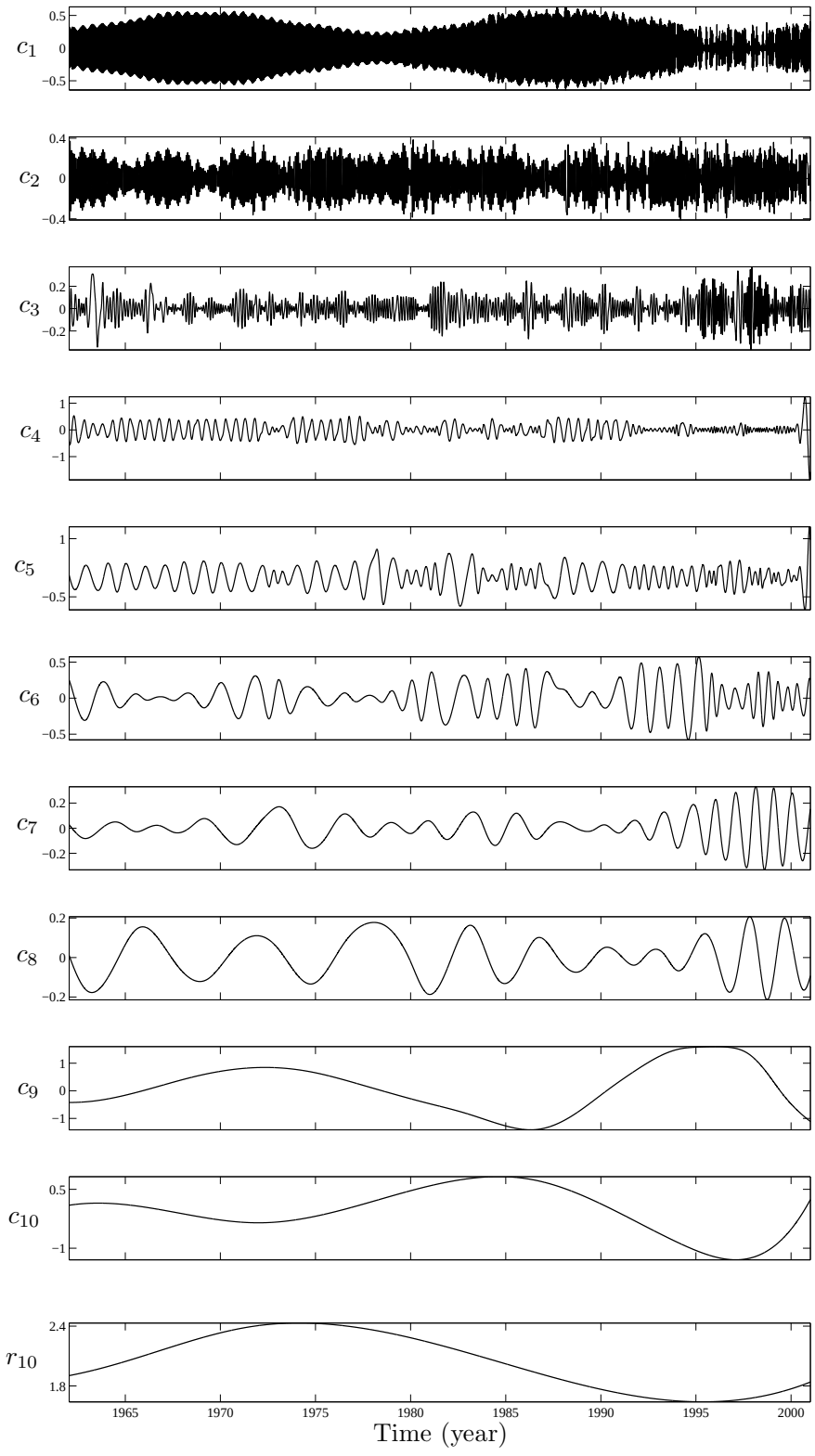

Figure D.2: The IMF components (multiplied by a factor 1000) of the case: `EMD([1962:2001; LOD data],4,[0.02,0.2,0.035])` with $N_{epn} = N$, $N_{avg} = 0.2N$ and $\kappa = 10^{-3}$ (*no intermittency test*). Too many IMFs have been created because of over-sifting due to too low thresholds. It can be noticed that c_9 and c_{10} are almost symmetrical and have the same frequency range, thus meaning that the last one is a fallacious IMF.

