A Plea for Adaptive Data Analysis:

An introduction to EMD

Norden E. Huang

Research Center for Adaptive Data Analysis

National Central University

Academia Sinica

2010/10/13

Traditional Data Analysis

In pursue of mathematatic rigor and certainty, however, we are forced to to live in a pseudo-real world, in which all processes are

Linear and Stationary

Available ‘Data Analysis’ Methods

for Nonstationary (but Linear) time series

- Spectrogram
- Wavelet Analysis
- Wigner-Ville Distributions
- Empirical Orthogonal Functions aka Singular Spectral Analysis
- Moving means
- Successive differentiations

Available ‘Data Analysis’ Methods

for Nonlinear (but Stationary and Deterministic) time series

- Phase space method
  - Delay reconstruction and embedding
  - Poincaré surface of section
  - Self-similarity, attractor geometry & fractals
- Nonlinear Prediction
- Lyapunov Exponents for stability

Typical Apologia

- Assuming the process is stationary ....
- Assuming the process is locally stationary ....
- As the nonlinearity is weak, we can use perturbation approach ....

Though we can assume all we want, but
the reality cannot be bent by the assumptions.

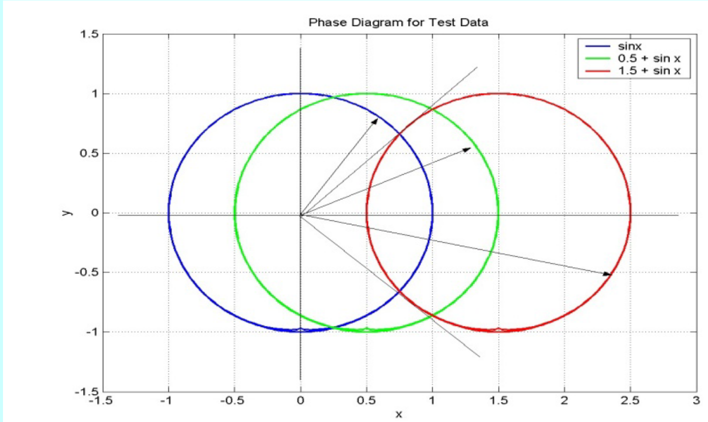
Motivations for alternatives:
Problems for Traditional Methods

- Physical processes are mostly nonstationary
- Physical Processes are mostly nonlinear
- Data from observations are invariably too short
- Physical processes are mostly non-repeatable.

Ensemble mean impossible, and temporal mean might not be meaningful for lack of stationarity and ergodicity.
Traditional methods are inadequate.

Hilbert Transform : Definition

$$For\ any\ \ x(t) \in L^p\ ,$$



$$y(t) = \frac{1}{\pi} \wp \int_{\tau} \frac{x(\tau)}{t-\tau} d\tau\ ,$$

then, $x(t)$ and $y(t)$ form the analytic pairs:

$$z(t) = x(t) + i\,y(t) = a(t)\,e^{i\theta(t)}\ ,$$

where

$$a(t) = \left(x^2 + y^2\right)^{1/2}\ \ and\ \ \theta(t) = tan^{-1}\frac{y(t)}{x(t)}\ .$$

Instantaneous Frequency

$$Velocity = \frac{distance}{time} \ ; \ mean \ \ velocity$$

$$Newton \Rightarrow v = \frac{dx}{dt}$$

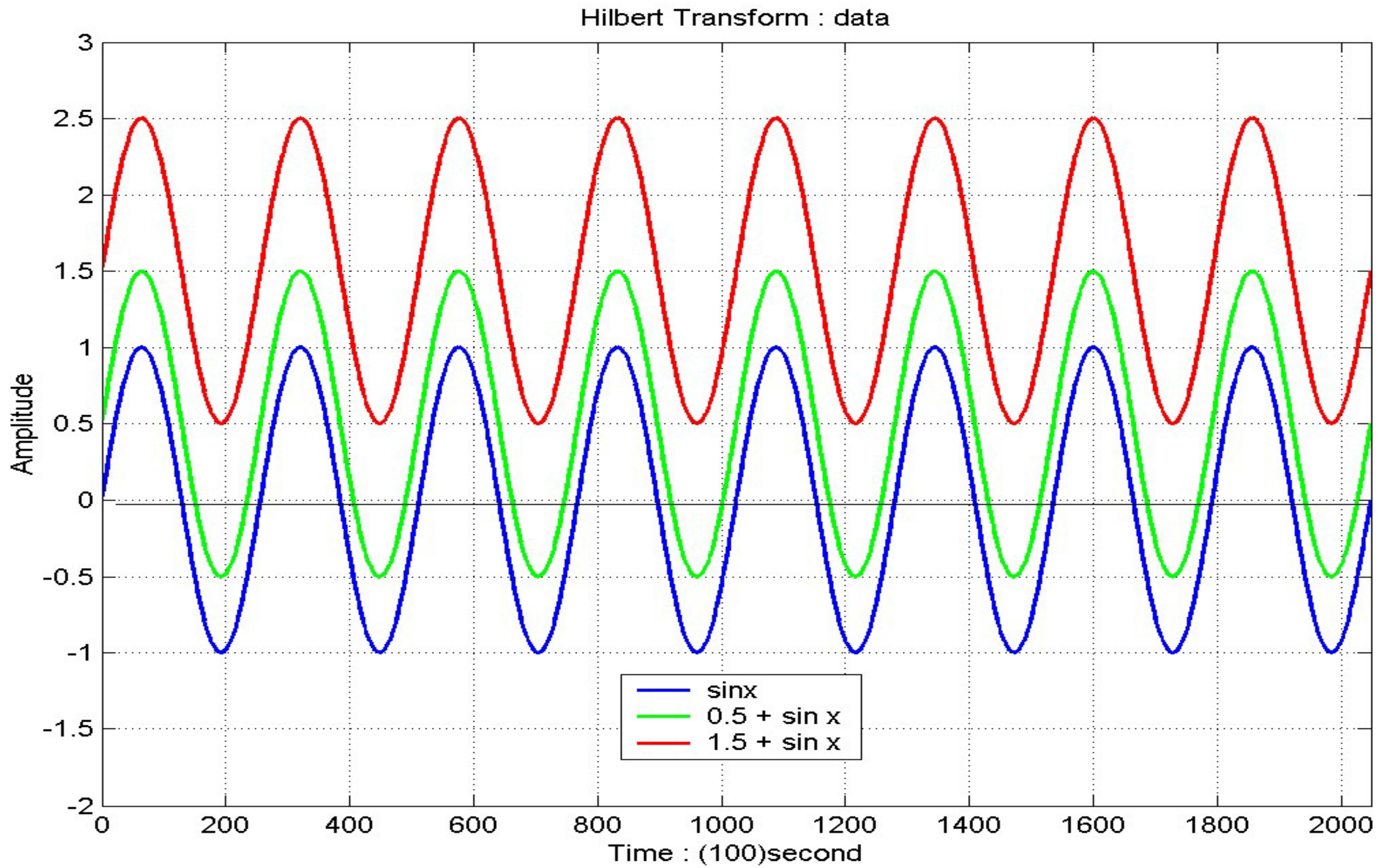
$$Frequency = \frac{1}{period} \ ; \ mean \ \ frequency$$

$$HHT \ defines \ the \ \ phase \ \ function \ \Rightarrow \ \omega = \frac{d\theta}{dt}$$

So that both v and ω can appear in differential equations.

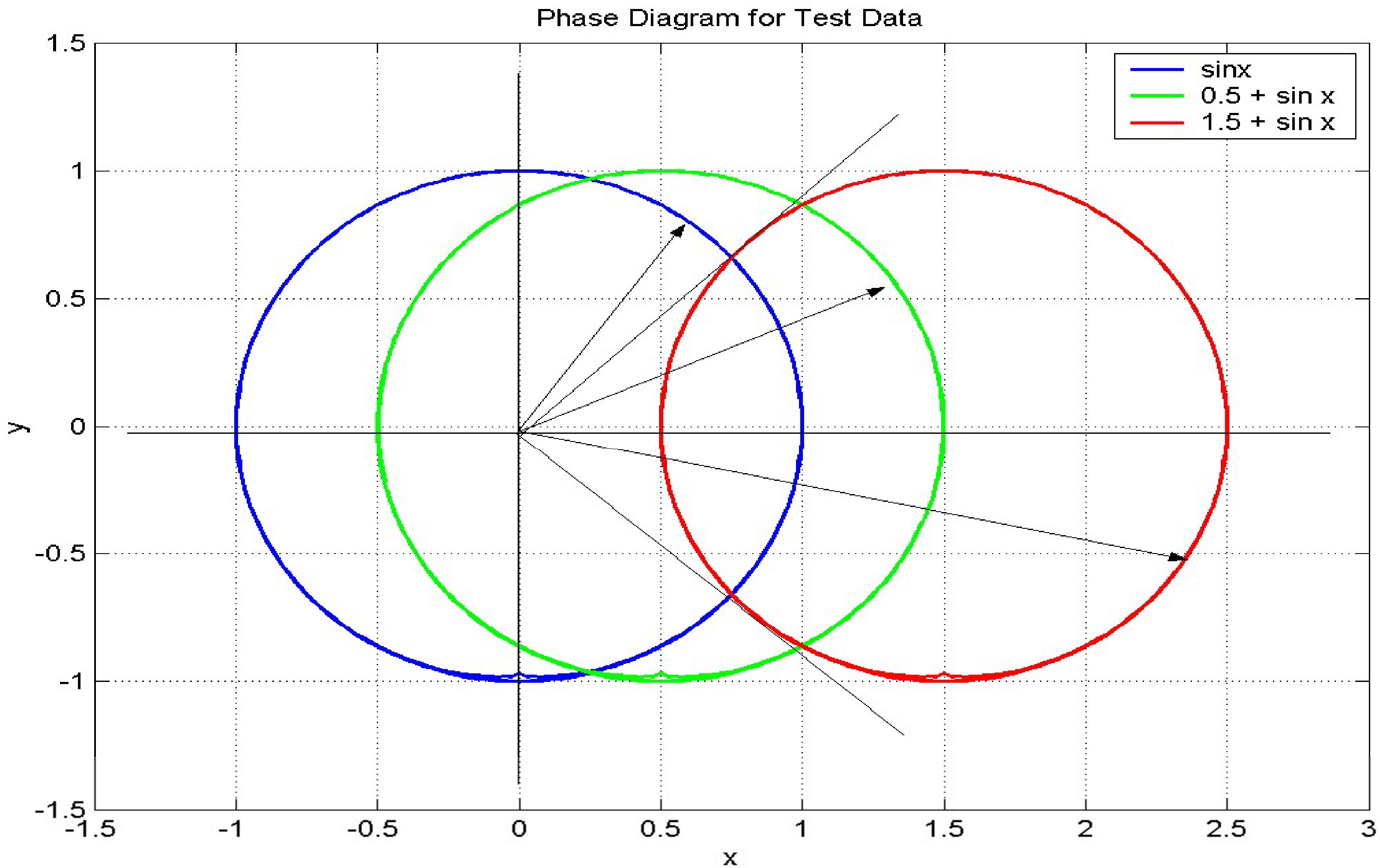
Why the traditional approach does not work?

Hilbert Transform *a cos* + *b* : *Data*



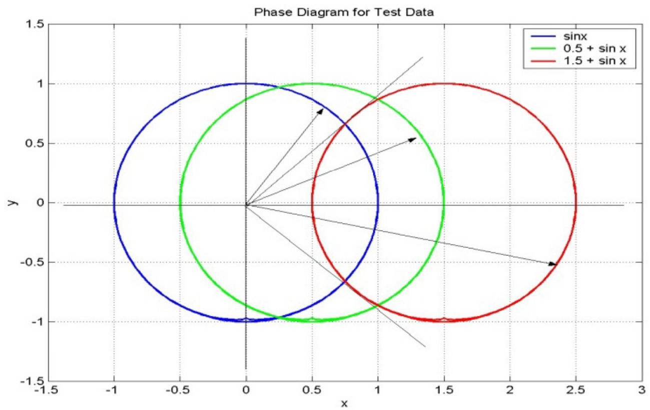
Hilbert Transform *a cos* + *b* :

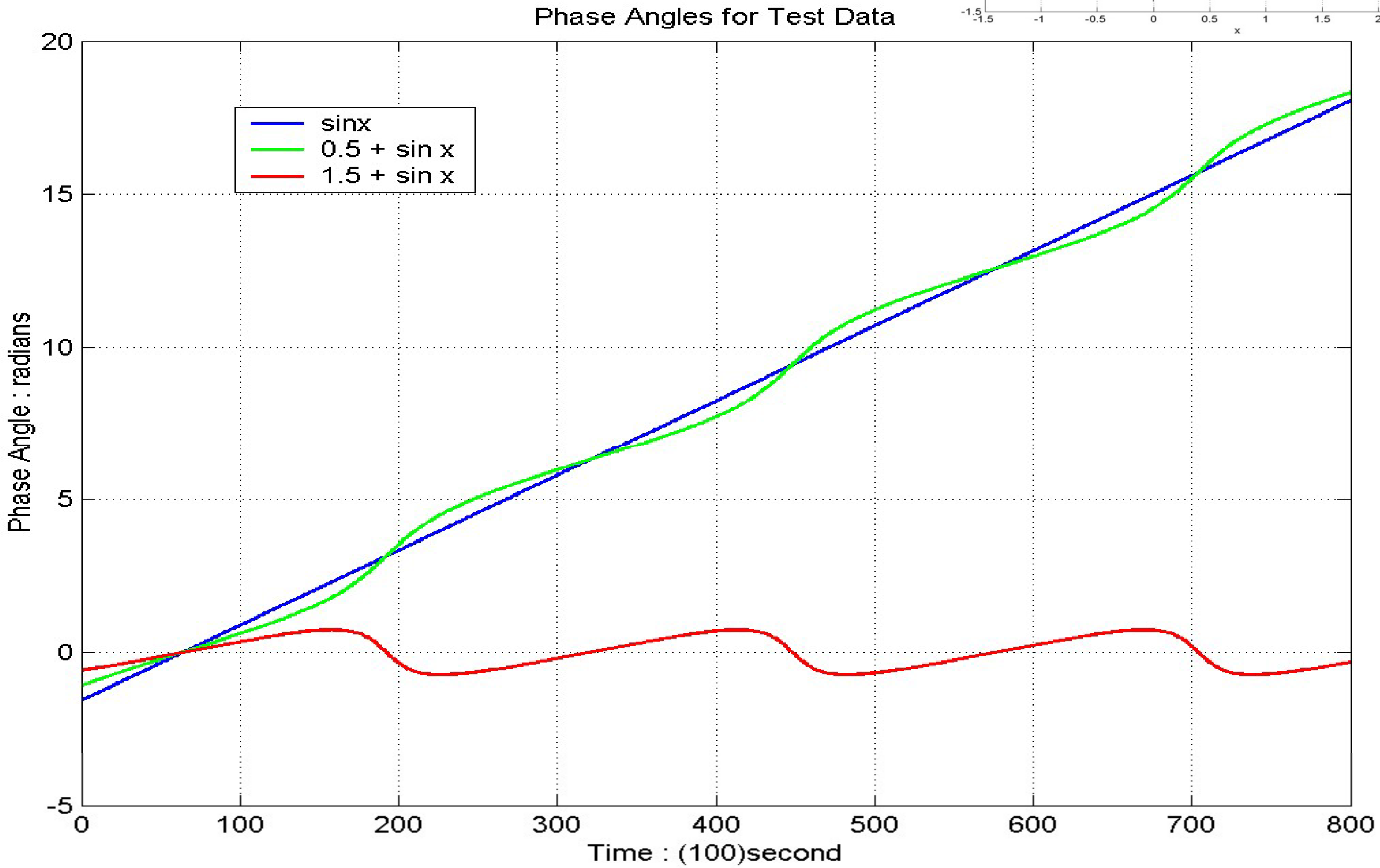
Phase Diagram



Hilbert Transform *a c*

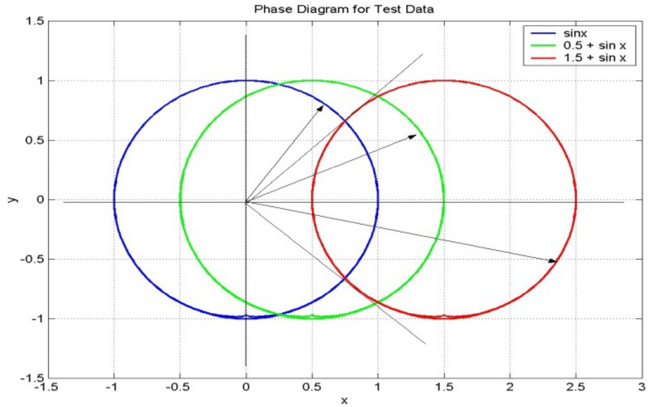
Phase Angle Deta

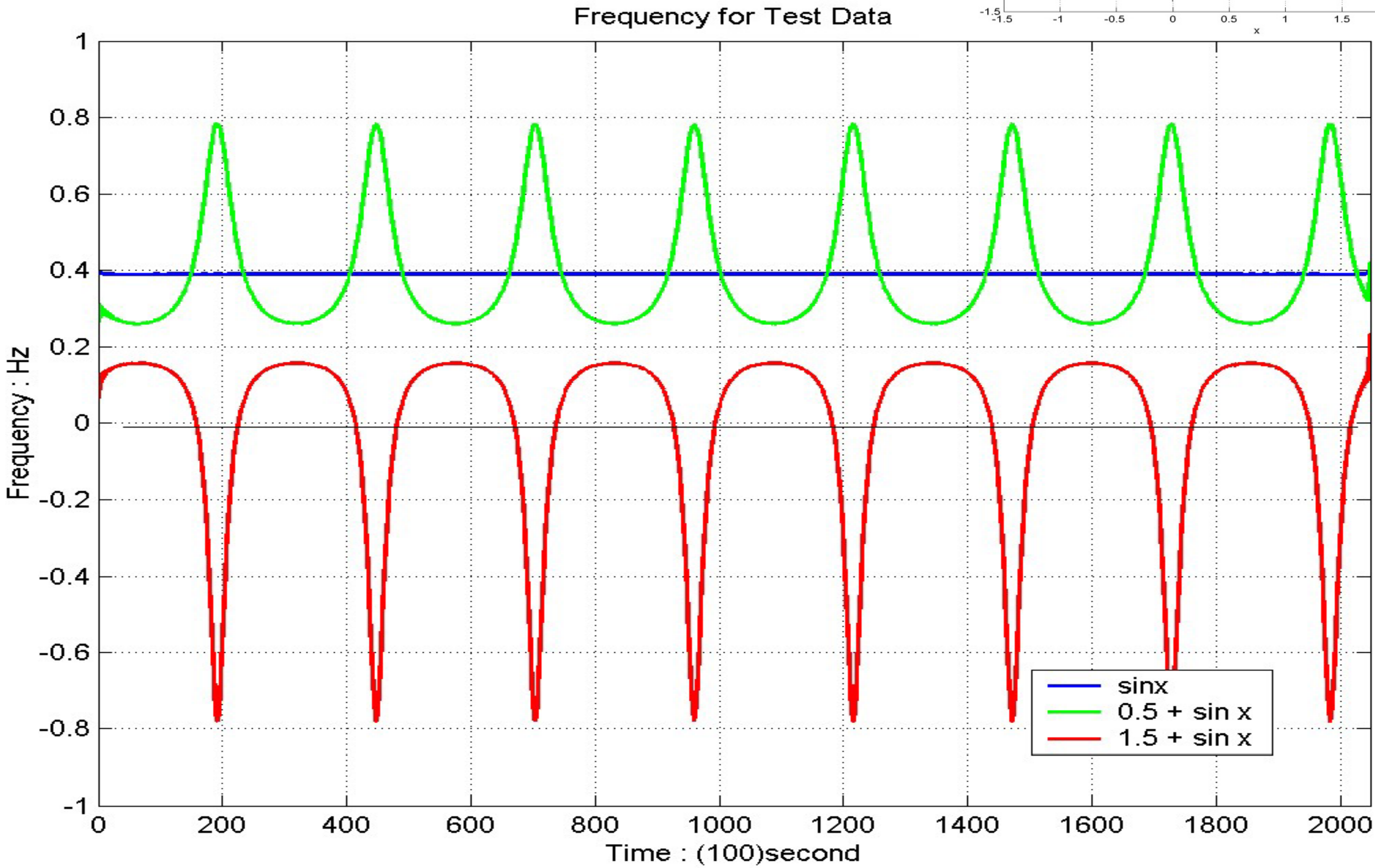




Hilbert Transform *a cos*

Frequency





Hilbert-Huang Transform (HHT)