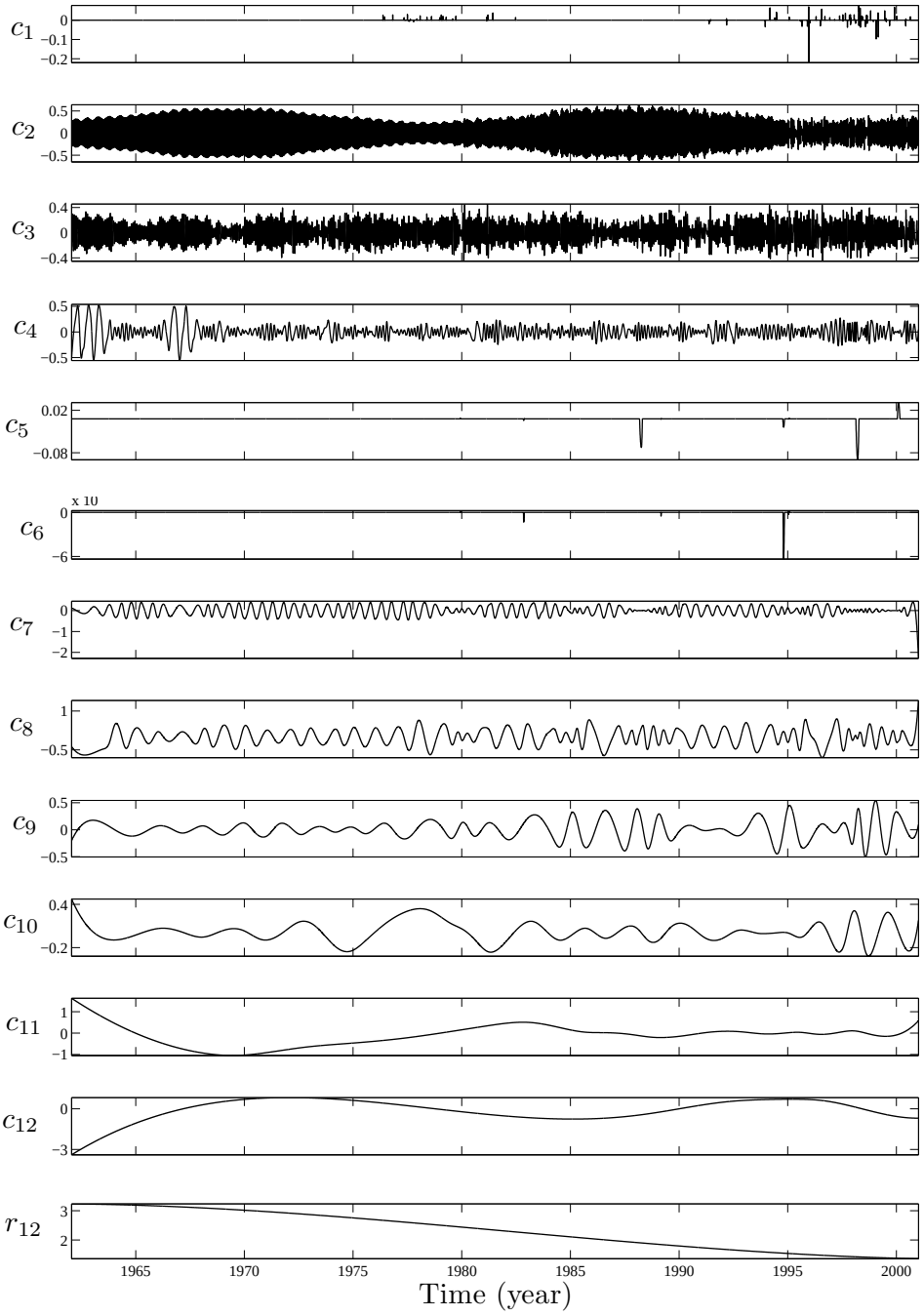

Figure D.3: The IMF components (multiplied by a factor 1000) of the case: `EMD([1962:2001; LOD data],2,[0.03,0.3,0.225],[4,0^3,45^2,-1])` (*invoking the intermittency test*). The intermittent IMFs are c_1, c_5 and c_6. Firstly, we can observe that c_7 and c_8 have a problem at one end: large steep swings terminate the two curves. Secondly, the begining of c_{11} and c_{12} are symmetrical and compensate each other. This problem is due to low first and second thresholds.

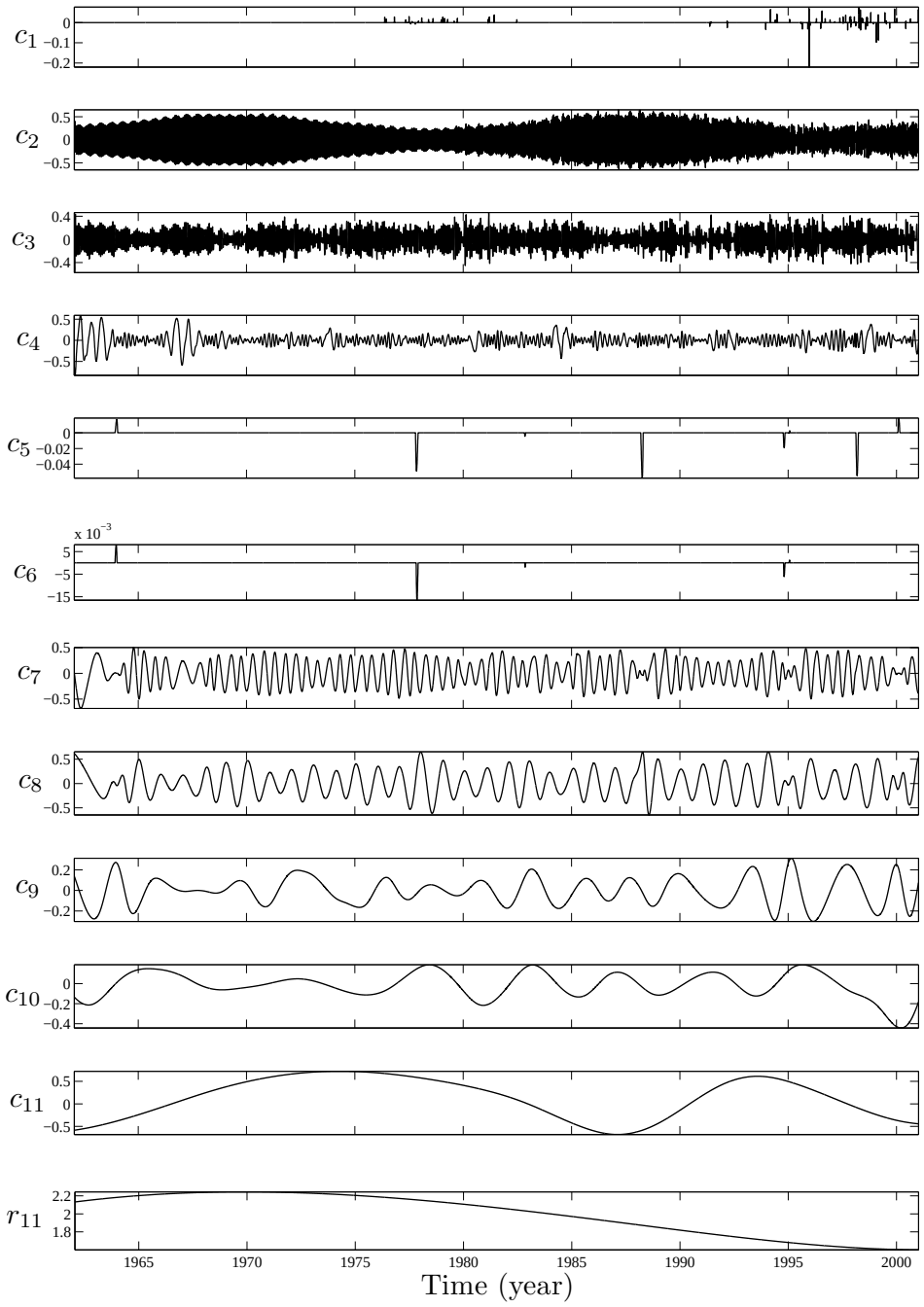

Figure D.4: The eleven IMF components (multiplied by a factor 1000) of the case: `EMD([1962:2001; LOD data],4,[0.12,1.2,0.1],[4,0^3,45^2,-1])` with $N_{epn} = N$, $N_{avg} = 0.2N$ and $\kappa = 10^{-3}$ (*with intermittency test*). The intermittent IMFs are c_1, c_5 and c_6. These parameters have produced a good decomposition without mode mixing and end-effect, and the IMFs were obtained in only 63 iterations.