- Empirical Mode Decomposition (EMD)
  - Decompose multiscale data into Intrinsic Mode Functions (IMF)
- Hilbert Spectrum Analysis
  - Hilbert Transform of IMF

    -> Analytic Signal

    -> Instantaneous Amplitude & Phase

    -> Instantaneous Frequency

    -> Hilbert Spectrum

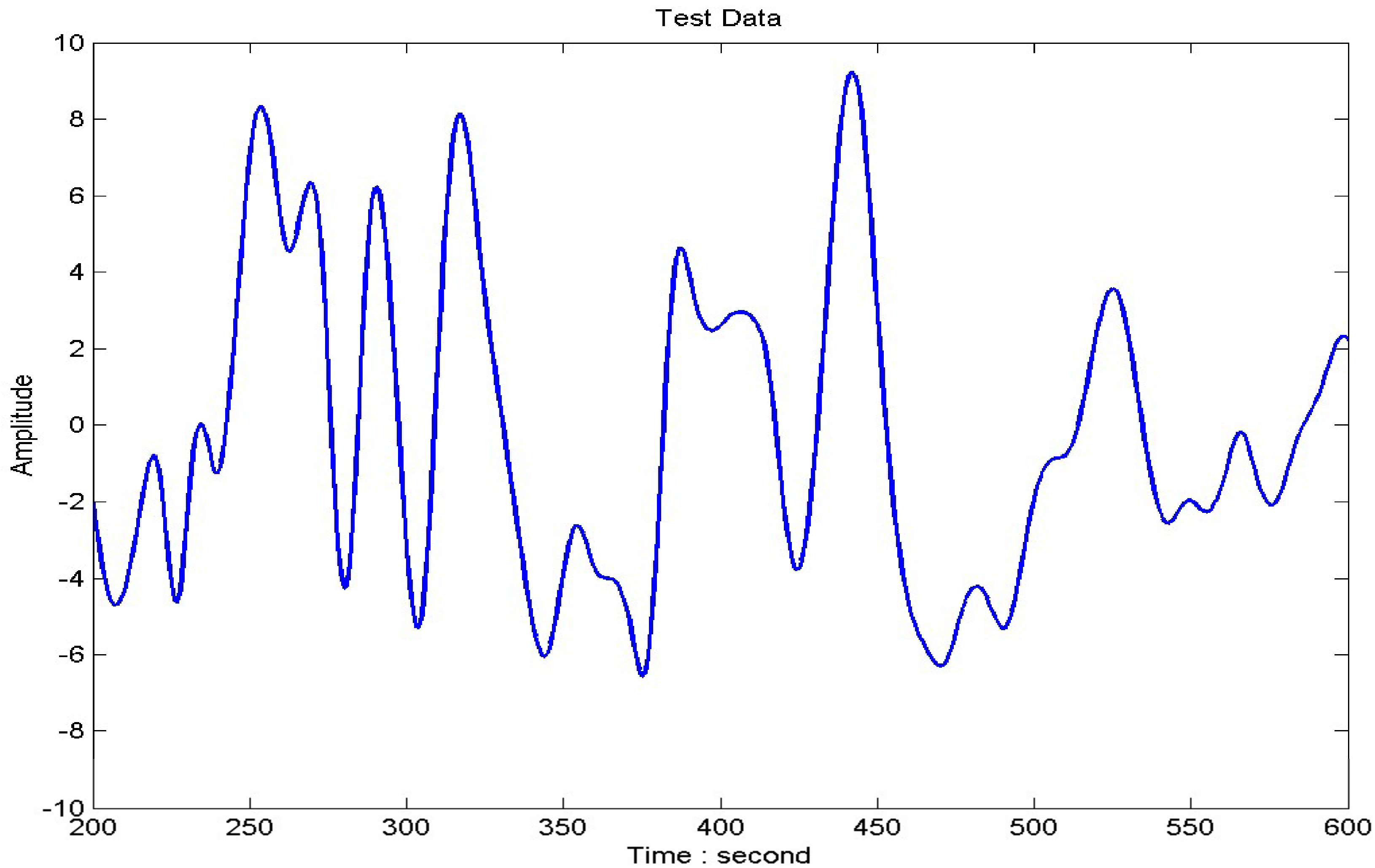
    -> Marginal Spectrum

The Empirical Mode Decomposition Method and Hilbert Spectral Analysis

Sifting

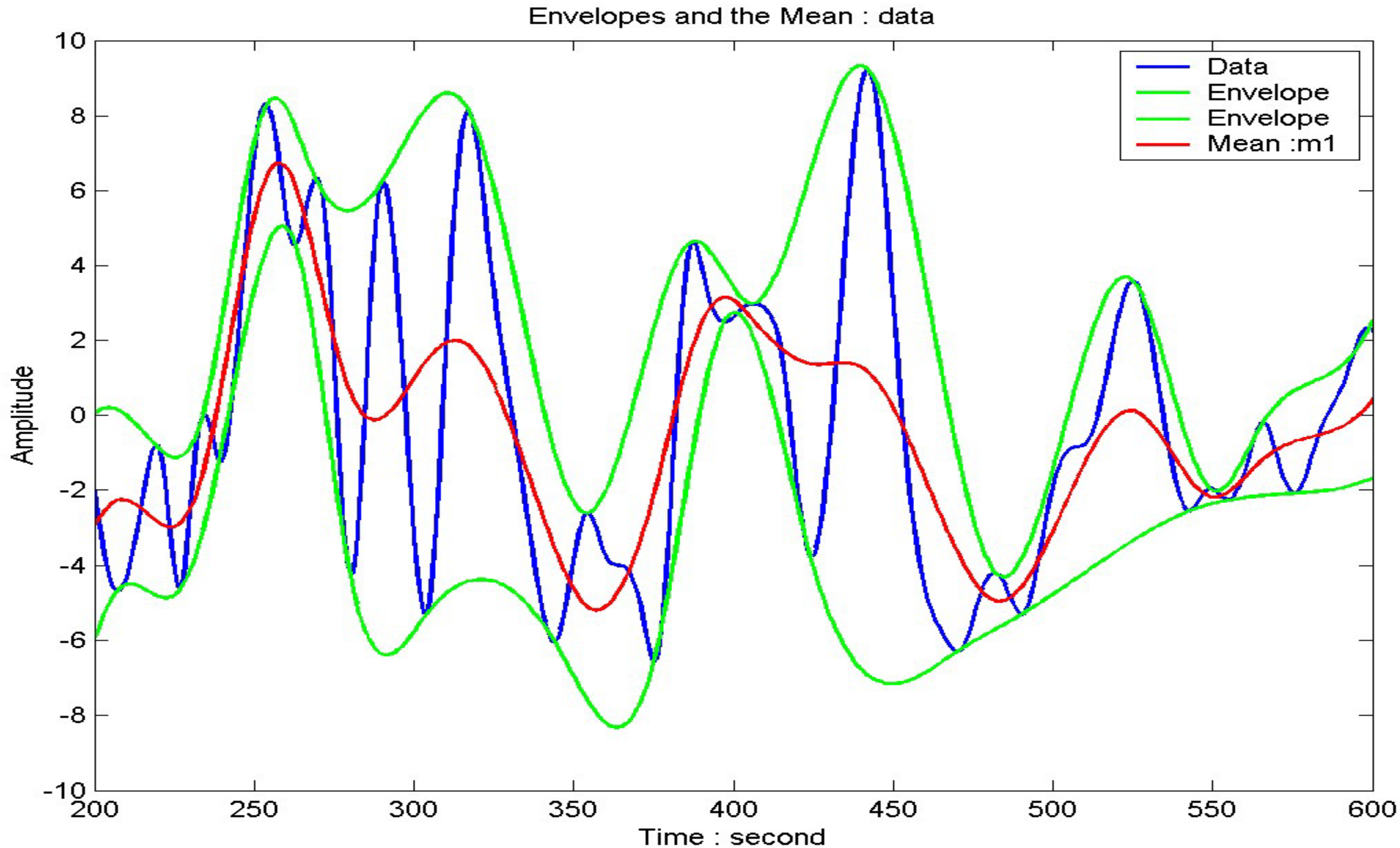
Empirical Mode Decomposition:

Methodology : Test Data



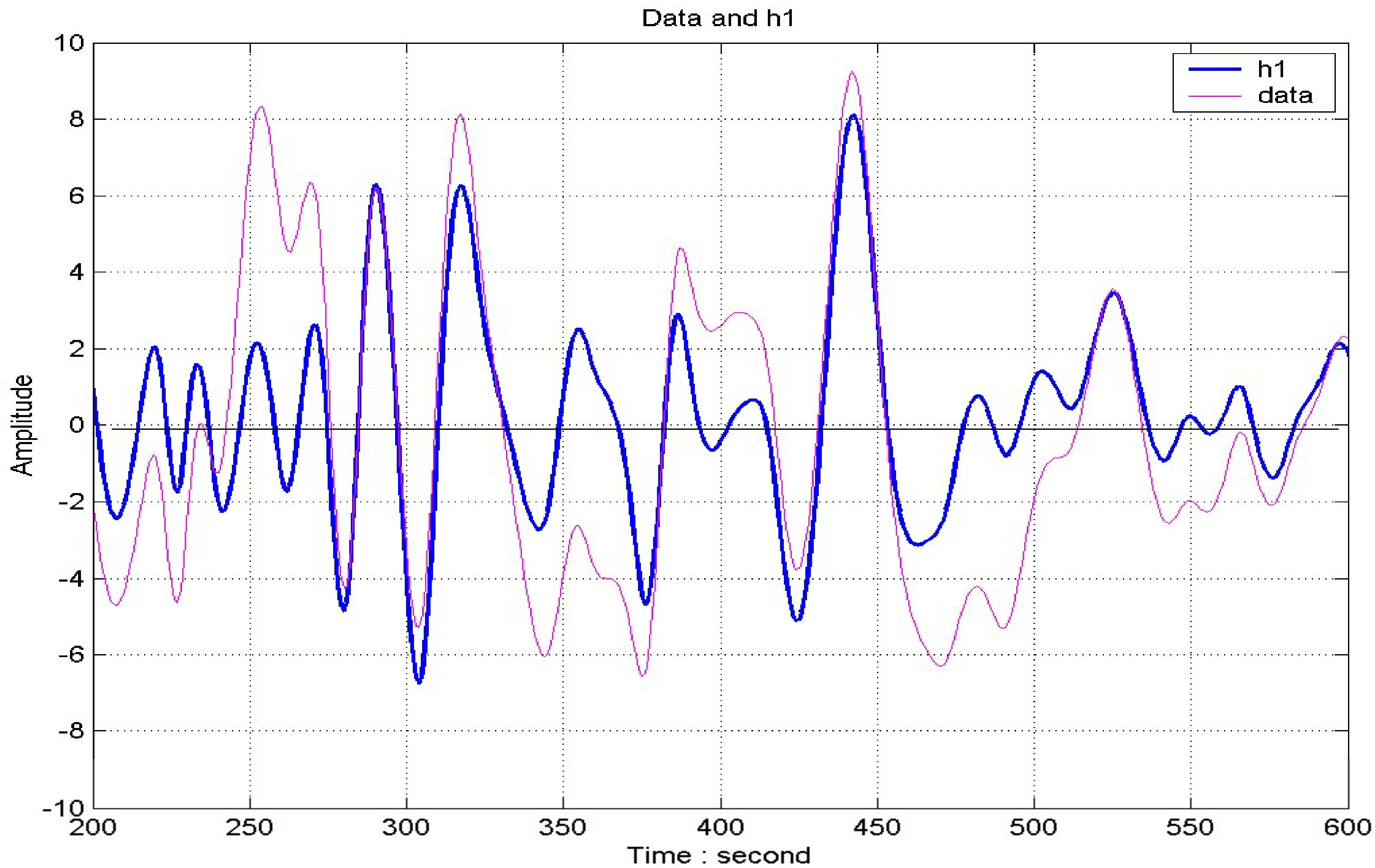
Empirical Mode Decomposition:

Methodology : data and m1



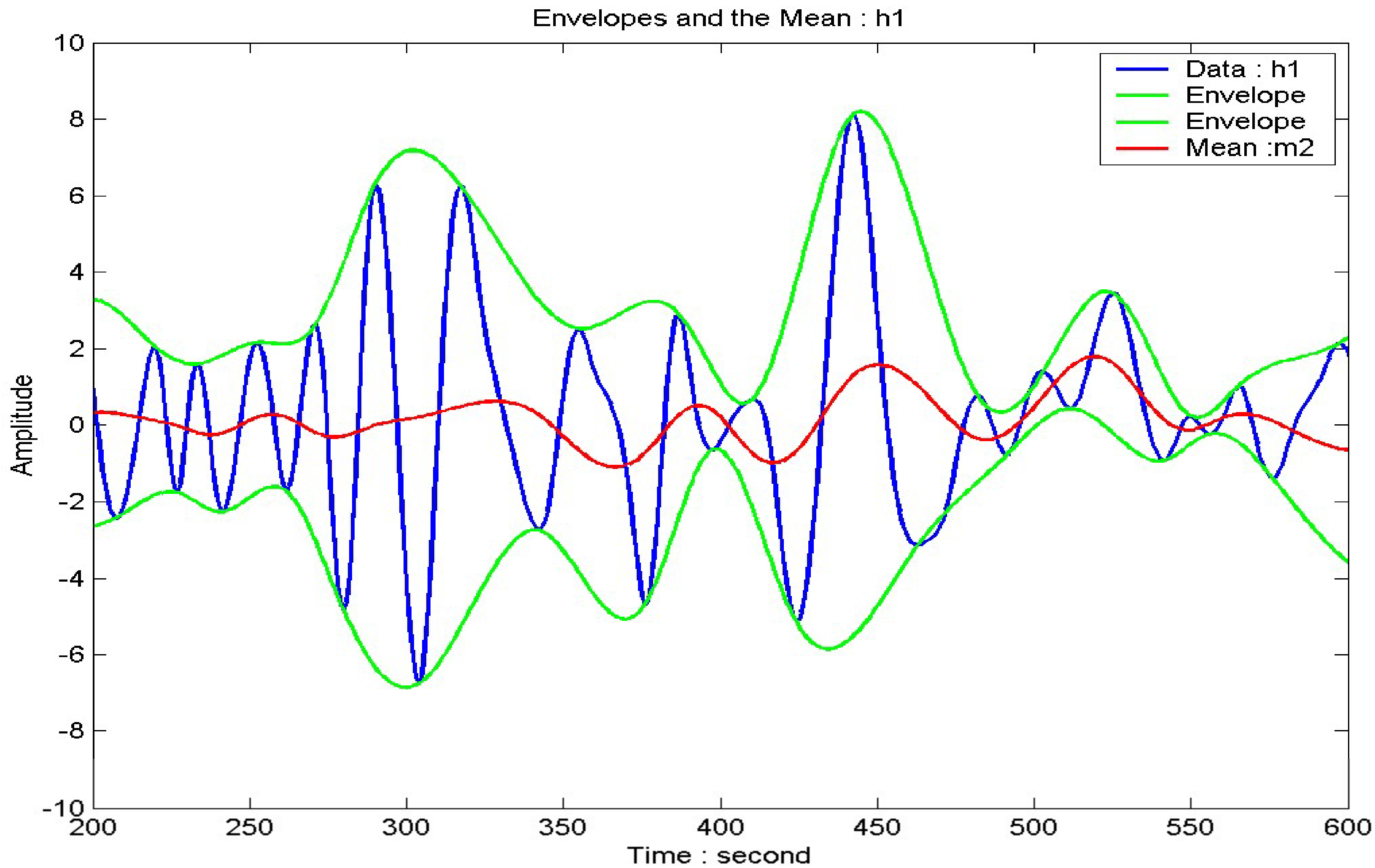
Empirical Mode Decomposition:

Methodology : data & h1



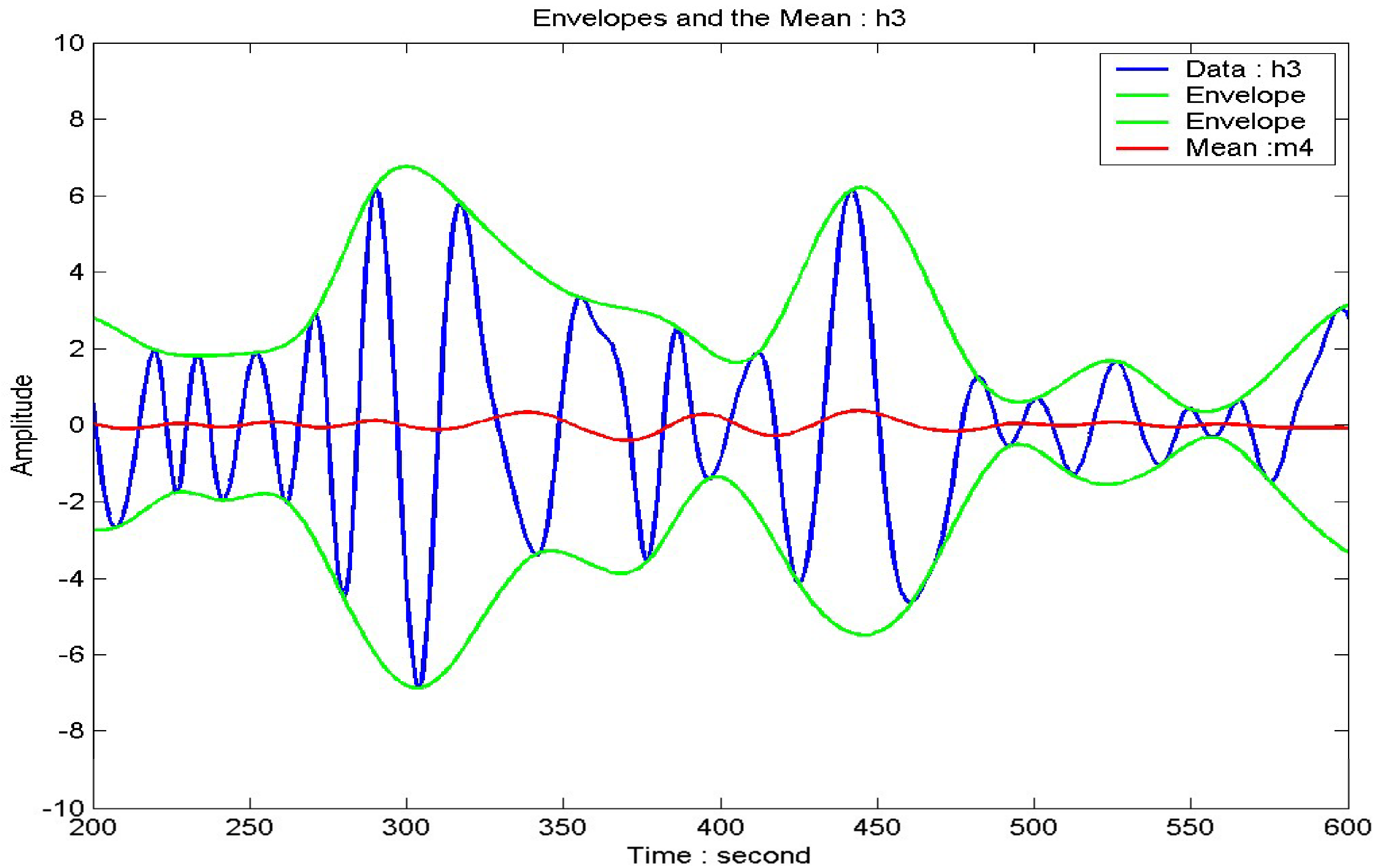
Empirical Mode Decomposition:

Methodology : h1 & m2



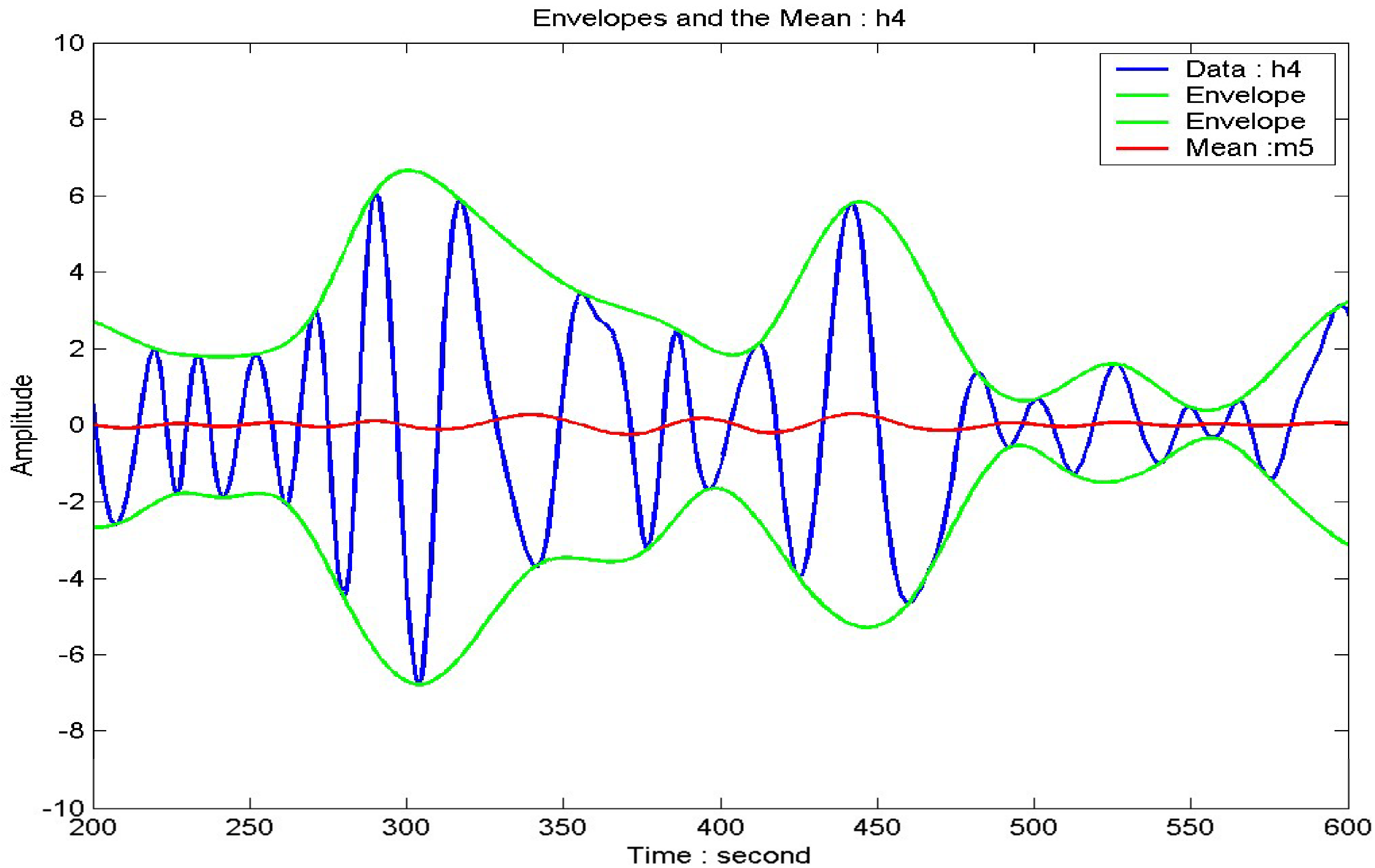
Empirical Mode Decomposition:

Methodology : h3 & m4



Empirical Mode Decomposition:

Methodology : h4 & m5



Empirical Mode Decomposition

Sifting : to get one IMF component

$$x(t) - m_1 = h_1 \ ,$$

$$h_1 - m_2 = h_2 \ ,$$

$$\cdots\cdots$$

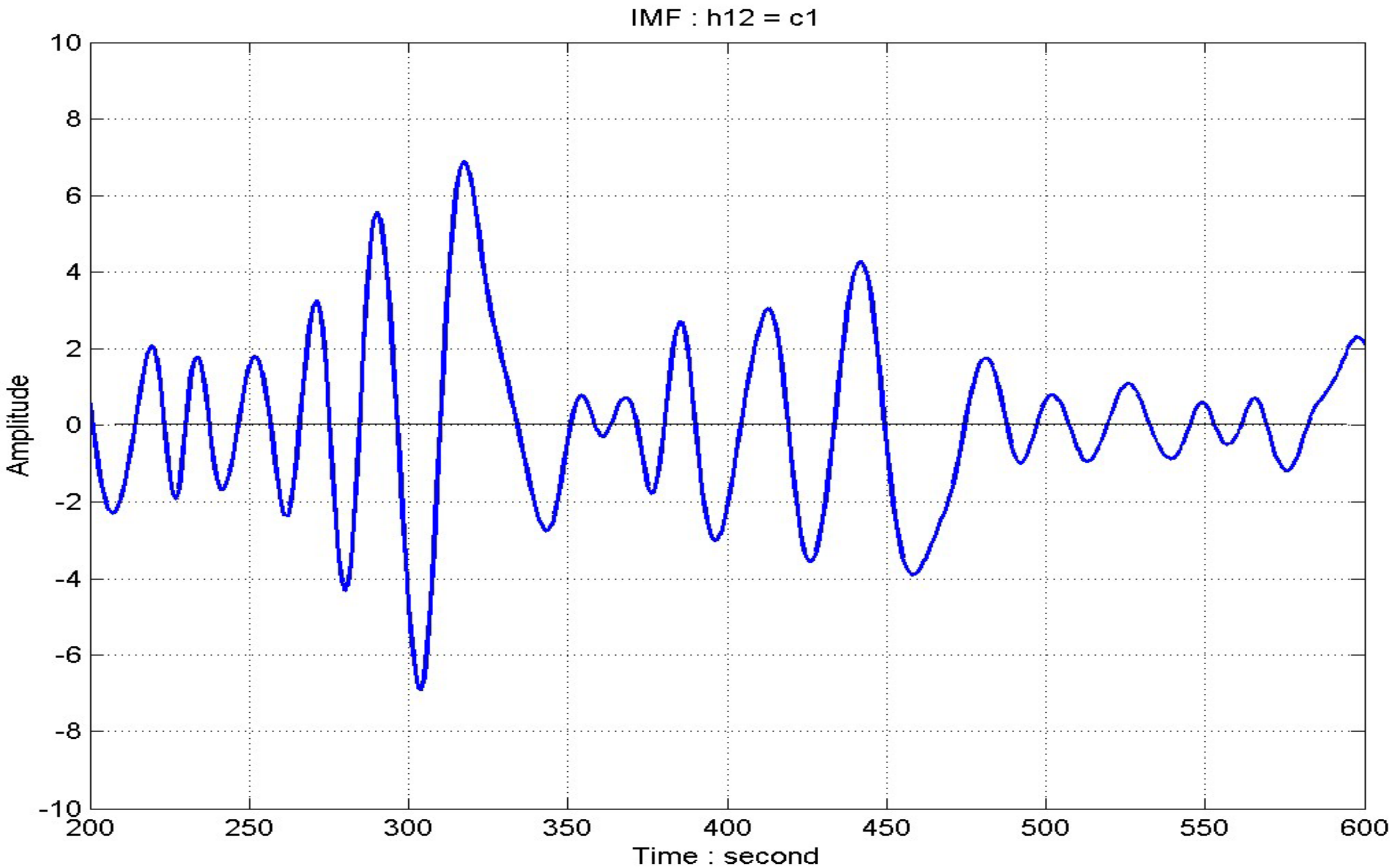
$$\cdots\cdots$$

$$h_{k-1} - m_k = h_k \ .$$

$$\Rightarrow \quad h_k = c_1 \ .$$

Empirical Mode Decomposition:

Methodology : IMF c1



The Stoppage Criteria

The Cauchy type criterion: when SD is small than a pre-set value, where

$$SD = \frac{\sum_{t=0}^{T} \left| h_{k-1}(t) - h_k(t) \right|^2}{\sum_{t=0}^{T} h_{k-1}^{\;2}(t)}$$

Or, **simply pre-determine the number of iterations and many other alternatives**.

Effects of Sifting

- To find the IMF component so that Instantaneous Frequency could be calculated through removing the ridding waves
- But it also reduces amplitude variations (Theorem by Wang Gang, et al. 2010)
- Most importantly, to find the mean for a non-stationary time series through a Local Reynolds Mean.

Definition of the Intrinsic Mode Function (IMF): a necessary condition only!

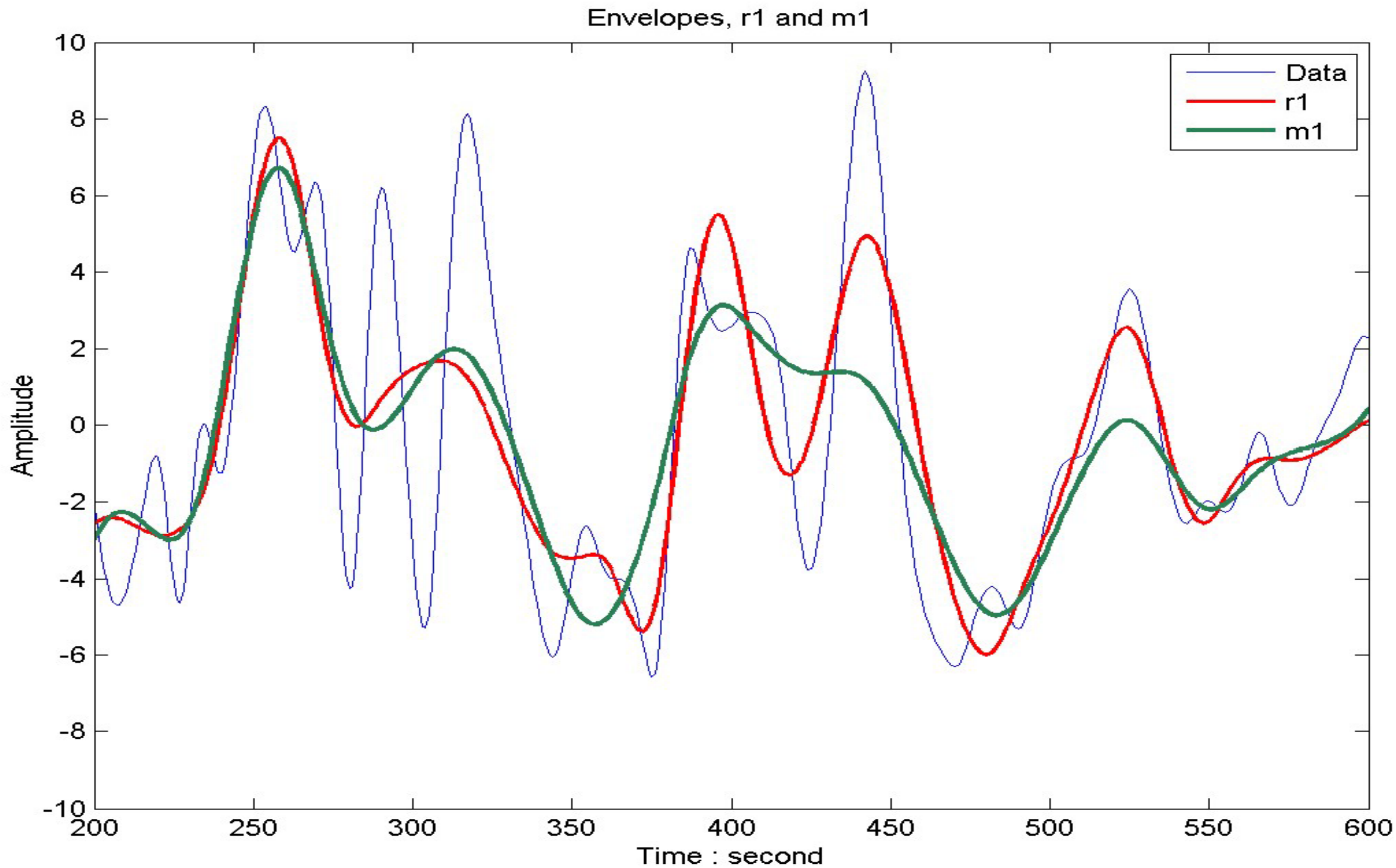
Any function having the same numbers of zero – cros sin gs and extrema, and also having symmetric envelopes defined by local max ima and min ima respectively is defined as an Intrinsic Mode Function (IMF).

All IMF enjoys good Hilbert Transform :

$$\Rightarrow \Rightarrow \quad c(t) = a(t) e^{i\theta(t)}$$

Empirical Mode Decomposition:

Methodology : data, r1 and m1



Empirical Mode Decomposition

Sifting : to get all the IMF components

$$x(t) - c_1 = r_1 \ ,$$

$$r_1 - c_2 = r_2 \ ,$$

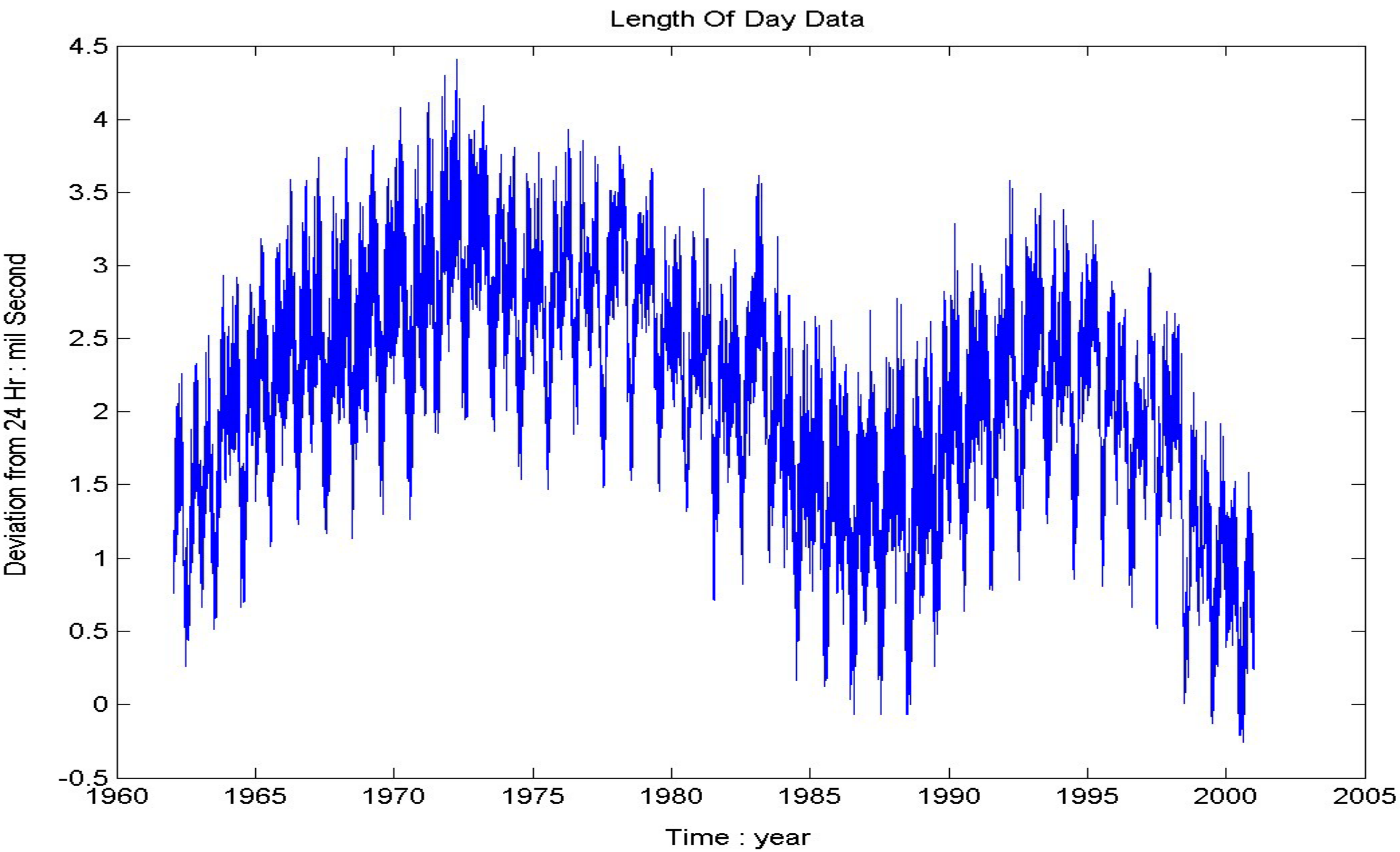
$$\cdots$$

$$r_{n-1} - c_n = r_n \ .$$

$$\Rightarrow \ x(t) - \sum_{j=1}^{n} c_j = r_n \ .$$

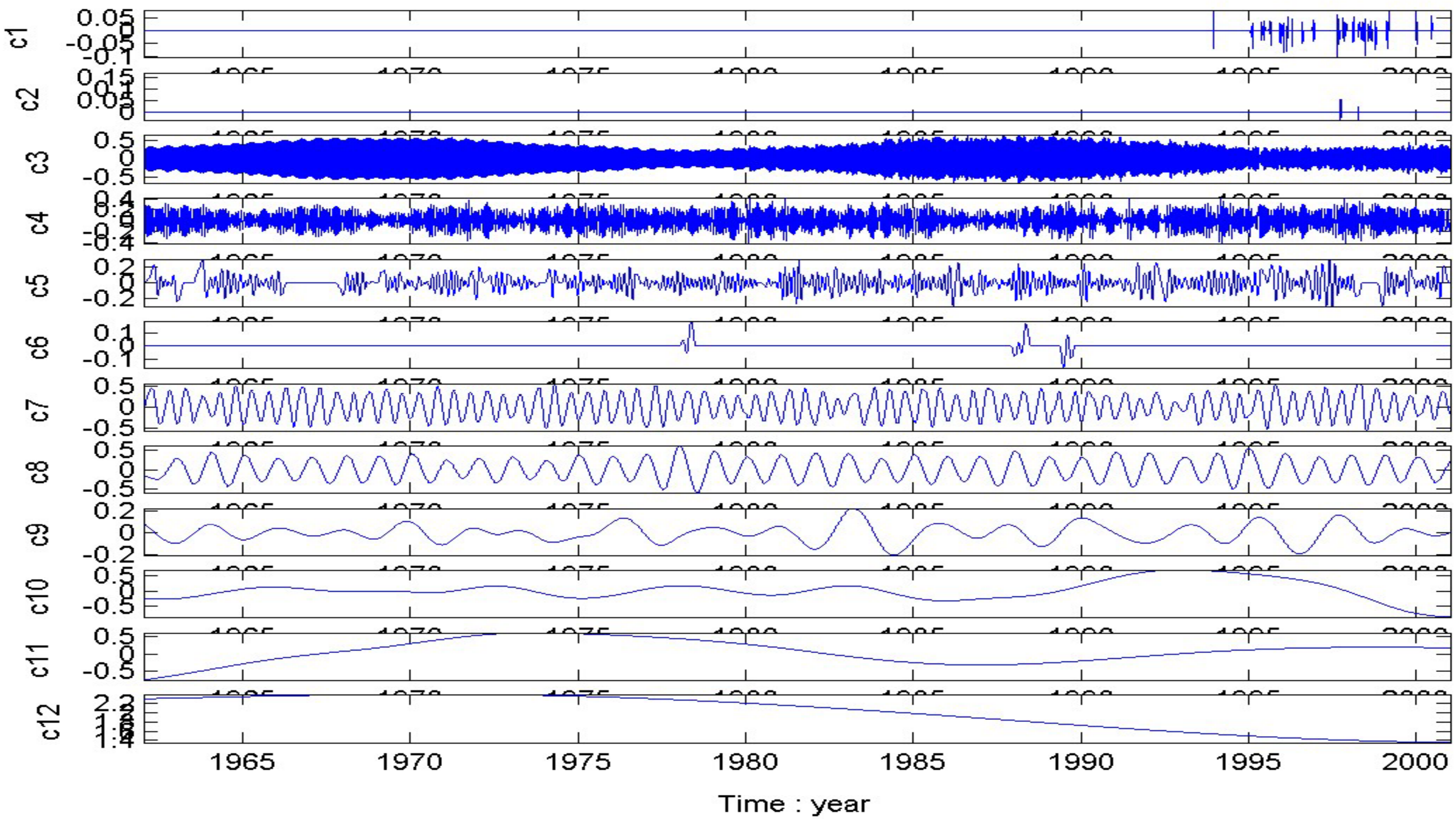
An Example of Sifting

Length Of Day Data

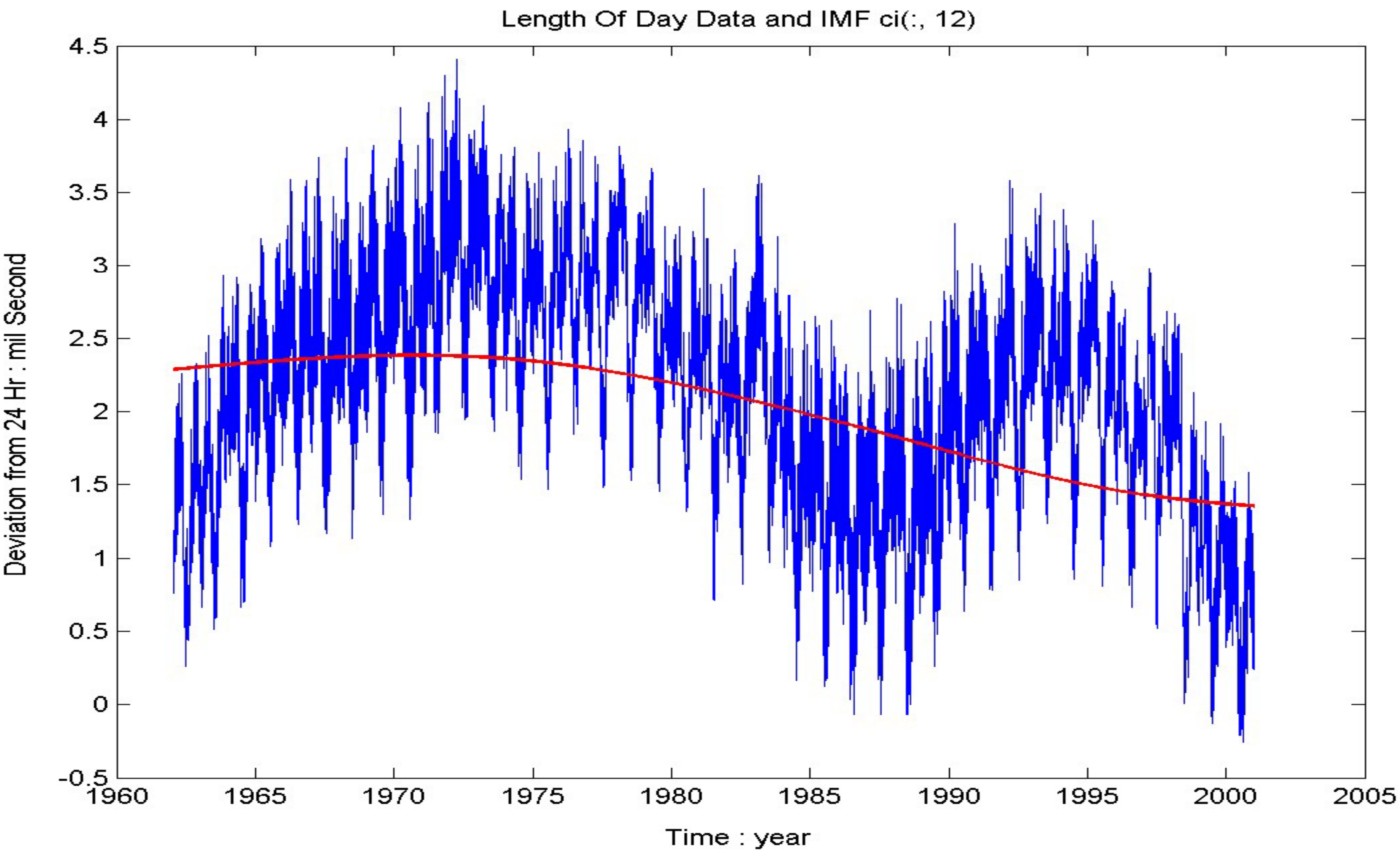


LOD： IMF

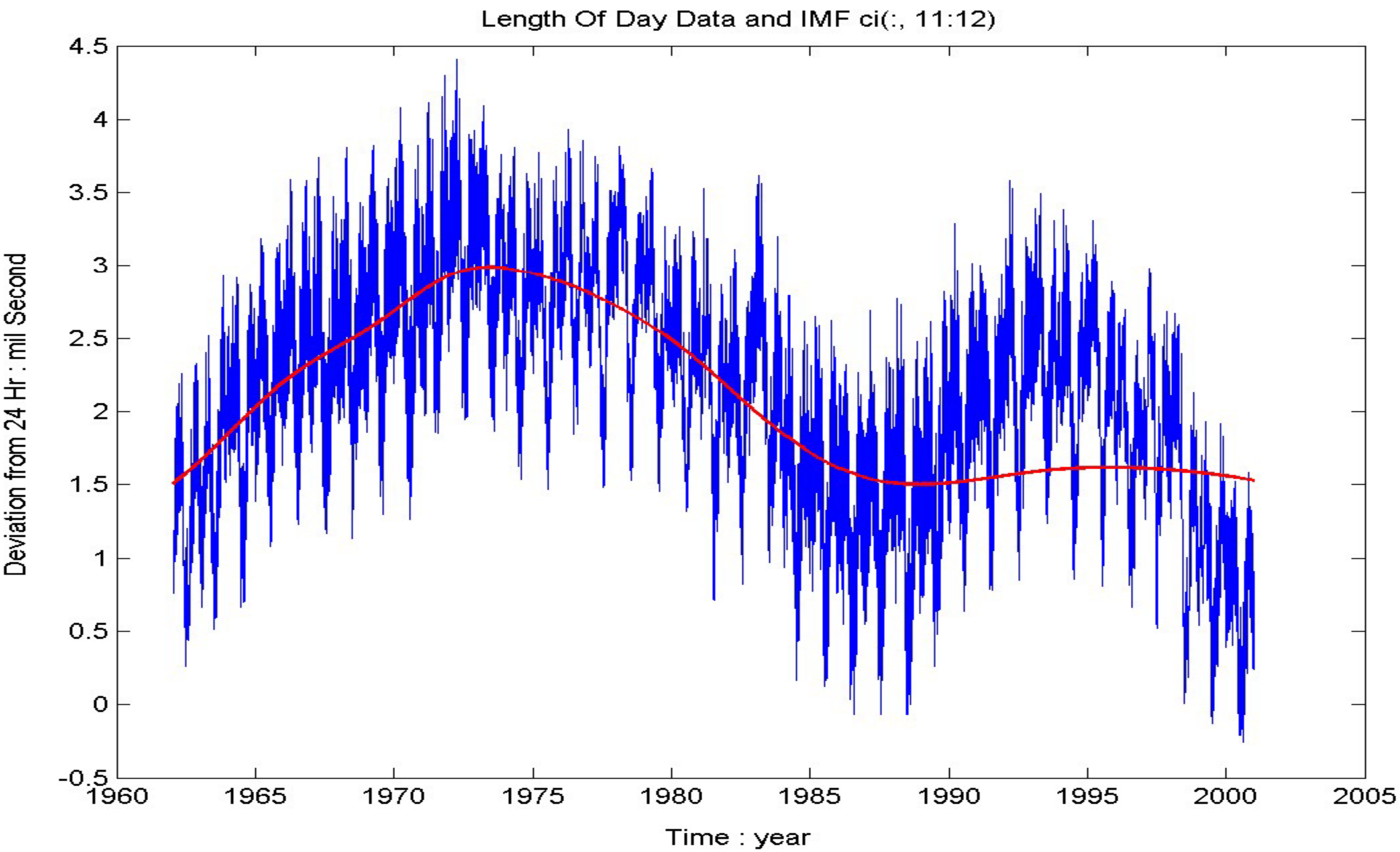
IMF LOD62 : ci(100,8,8; 3ª,: 50,3,3;-1²,45ª, -10)



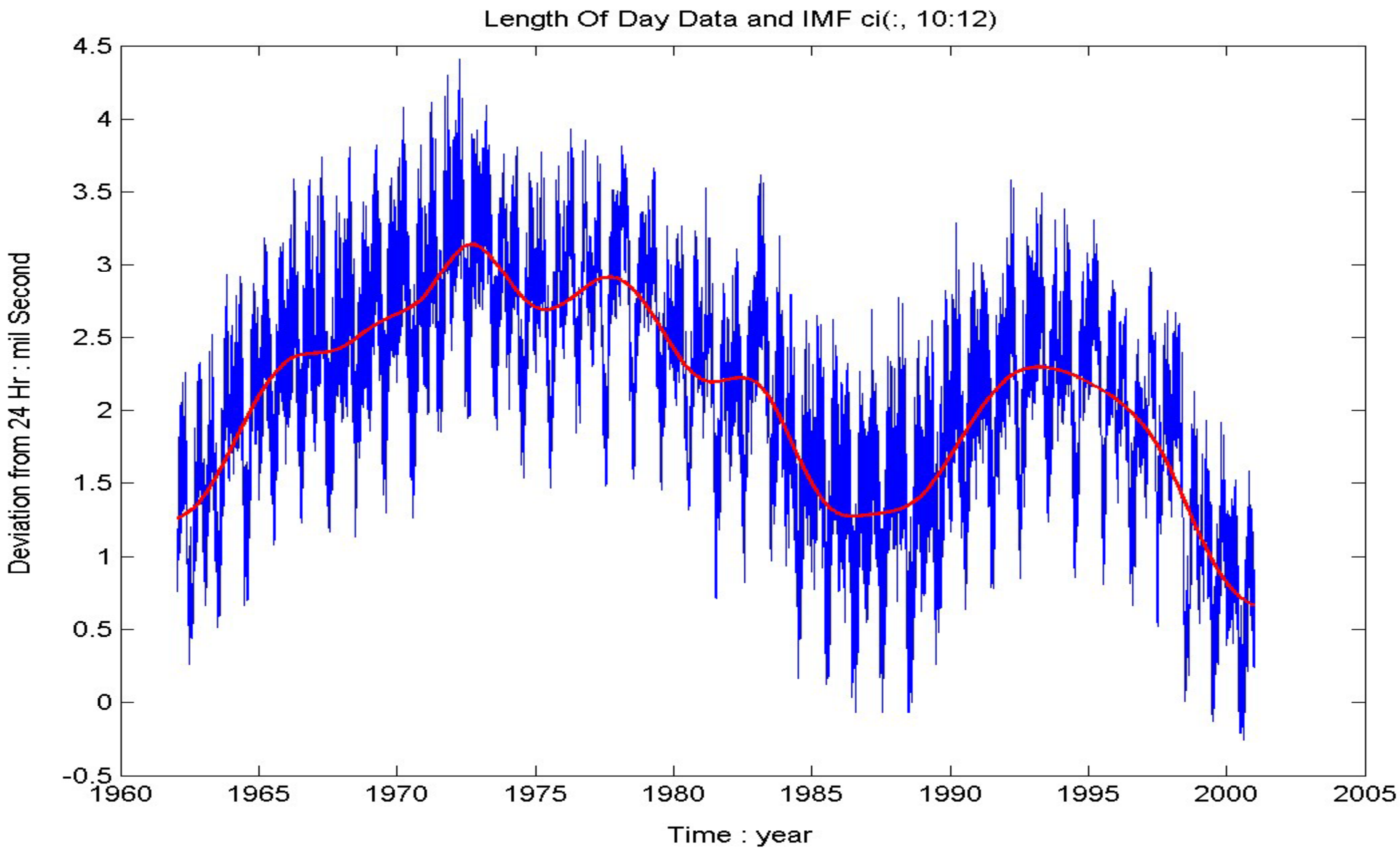
LOD : Data & c12


Length Of Day Data and IMF ci(:, 12)
Deviation from 24 Hr : mil Second
Time : year
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
-0.5
1960
1965
1970
1975
1980
1985
1990
1995
2000
2005