D.2 Marginal spectrum

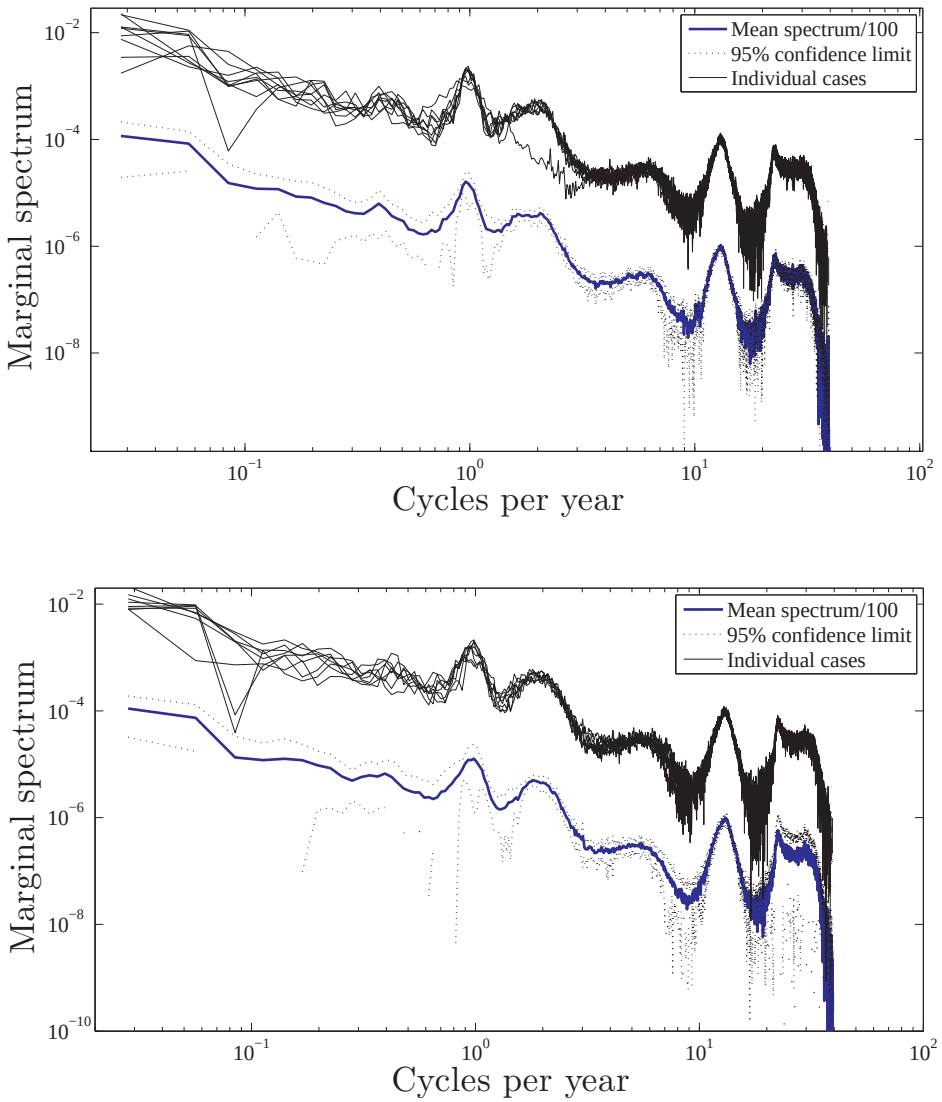

Figure D.5: Marginal spectra of a few individual cases selected randomly, mean marginal spectrum (divided by a factor 100) and 95% CL of the LOD data *without intermittency test* (top) and *with intermittency test* (bottom). As can be seen, the mean marginal spectrum, the 95% CL and the individual cases are all very similar between these graphs.

E. Vortex-shedding results

E.1 Vortex-shedding signal at Re=105

Figure E.1 to E.7 show the results of the quantitative indexes, IO, N_{IMF}, $N_{ite,T}$ and $mean(\overline{ICS})$, in the space $(0.02 \leq \theta_1 \leq 0.3, 0.02 \leq \alpha \leq 0.3)$ and with the second, third and fourth end-point options.

Figure E.8 shows the squared deviation between the marginal spectrum of each individual cases and the mean marginal spectrum.

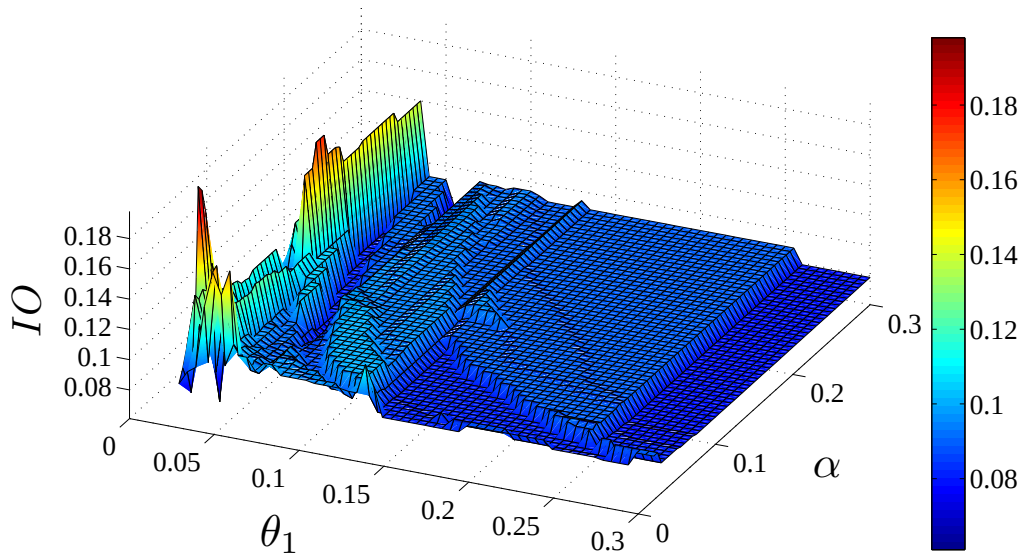

Figure E.1: Index of orthogonality versus (θ_1, α) for the study of the vortex shedding data with the *second end-point option* and *without intermittency test*.

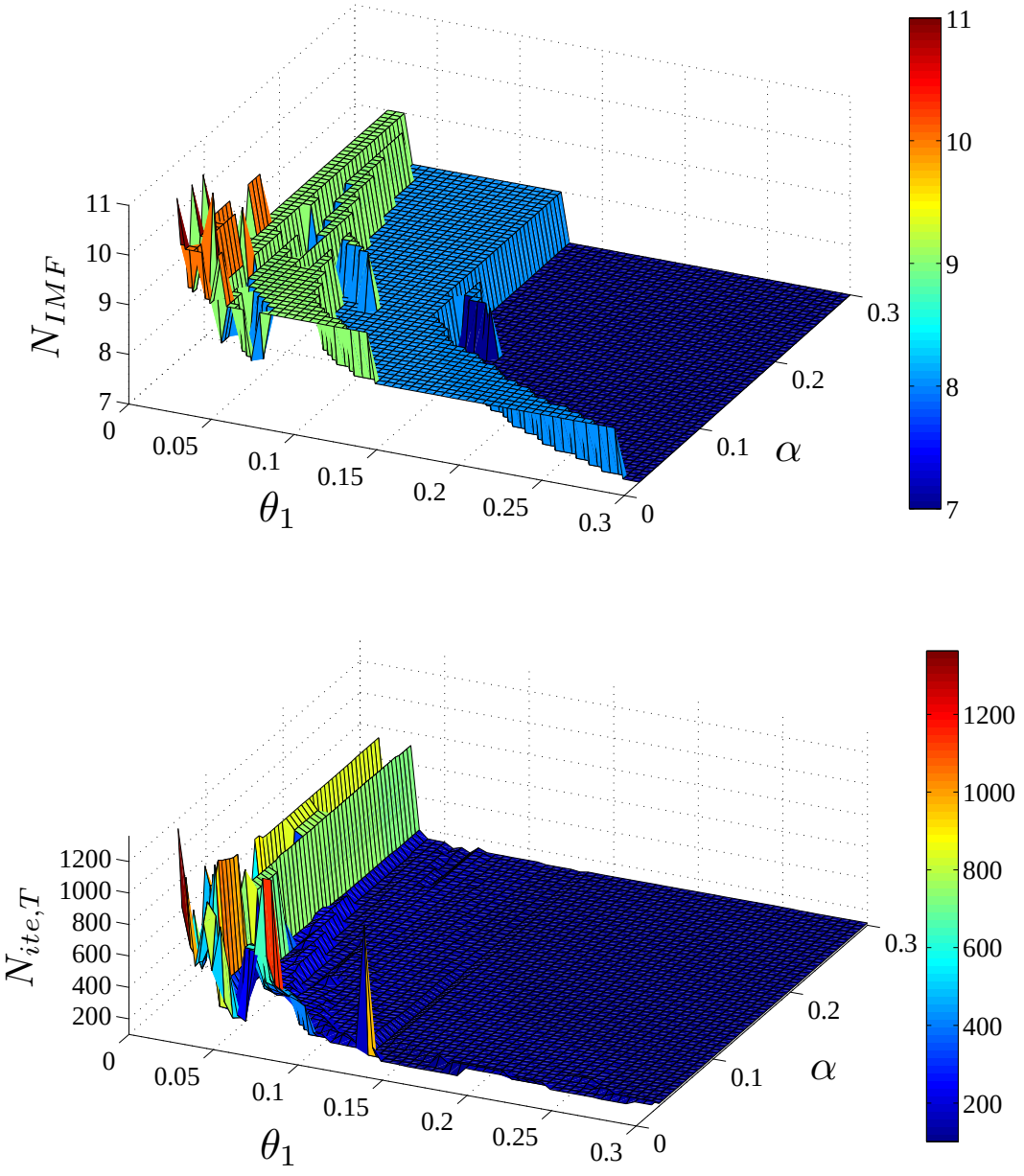

Figure E.2: Number of IMFs (top) and total number of iterations (bottom) versus (θ_1, α) for the study of the vortex-shedding data with the *second end-point option* and *without intermittency test*.

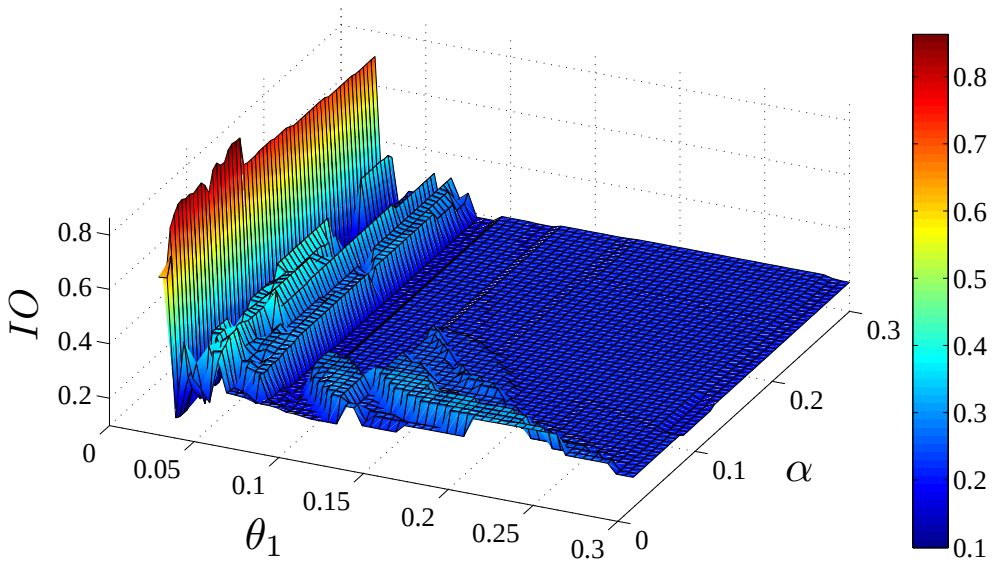

Figure E.3: Index of orthogonality versus (θ_1, α) for the study of the vortex-shedding data with the *third end-point option* and *without intermittency test*.