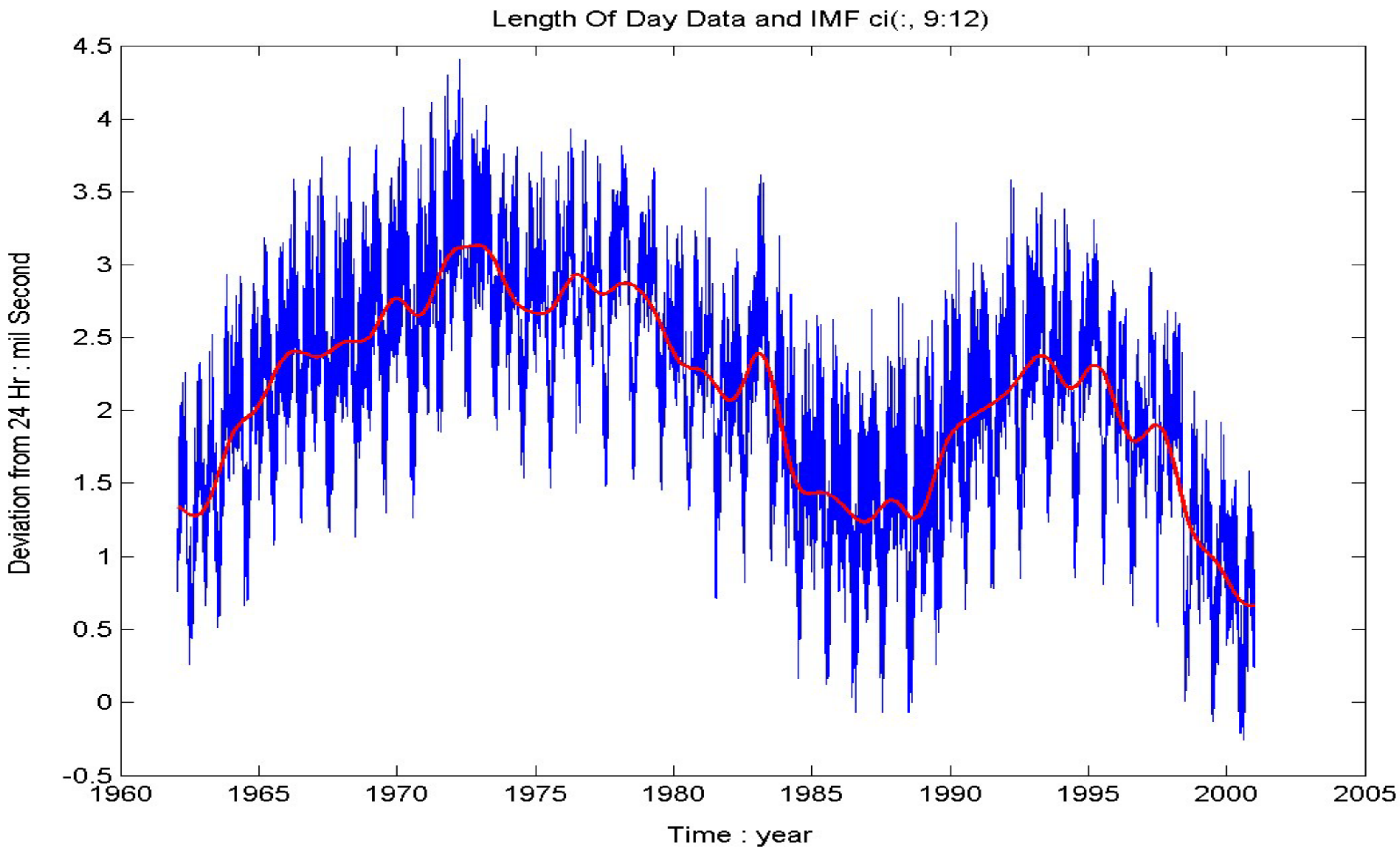
LOD : Data & Sum c11-12



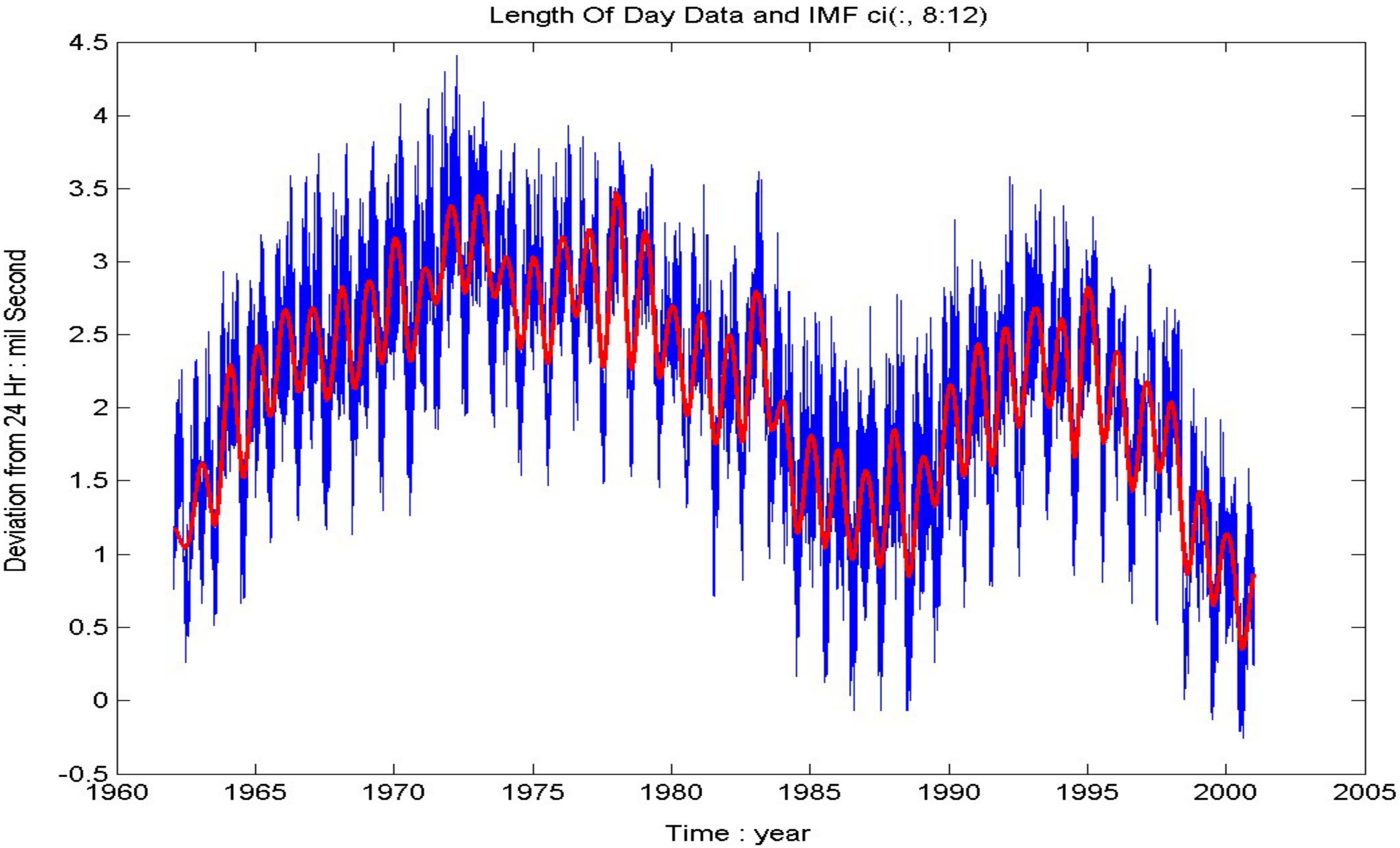
LOD : Data & sum c10-12



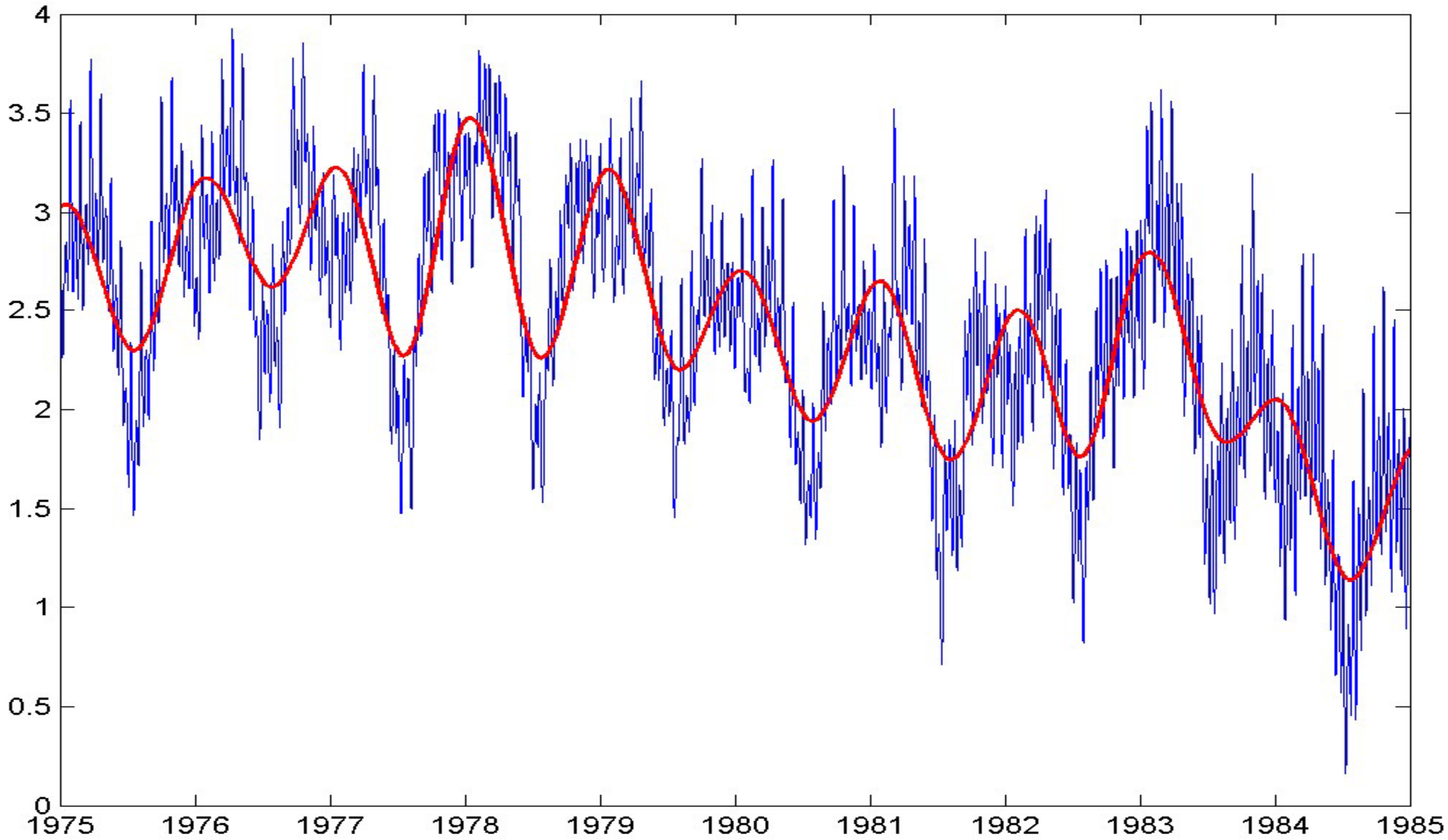
LOD : Data & c9 - 12



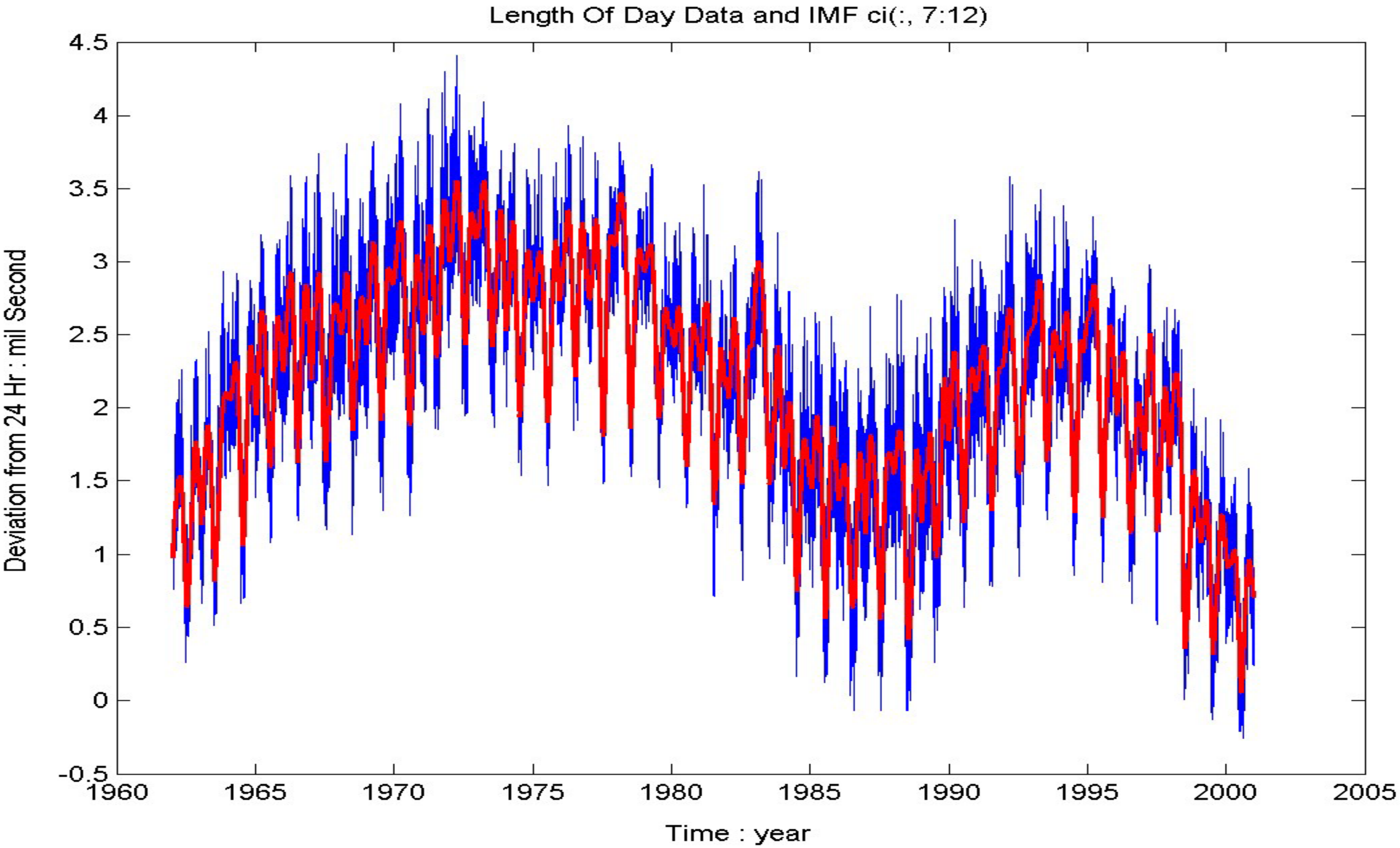
LOD : Data & c8 - 12



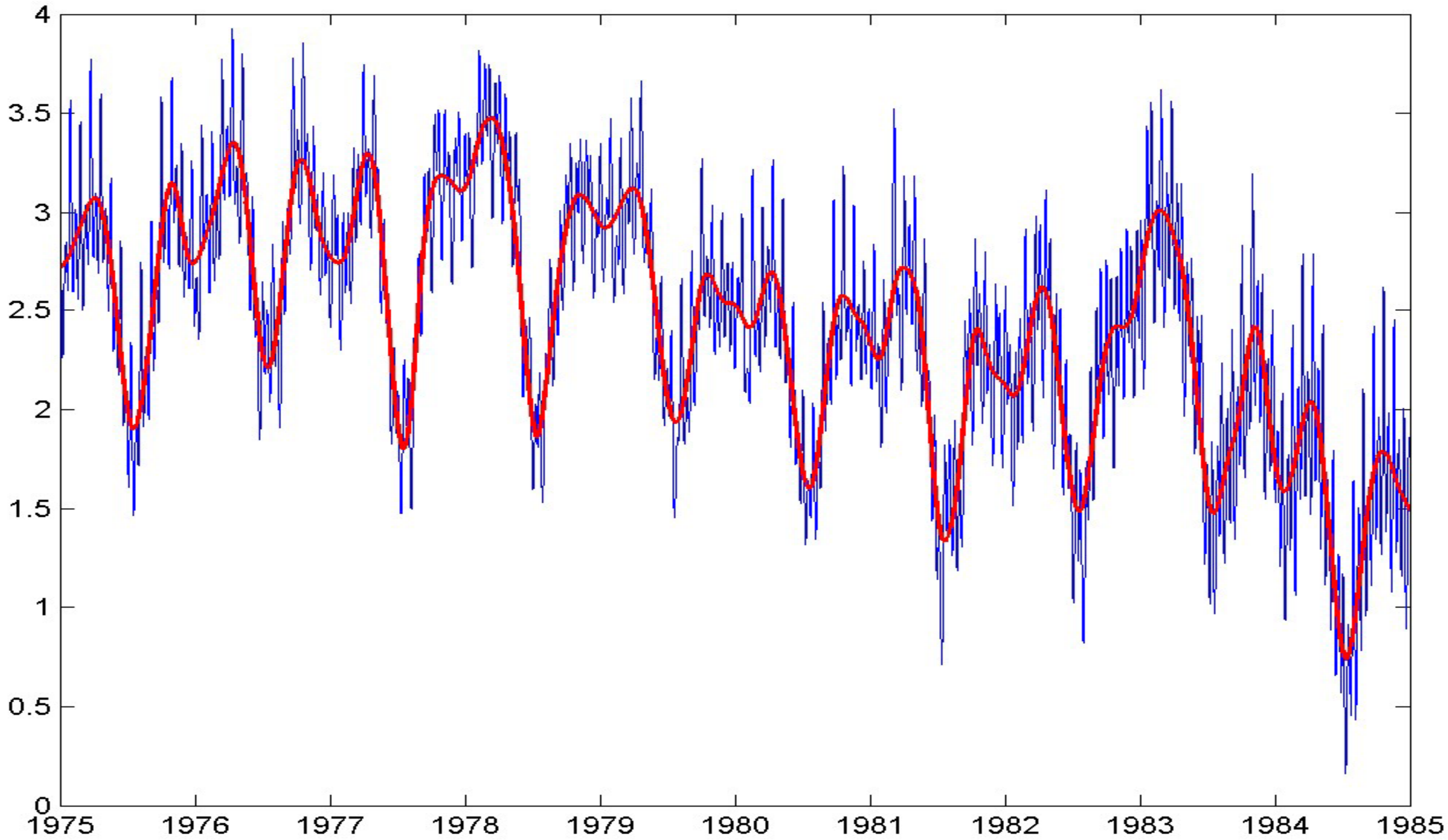
LOD : Detailed Data and Sum c8-c12



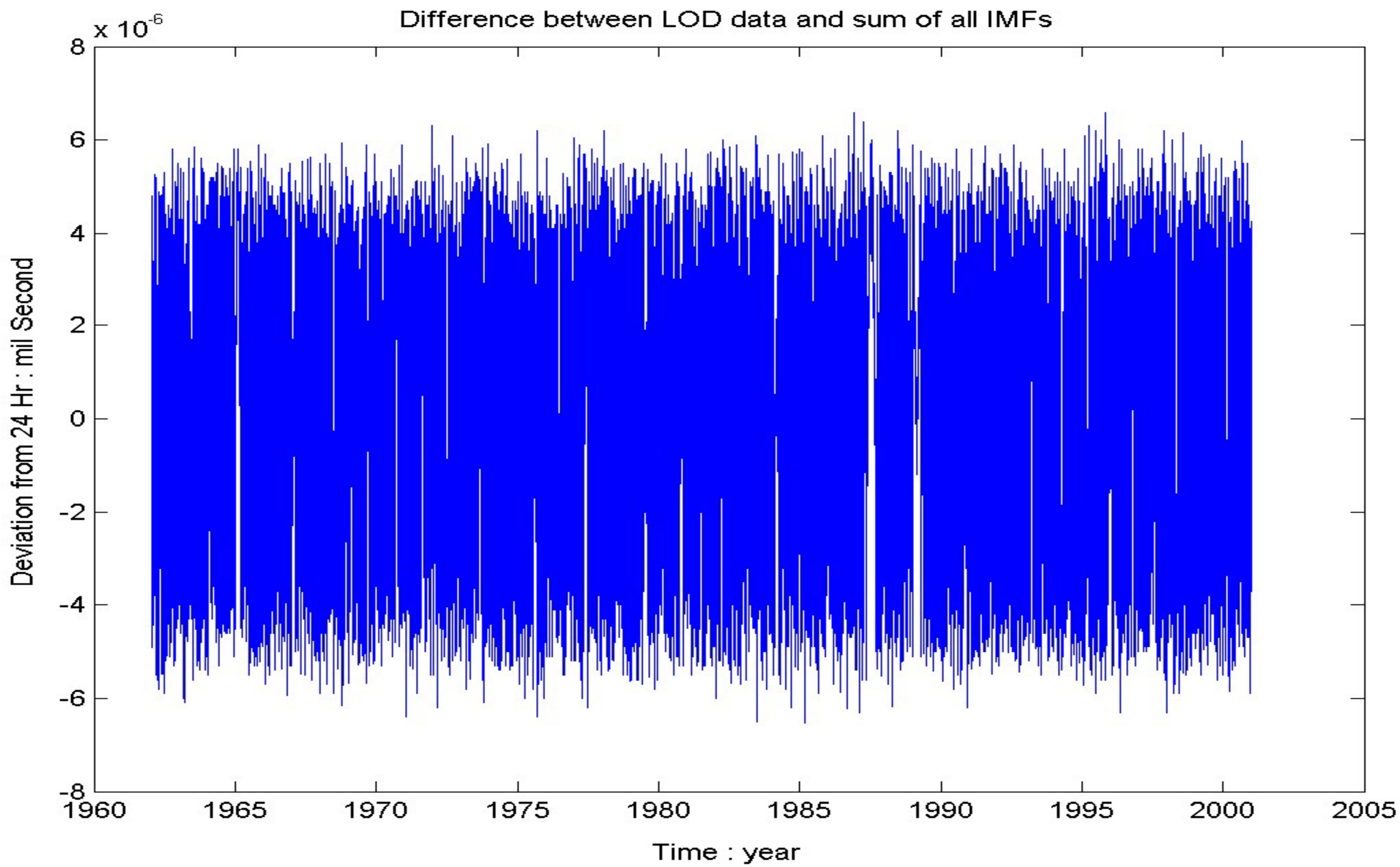
LOD : Data & c7 - 12



LOD : Detail Data and Sum IMF c7-c12



LOD : Difference Data − sum all IMFs



Definition of Instantaneous Frequency

The Fourier Transform of the Instrinsic Mode Funnction, c(t), gives

$$W(\omega) = \int_t a(t)\, e^{i(\theta - \omega t)}\, dt$$

By Stationary phase approximation we have

$$\frac{d\theta(t)}{dt} = \omega \ ,$$

This is defined as the Instantaneous Frequency.