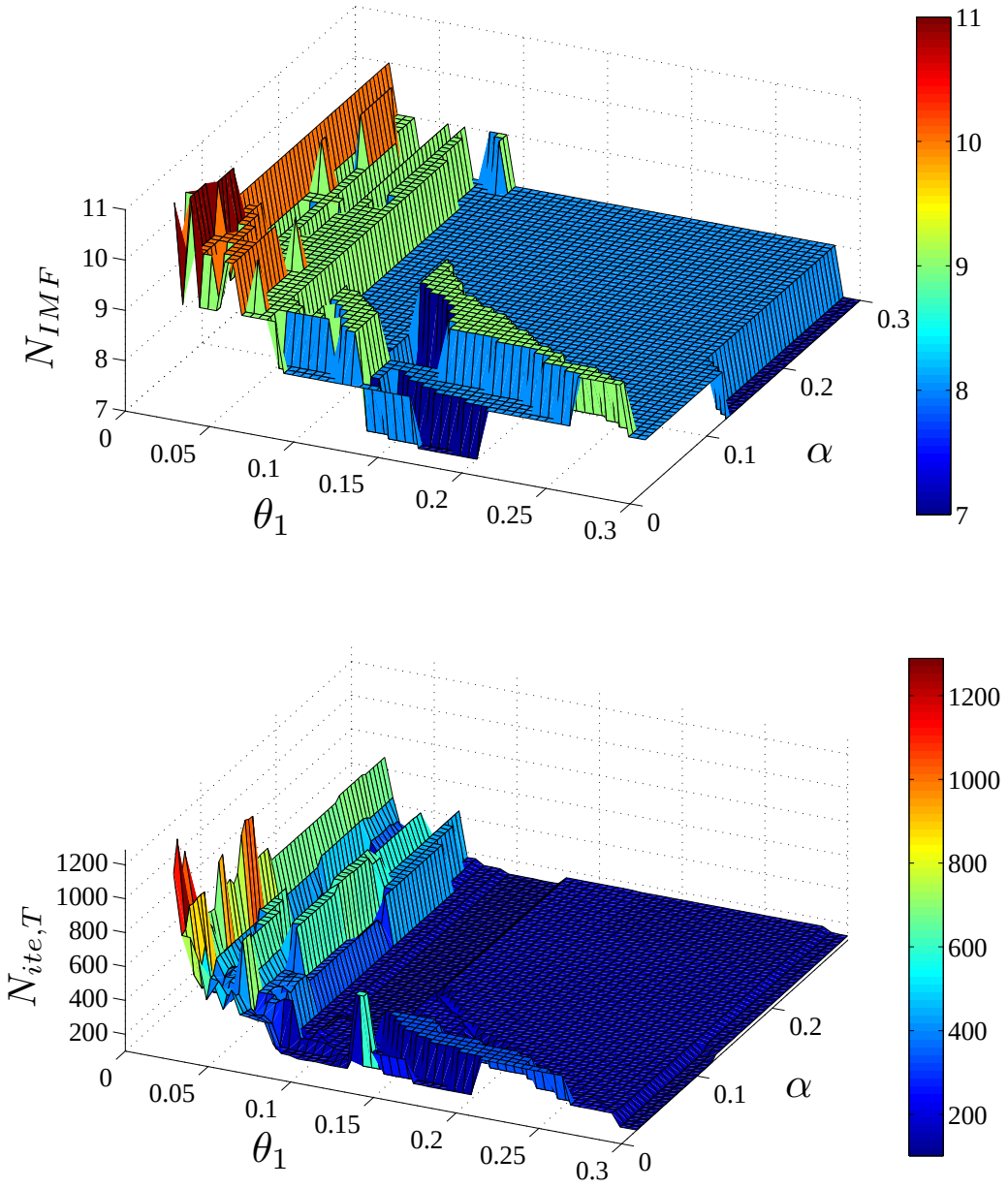

Figure E.4: Number of IMFs (top) and total number of iterations (bottom) versus (θ_1, α) for the study of the vortex-shedding data with the *third end-point option* and *without intermittency test*.

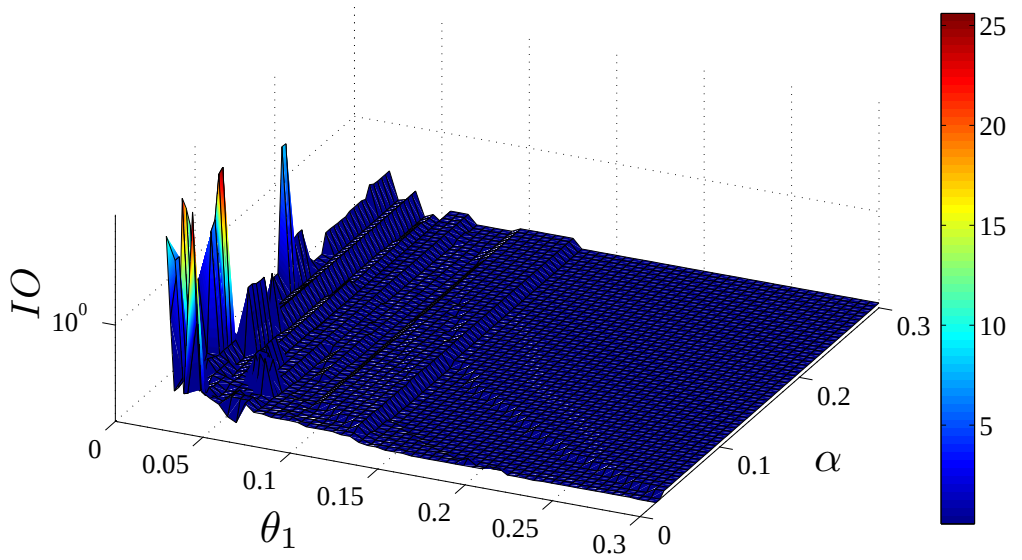

Figure E.5: Index of orthogonality versus (θ_1, α) for the study of the vortex-shedding data with the *fourth end-point option* and *without intermittency test*.

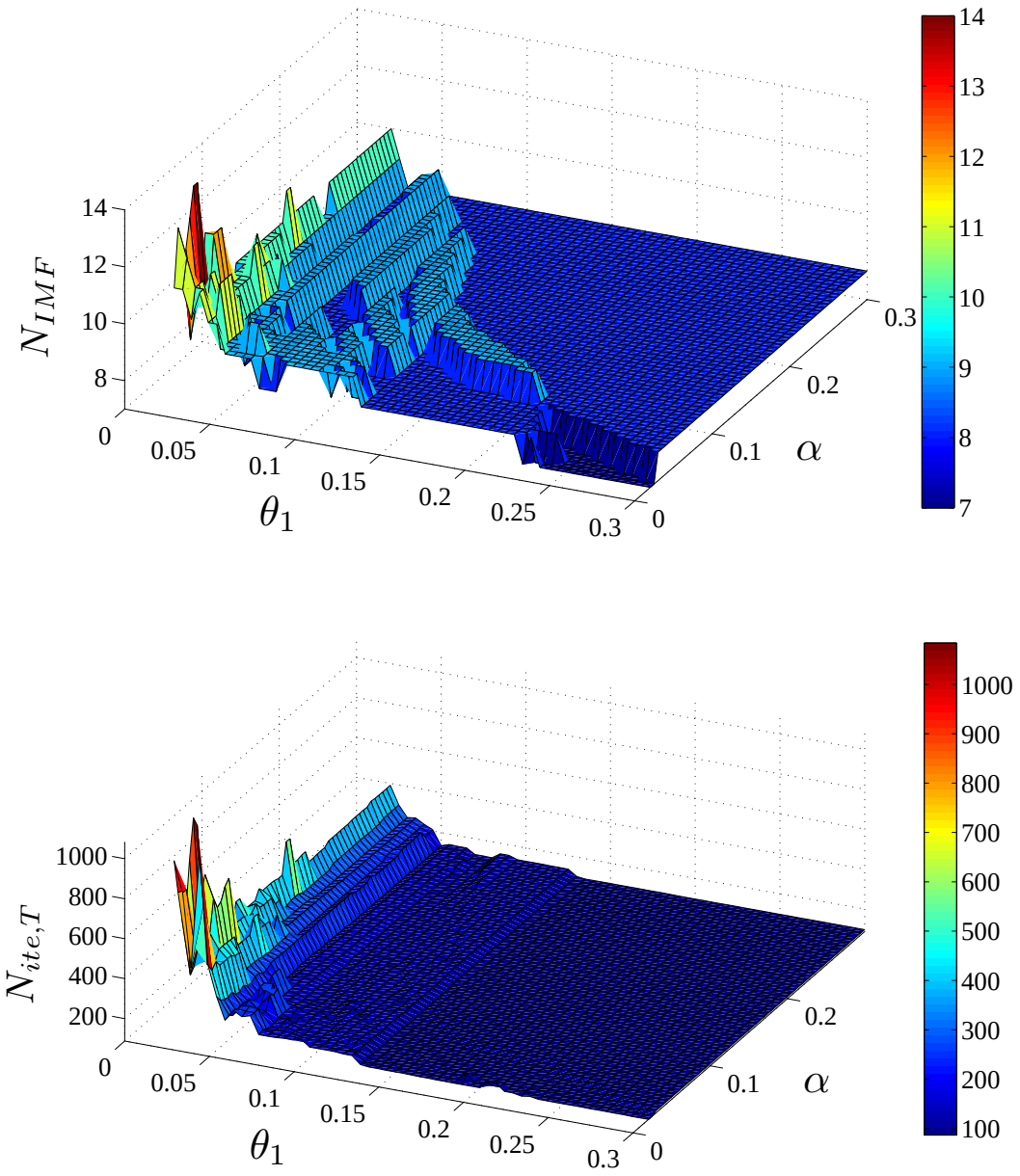

Figure E.6: Number of IMFs (top) and total number of iterations (bottom) versus (θ_1, α) for the study of the vortex-shedding data with the *fourth end-point option* and *without intermittency test*.

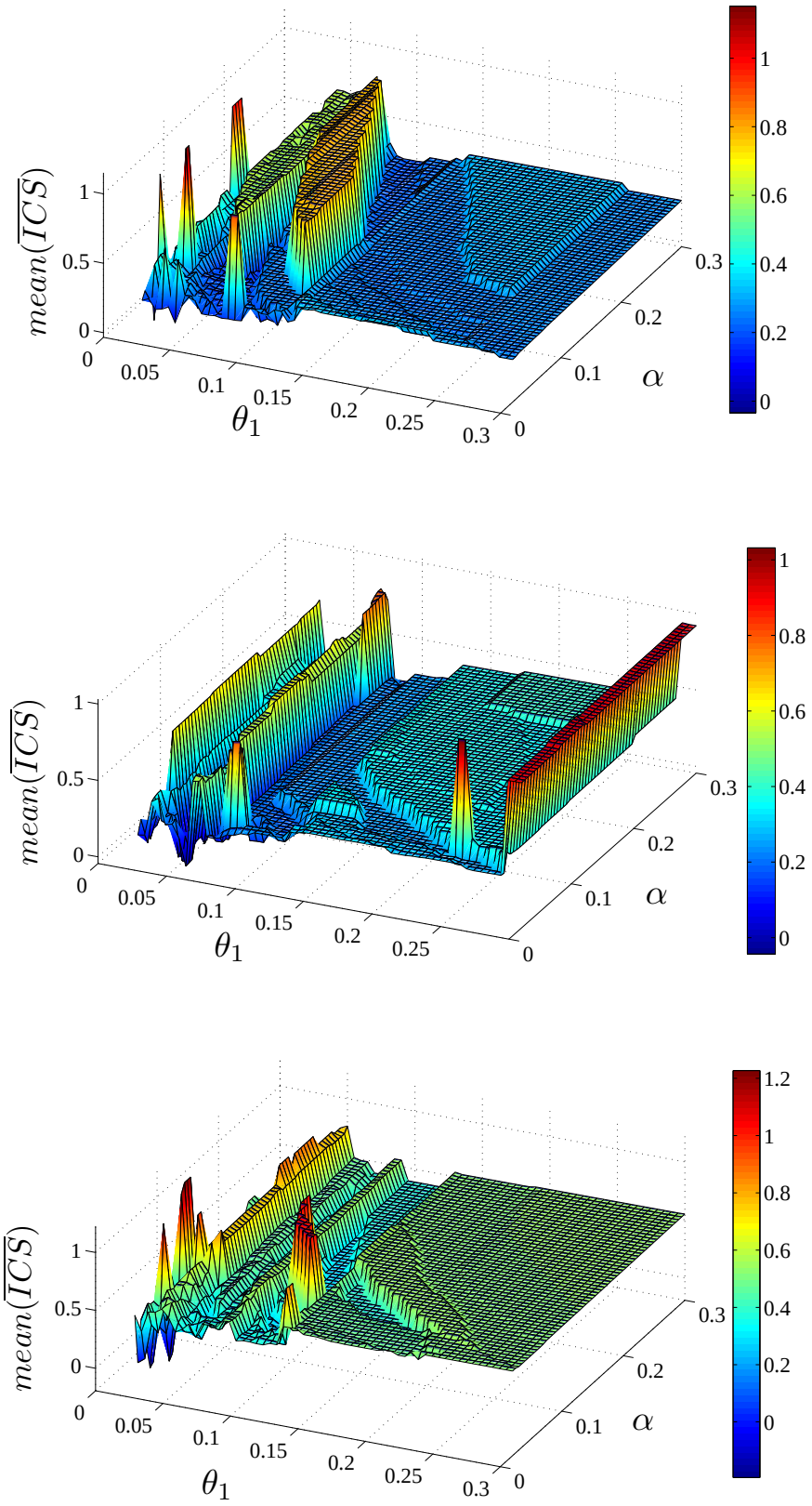

Figure E.7: Results of the mean of the average index of component separation, $mean(\overline{ICS})$, versus (θ_1, α) for the study of the vortex-shedding data with each end-point option and *without intermittency test*: top, second end-point option; middle, third end-point option; bottom, fourth end-point option.