The combination of Hilbert Spectral Analysis and Empirical Mode Decomposition has been

designated by NASA as

HHT

(HHT vs. FFT)

Jean-Baptiste-Joseph Fourier

1807 ***"On the Propagation of Heat in Solid Bodies"***

1812 ***Grand Prize of Paris Institute***

"Théorie analytique de la chaleur"

'... the manner in which the author arrives at these equations is not exempt of difficulties and that his analysis to integrate them still leaves something to be desired on the score of generality and even rigor.'

1817 ***Elected to Académie des Sciences***

1822 ***Appointed as Secretary of Math Section***

paper published

Fourier's work is a great mathematical poem.

Lord Kelvin

Comparison between FFT and HHT

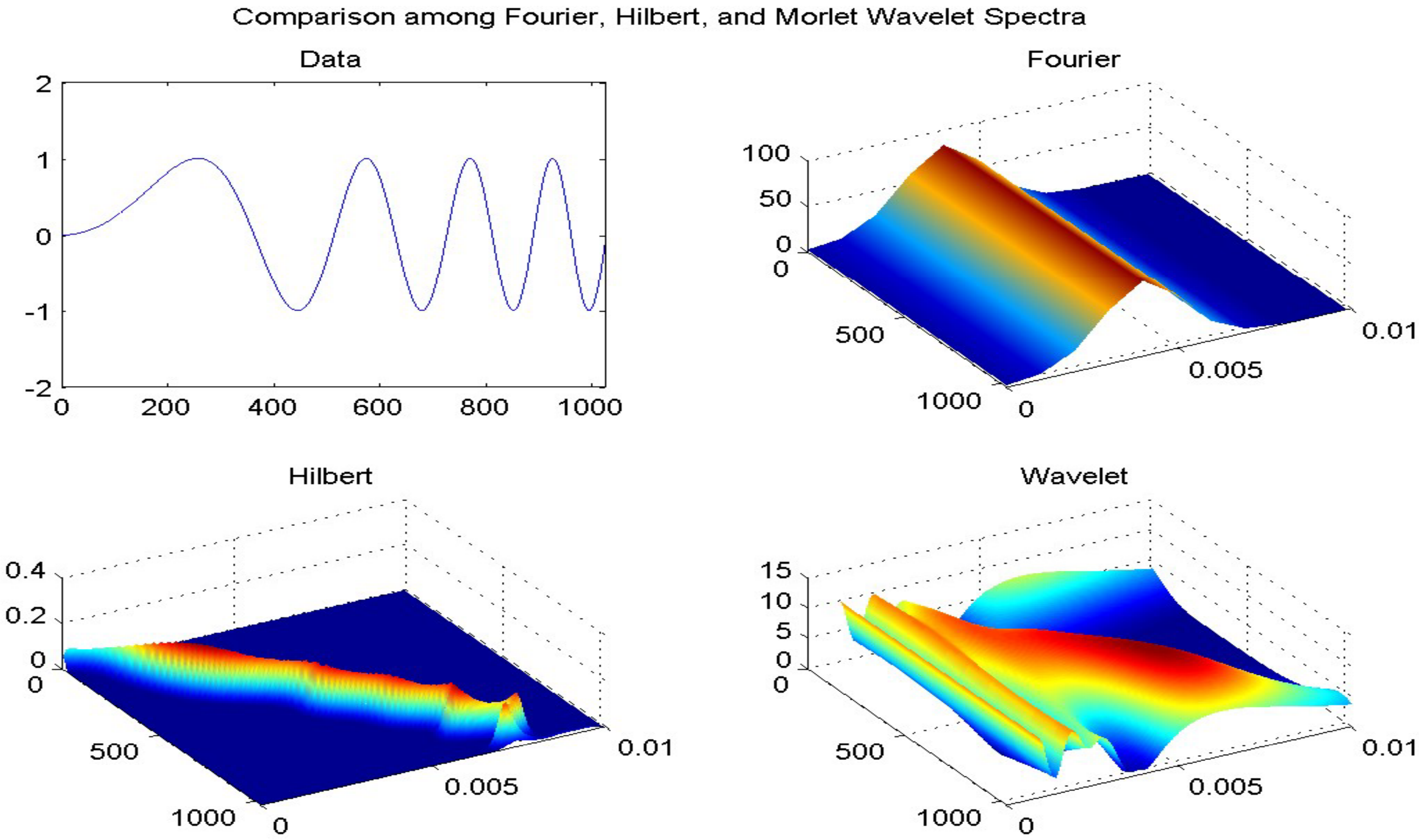
1. FFT :

$$x(t) = \Re \sum_j a_j e^{i\omega_j t} .$$

2. HHT :

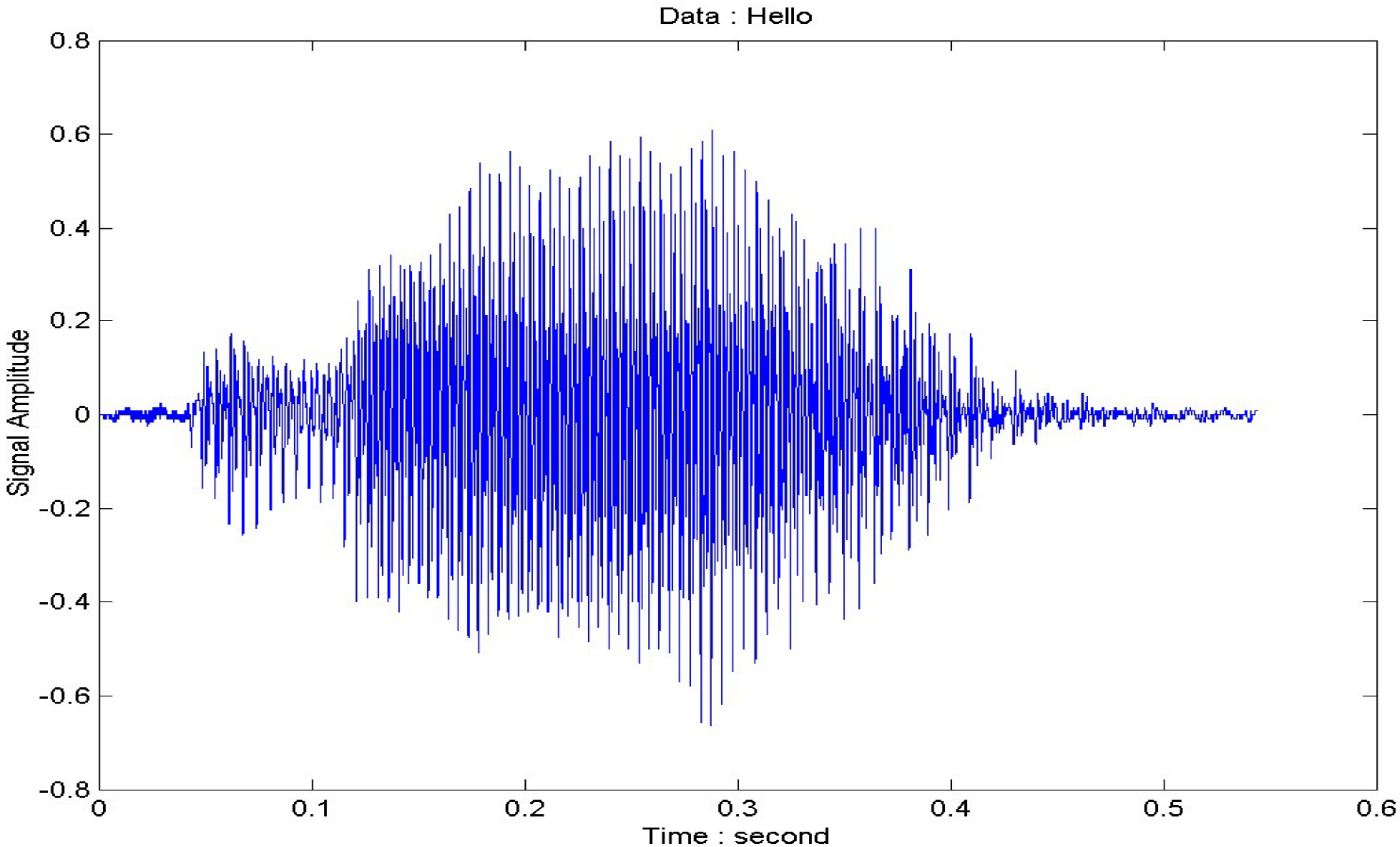
$$x(t) = \Re \sum_j a_j(t) e^{i\int_t \omega_j(\tau)d\tau} .$$

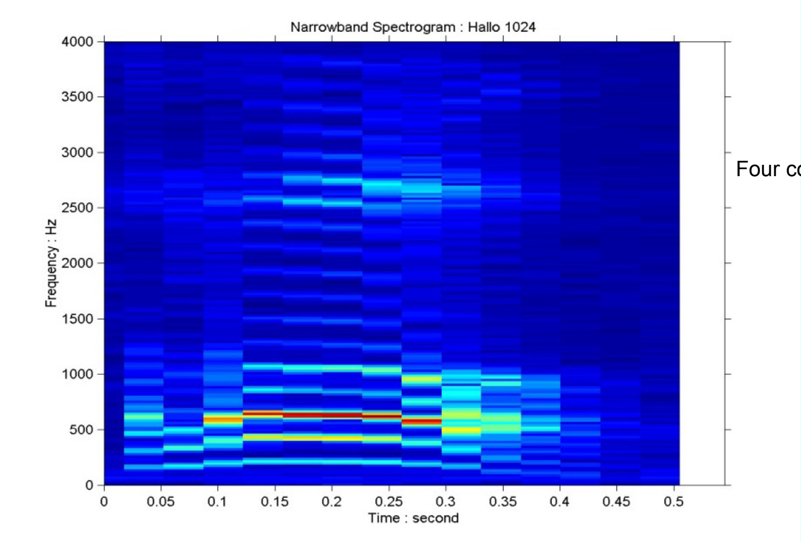
Comparisons: Fourier, Hilbert & Wavelet



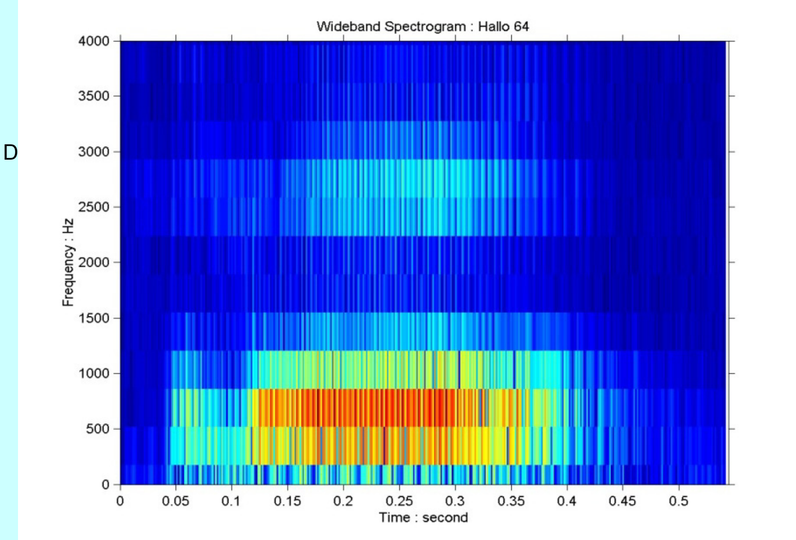
Speech Analysis

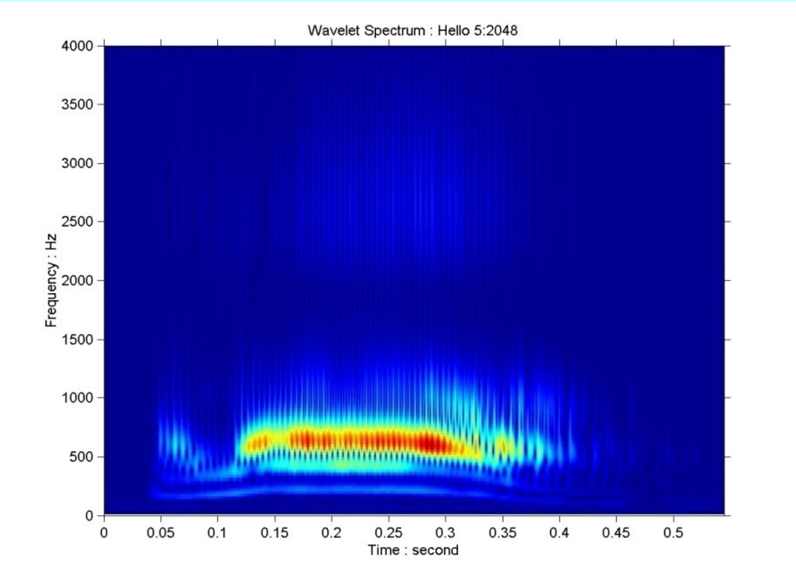
Hello : Data

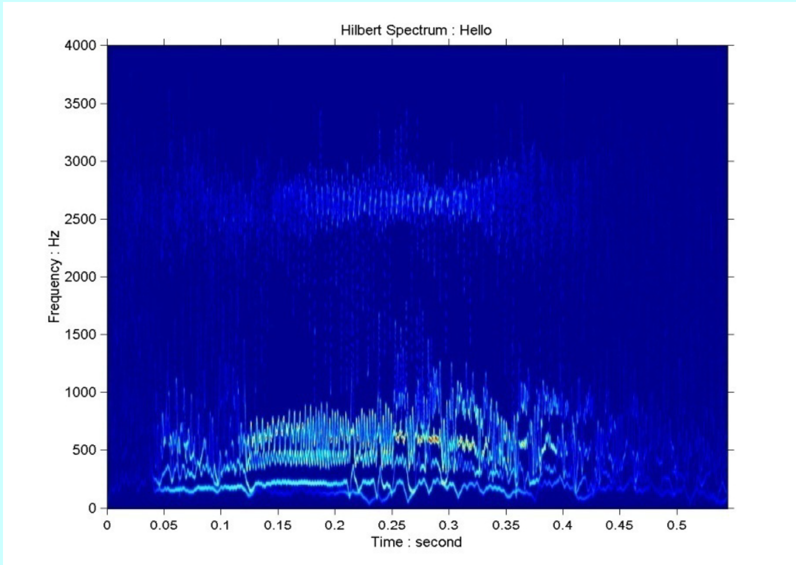




Four comparsions D







Comparisons

	Fourier	Wavelet	Hilbert
Basis	a priori	a priori	Adaptive
Frequency	Integral transform: Global	Integral transform: Regional	Differentiation: Local
Presentation	Energy-frequency	Energy-time-frequency	Energy-time-frequency
Nonlinear	no	no	yes
Non-stationary	no	yes	yes
Uncertainty	yes	yes	no
Harmonics	yes	yes	no

Recent Advances in Hilbert-Huang Transform and Its Applications

Norden E. Huang
Research Center for Adaptive Data Analysis
National Central University

Academia Sinica
2010/10/13

History of HHT

1998: The Empirical Mode Decomposition Method and the Hilbert Spectrum for Non-stationary Time Series Analysis, Proc. Roy. Soc. London, A454, 903-995. The invention of the basic method of EMD, and Hilbert transform for determining the Instantaneous Frequency and energy.

1999: A New View of Nonlinear Water Waves – The Hilbert Spectrum, Ann. Rev. Fluid Mech. 31, 417-457.

Introduction of the intermittence in decomposition.

2003: A confidence Limit for the Empirical mode decomposition and the Hilbert spectral analysis, Proc. of Roy. Soc. London, A459, 2317-2345.

Establishment of a confidence limit without the ergodic assumption.

2004: A Study of the Characteristics of White Noise Using the Empirical Mode Decomposition Method, Proc. Roy. Soc. London, A460, 1597-1611

Defined statistical significance and predictability.

2007: On the trend, detrending, and variability of nonlinear and nonstationary time series. Proc. Natl. Acad. Sci., 104, 14,889-14,894.

The correct adaptive trend determination method

2009: On Ensemble Empirical Mode Decomposition. Advances in Adaptive Data Analysis, 1, 1-41, 2009.

2009: On instantaneous Frequency. Advances in Adaptive Data Analysis 1, 177-229, 2009.