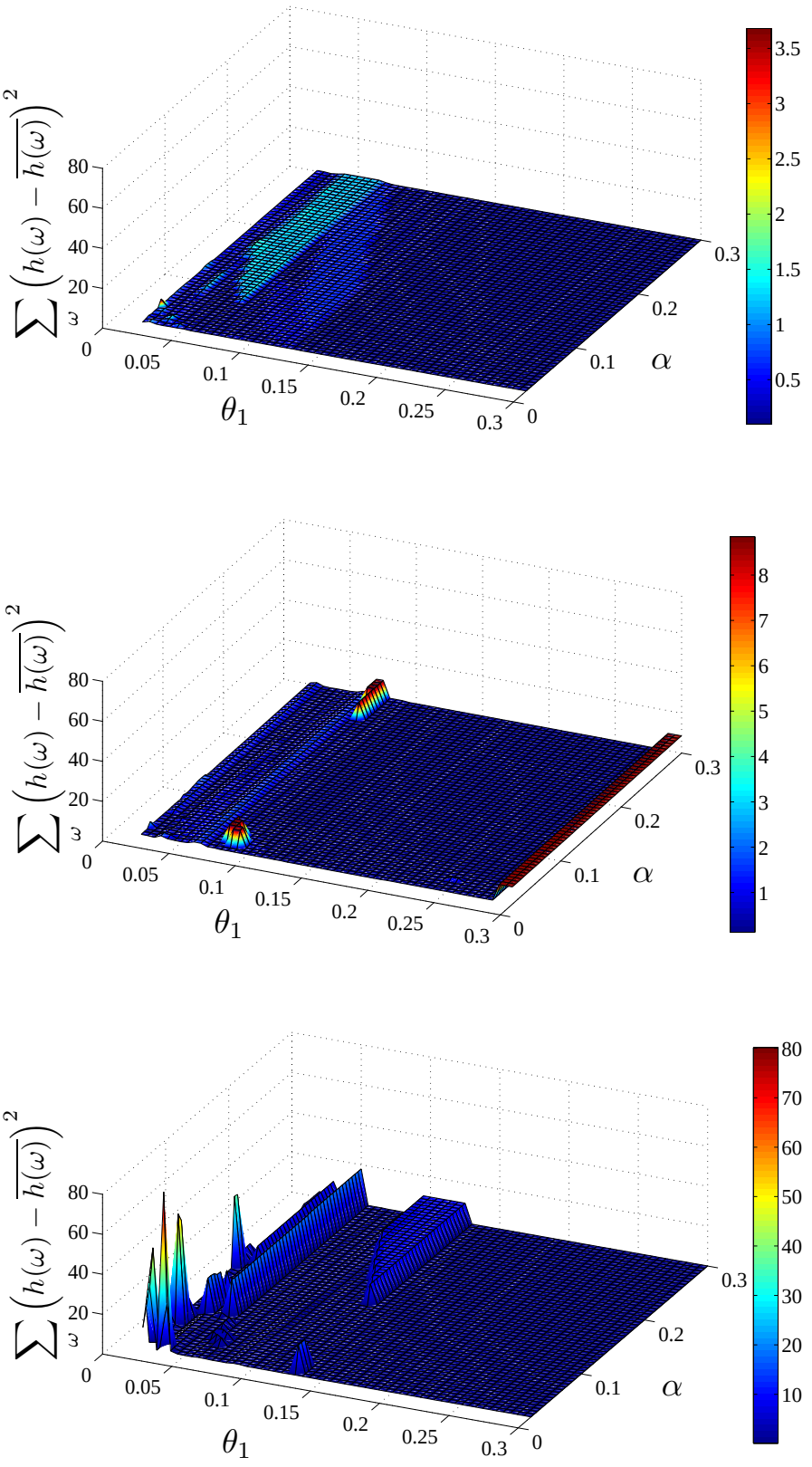

Figure E.8: Cumulative squared deviation between the mean marginal spectrum and marginal spectra of the vortex-shedding data according to the end-point option and *without intermittency test*: top, second end-point option; middle, third end-point option; bottom, fourth end-point option.

E.2 Vortex-shedding signal at Re=145

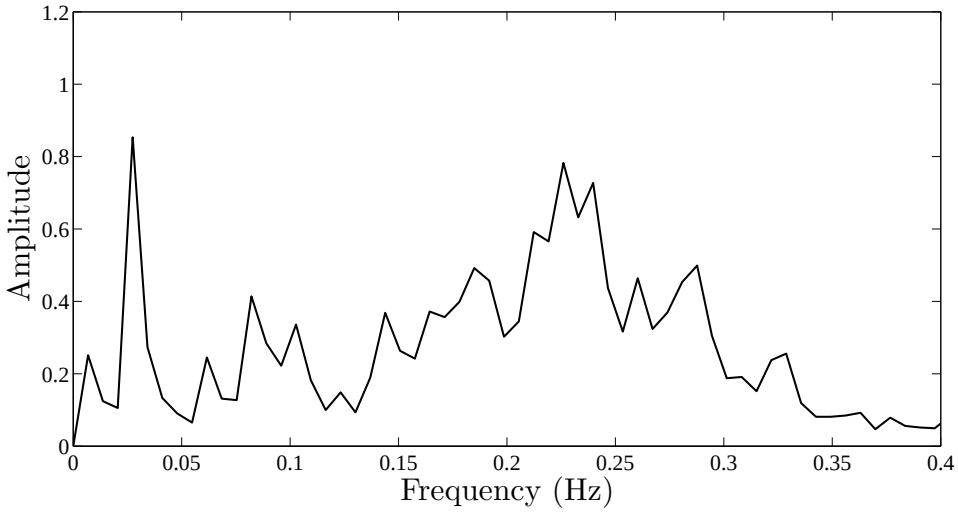

Figure E.9: Marginal spectrum of the vortex-shedding signal at $Re = 145$ obtained with `EMD([0:0.175:131], V-S 145,2,[0.05,0.5,0.05],[0,1.58])`. The vortex-shedding frequency is well retrieved at approximately $2F_{S,HHT} = 0.23$ Hz. However, likewise the signal at $Re = 105$, the main peak is wide, thus showing some frequency modulation.

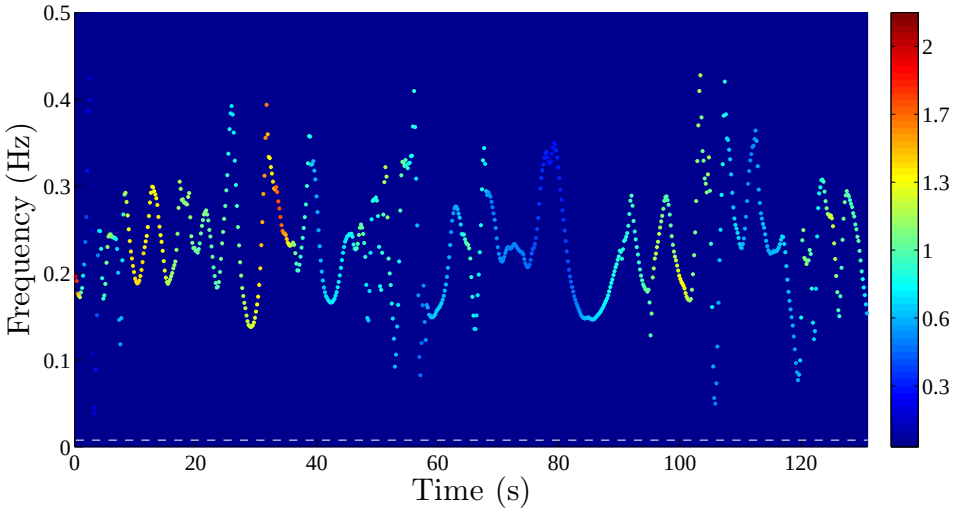

Figure E.10: Marginal spectrum of the third IMF of the vortex-shedding signal at $Re = 145$. Though the resolution is lower than with the signal at $Re = 105$, we can visualise the periodical frequency modulation of the instantaneous frequency ω_3 of $\pm 27\%$ with respect to the mean frequency $\overline{\omega_3} = 0.233$ Hz.

F. Frequency-modulated signal

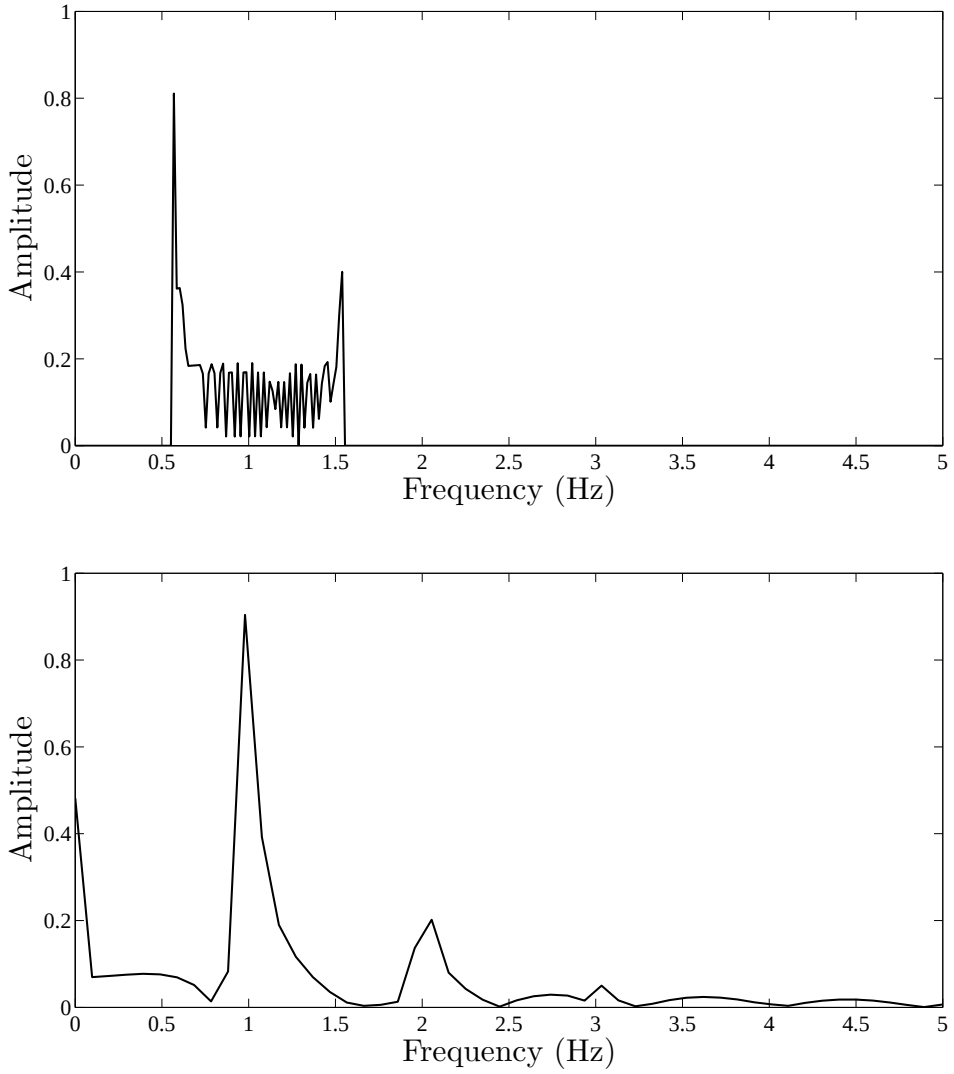

Figure F.1: Marginal spectrum (top) and Fourier spectrum (bottom) of the frequency-modulated signal presented in Paragraph 3.2.3. The marginal spectrum has recovered the whole bandwidth of the signal which varies from 0.5 Hz to 1.5 Hz accordingly to Equation 3.3. On the other hand, the Fourier spectrum fails to retrieve the nonlinear character of the signal, the instantaneous frequency is averaged, as shown by the main peak at 1 Hz, and spurious harmonics are created, as shown by the secondary peaks at 2 and 3 Hz.

G. Optimal implementation options

Table G.1 shows the optimal implementation options found for each signal studied. These findings are based on the results obtained with the four quantitative criteria, the analysis of the IMFs and the study of the Hilbert spectrum. The end-point options 1 to 4 are respectively the clamped end-point technique, the extrema extension technique, the mirror imaging extension and the damped sinusoidal extension based on an auto-regressive model (AR model). IT means intermittency test.

Table G.1: Optimal implementation options for each signal studied.

Signal	End-point option	Stopping criteria	Extension option	Comments
2-component signal	4	strict e.g. $(0.01, 0.1, 0.01)$	AR model	-
AM signal	2 or 4	loose e.g. $(0.1, 1, 0.1)$	AR model	-
FM signal	2 or 3	any	AR model	-
amplitude step signal	2	loose e.g. $(0.1, 1, 0.1)$	AR model	-
frequency step signal	2 or 3	any	AR model	-
LOD data	4	intermediate e.g. $(0.05, 0.5, 0.05)$	AR model	with IT
vortex shedding signal	4	loose e.g. $(0.095, 0.95, 0.125)$	AR model	with IT