………………………..

new Journal dedicated the new data analysis proaches

ADVANCES IN ADAPTIVE DATA ANALYSIS

Theory and Applications

ADVANCES IN ADAPTIVE DATA ANALYSIS

Volume 1 · Number 1 · January 20

Editors-in-Chief

Norden E. Huang
National Central University, Taiwan

Thomas Yizhao Hou
California Institute of Technology, USA

Vol. 1 No. 1 (January 2009) P

Some Recent Advances

- Ensemble EMD (EEMD)
- Instantaneous Frequency (Quadrature)
- Quantification of Hilbert Spectrum
- Multi-Dimensional EEMD (MDEEMD)
- Time Dependent Intrinsic Correlation (TDIC)

Some Recent Advances

- Ensemble EMD (EEMD)
- EEMD Instantaneous Frequency (Quadrature)
- Quantification of Hilbert Spectrum
- Multi-Dimensional (MDEEMD)
- Time Dependent Intrinsic Correlation (TDIC)

Advances in Adaptive Data Analysis
Vol. 1, No. 1 (2009) 1–41

World Scientific
www.worldscientific.com

ENSEMBLE EMPIRICAL MODE DECOMPOSITION: A NOISE-ASSISTED DATA ANALYSIS METHOD

ZHAOHUA WU* and NORDEN E. HUANG†

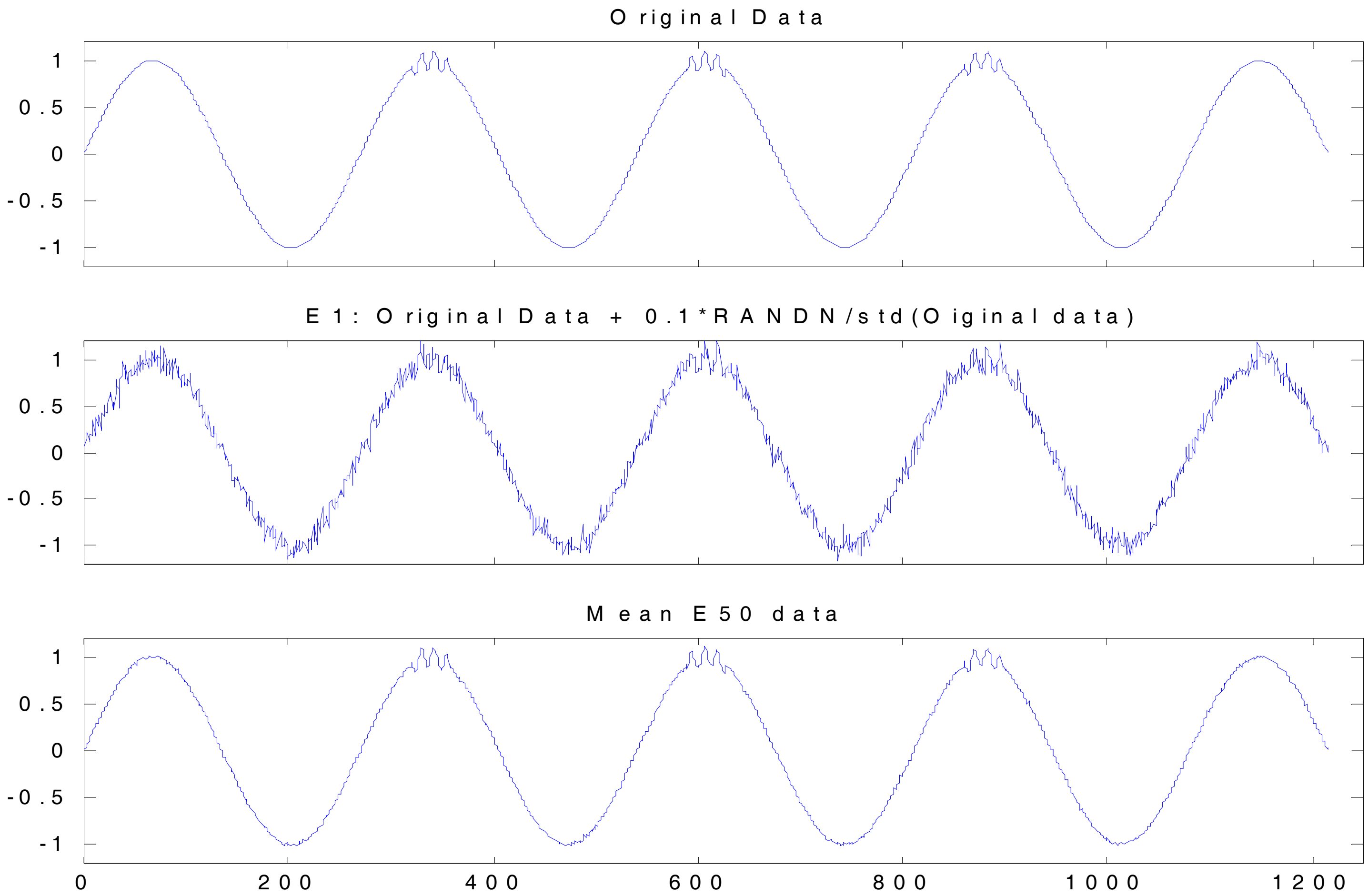
*Center for Ocean–Land–Atmosphere Studies
4041 Powder Mill Road, Suite 302
Calverton, MD 20705, USA*

†*Research Center for Adaptive Data Analysis
National Central University
300 Jhongda Road, Chungli, Taiwan 32001*

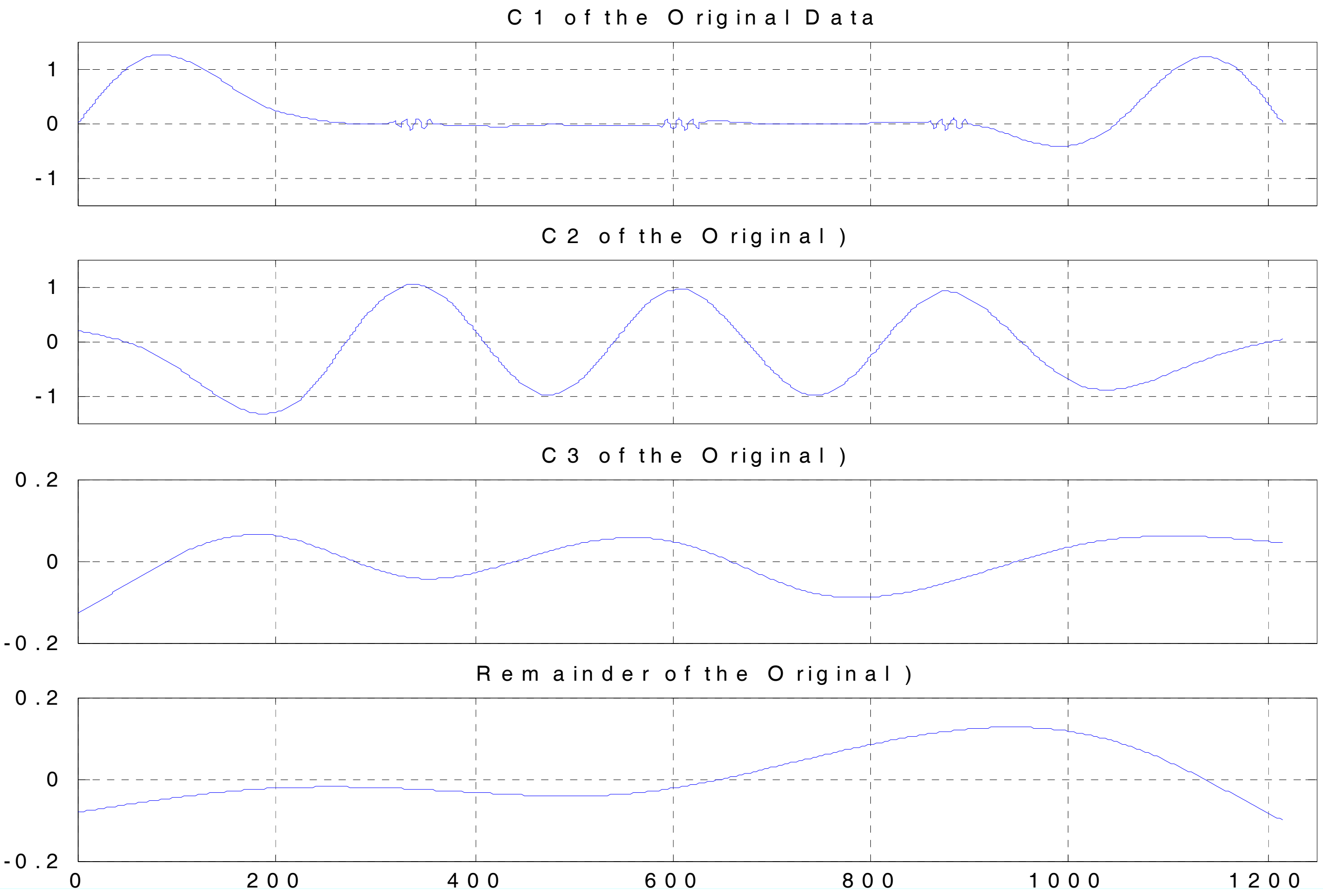
Procedures for EEMD

- Add a white noise series to the targeted data;
- Decompose the data with added white noise into IMFs;
- Repeat step 1 and step 2 again and again, but with different white noise series each time; and
- Obtain the (ensemble) means of corresponding IMFs of the decompositions as the final result.

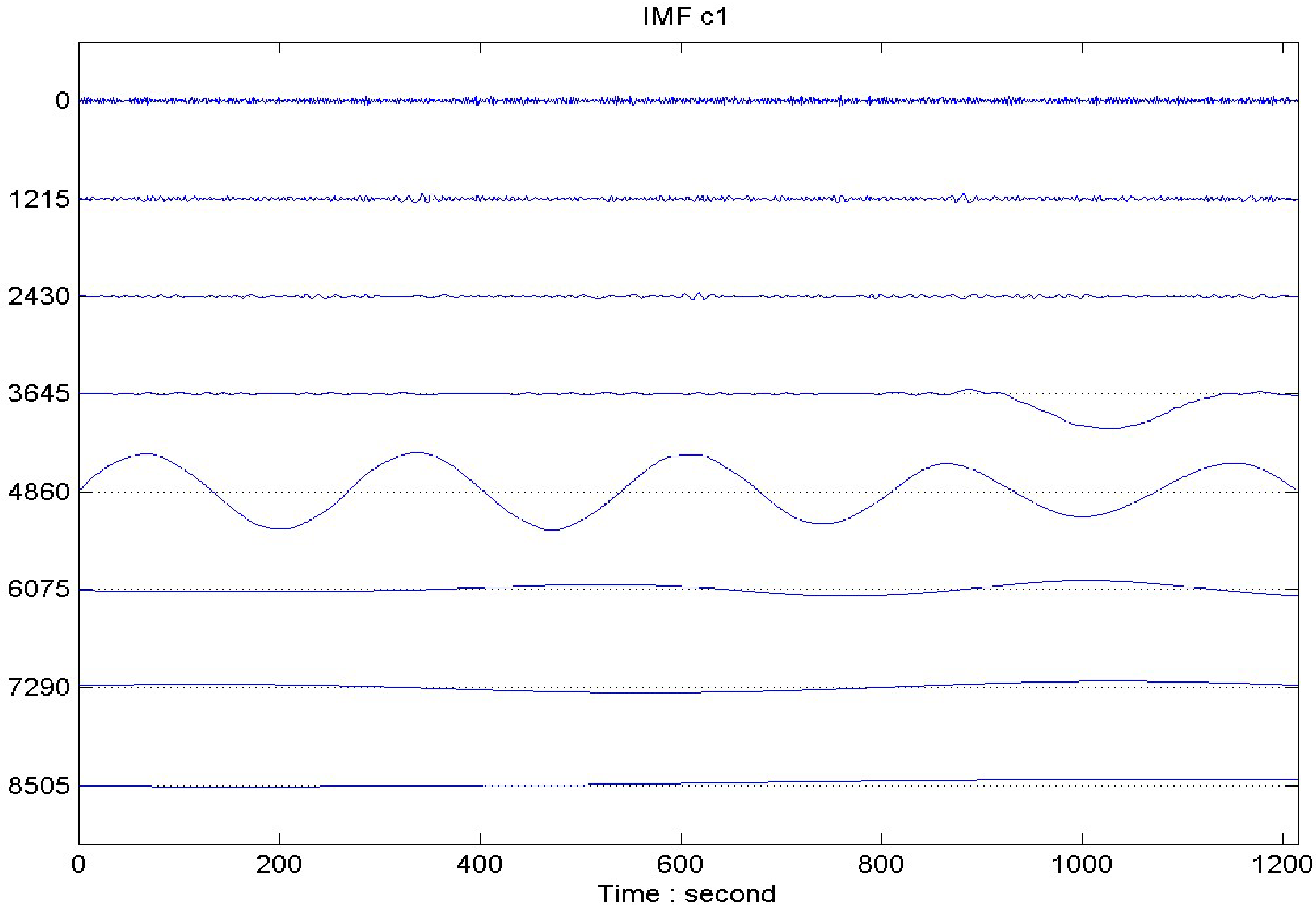
EXAMPLE : ORIGINAL DATA



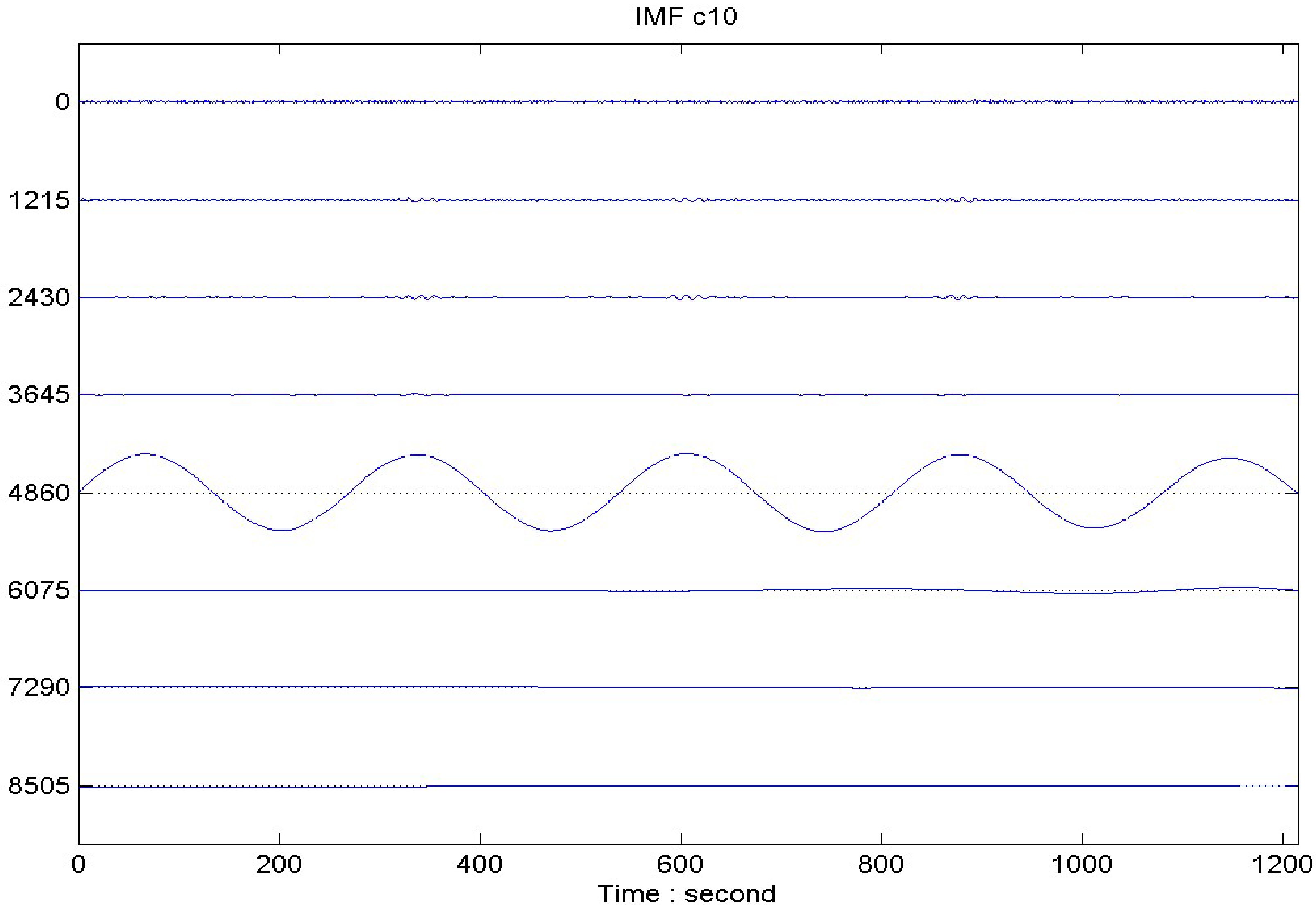
EXAMPLE : ORIGINAL DECOMP.



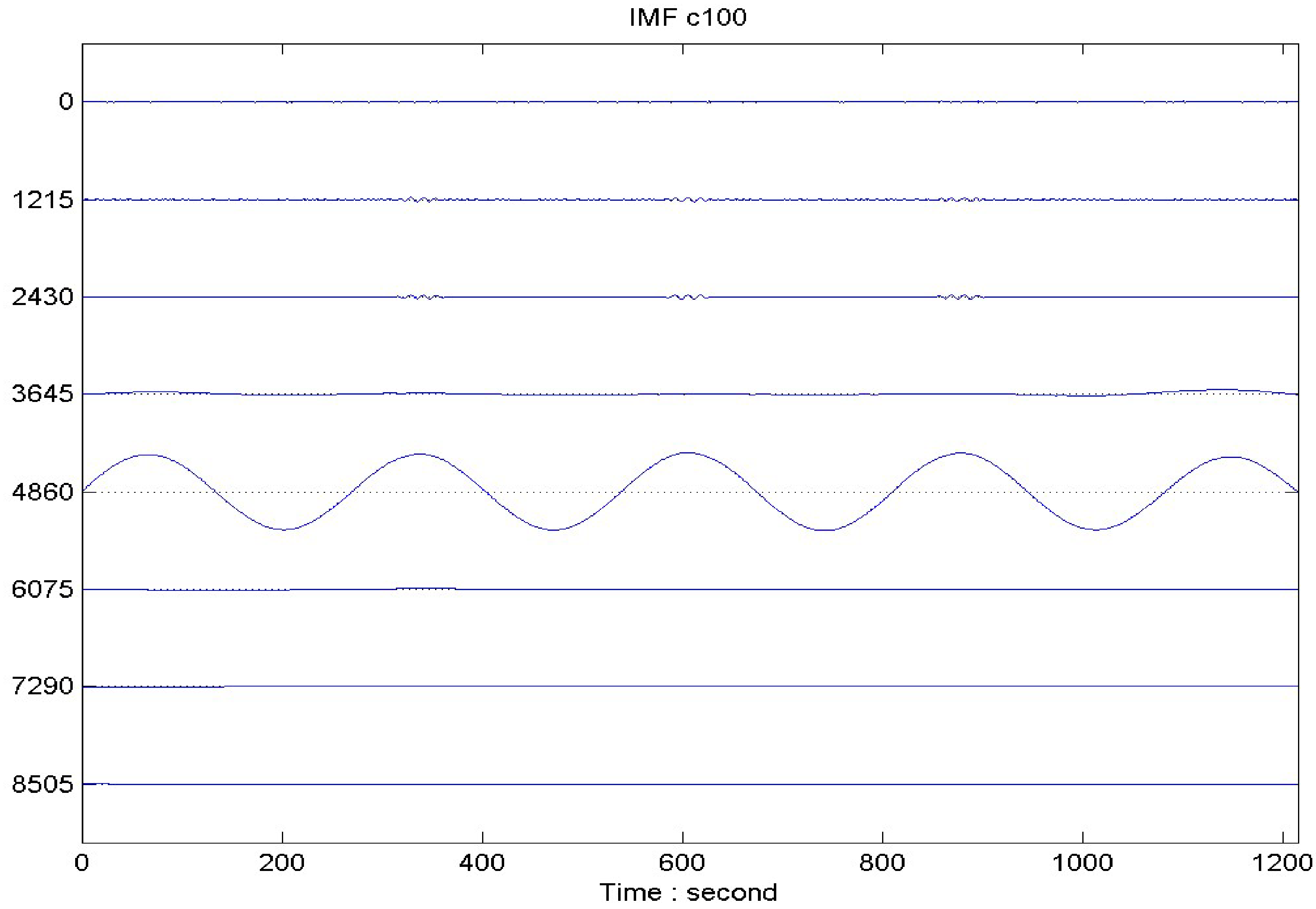
EXAMPLE : E1 DECOMP.



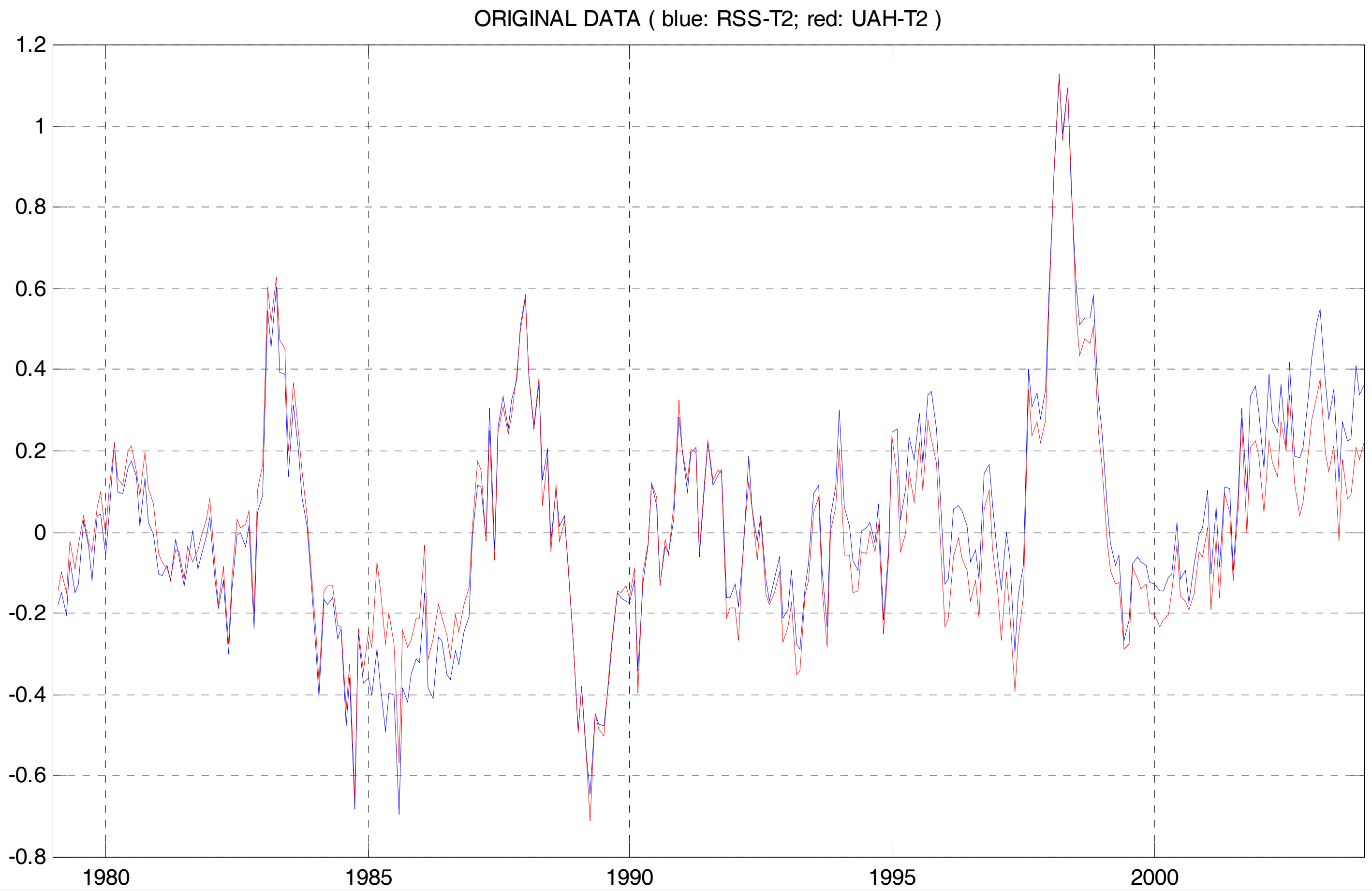
EXAMPLE : E10 DECOMP.



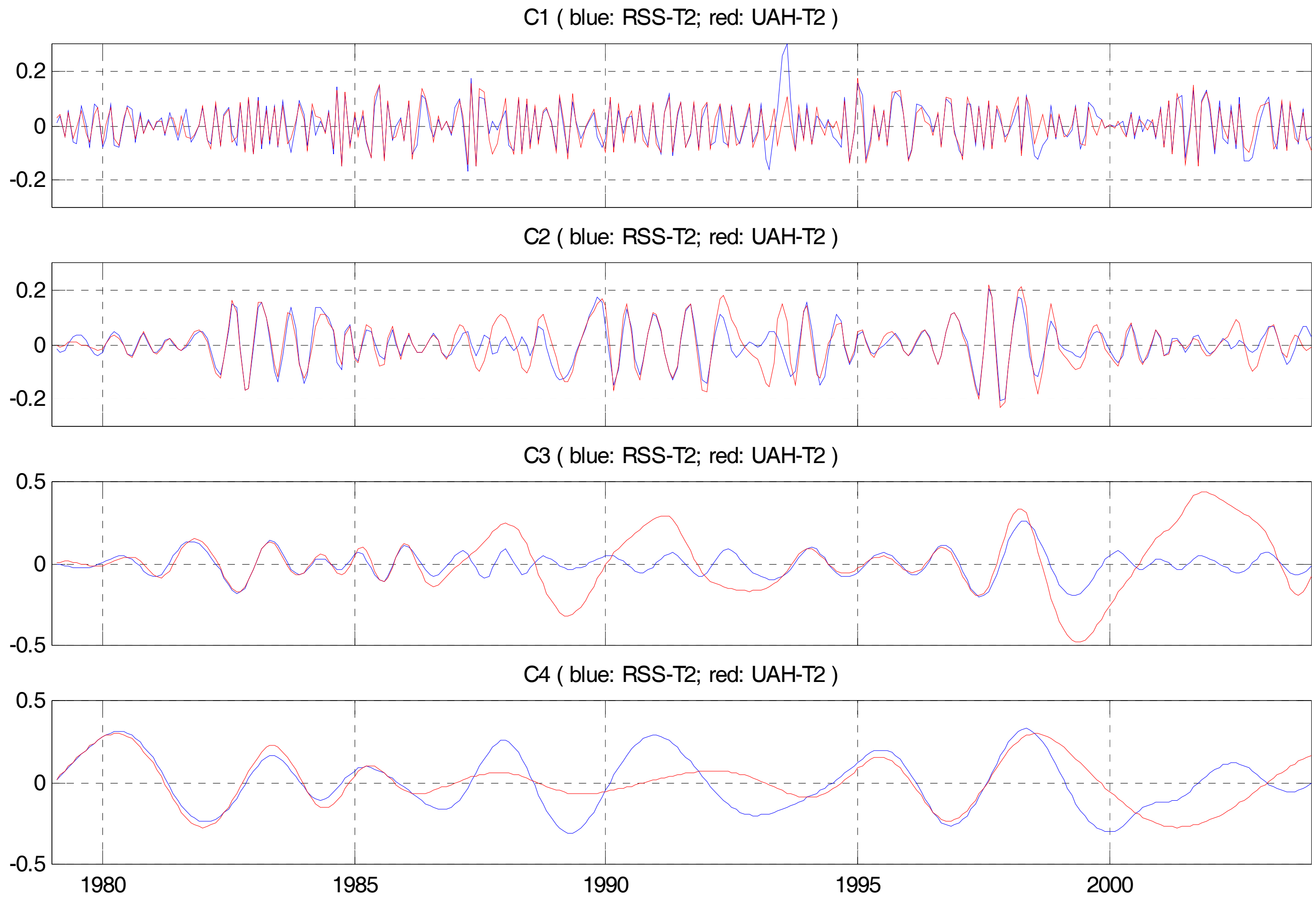
EXAMPLE : E100 DECOMP.



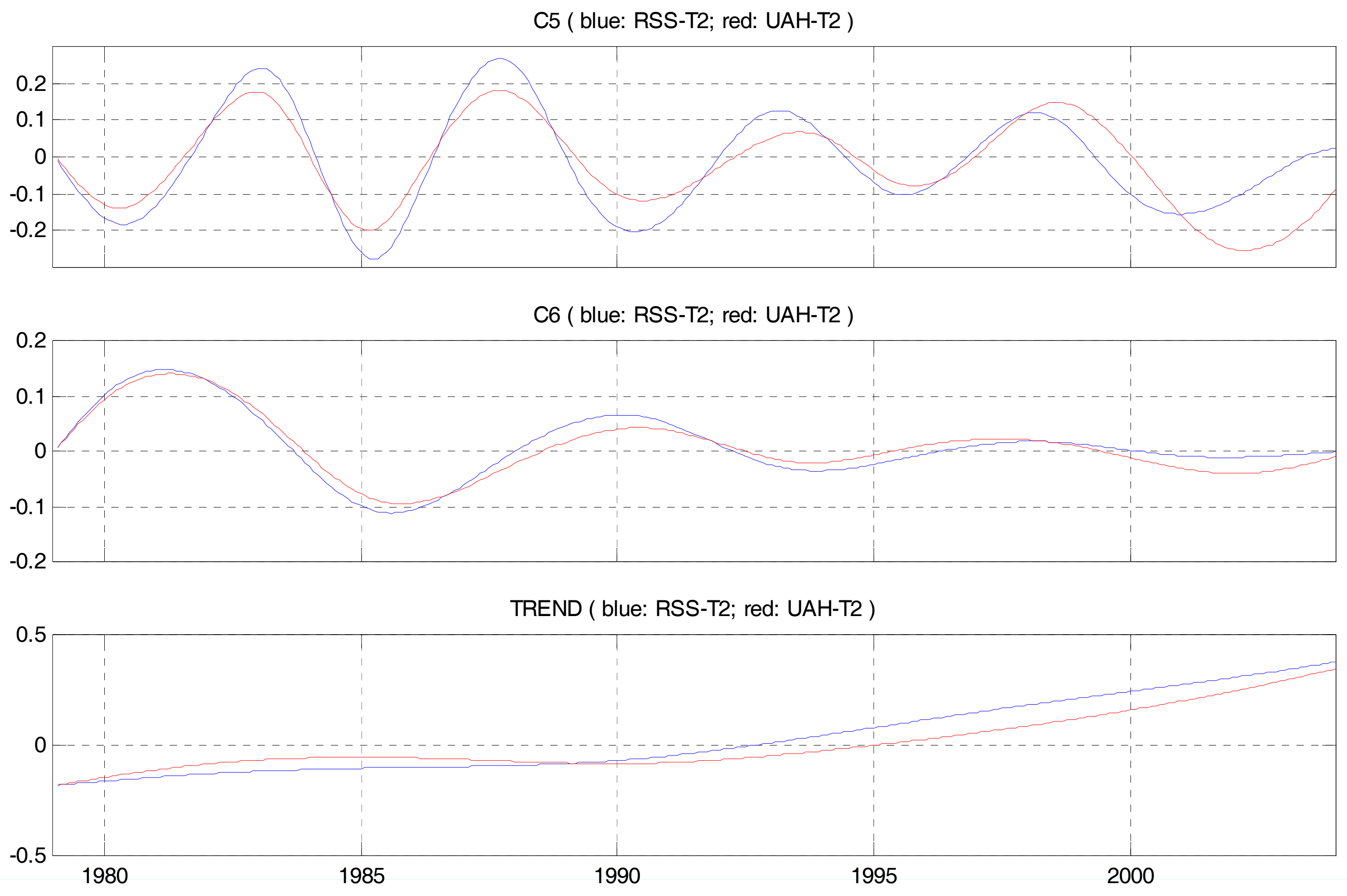
EXAMPLE I: ORIGINAL DATA



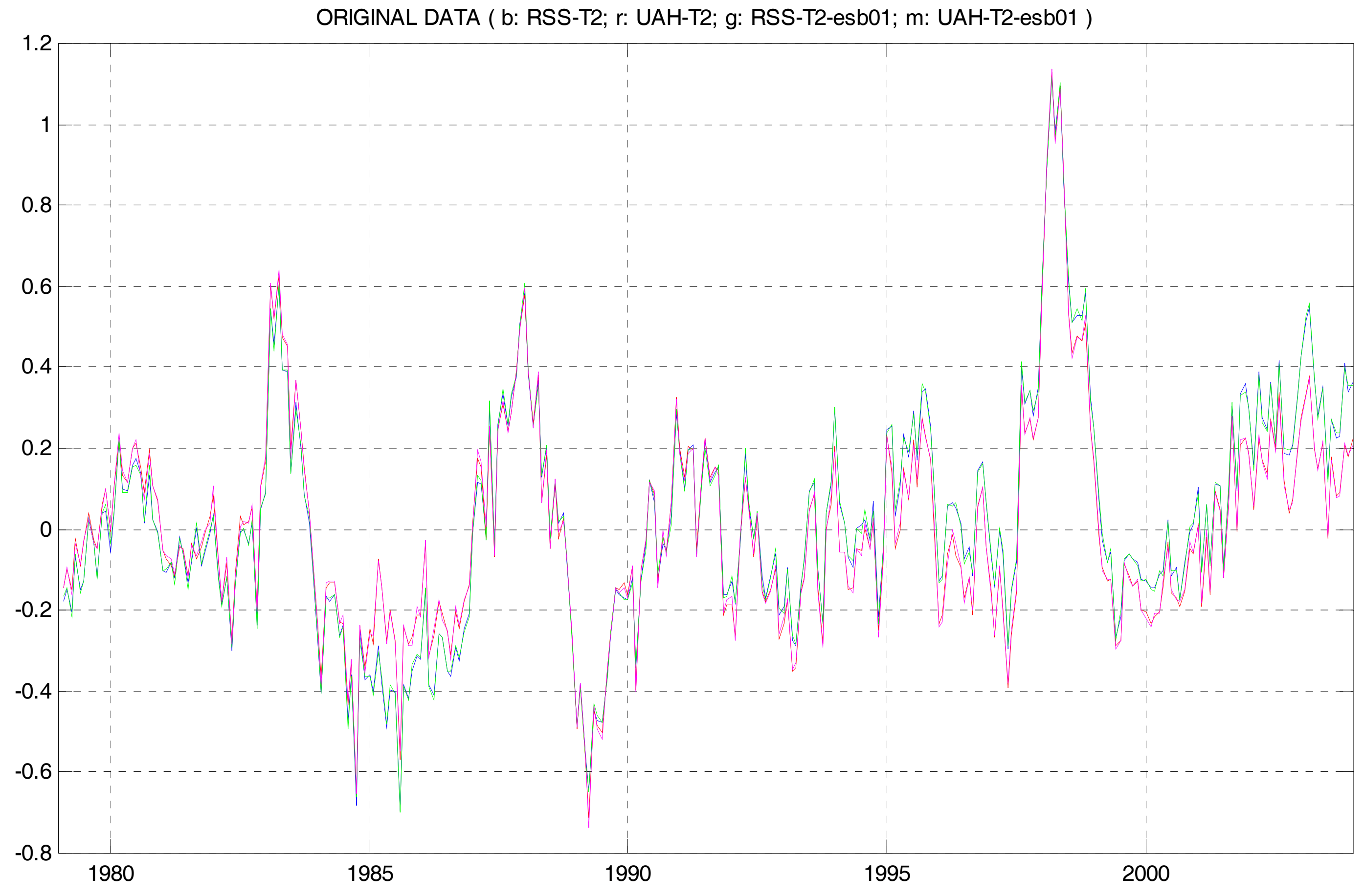
EXAMPLE I: DECOMPOSITION (I)



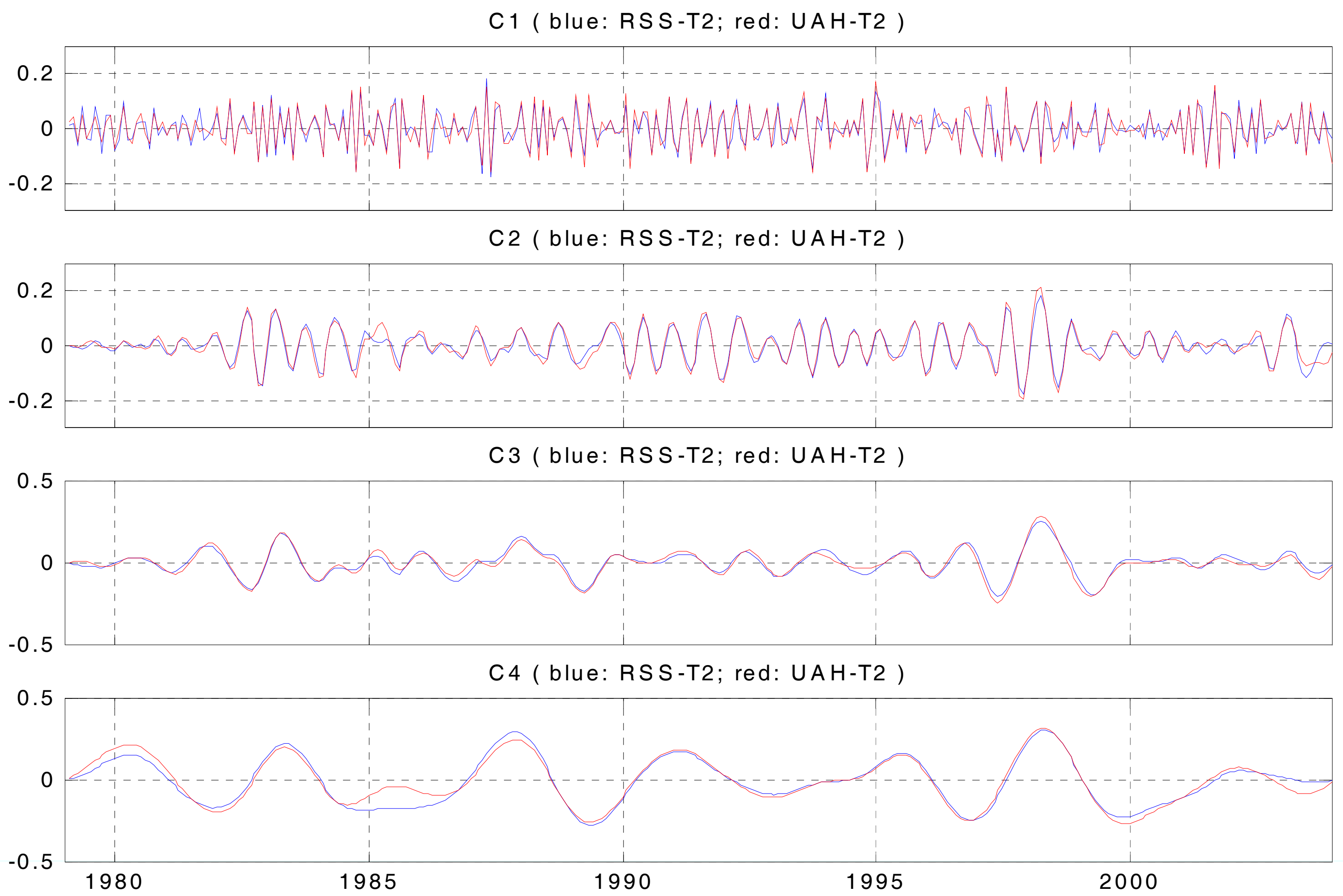
EXAMPLE I: DECOMPOSITION (II)



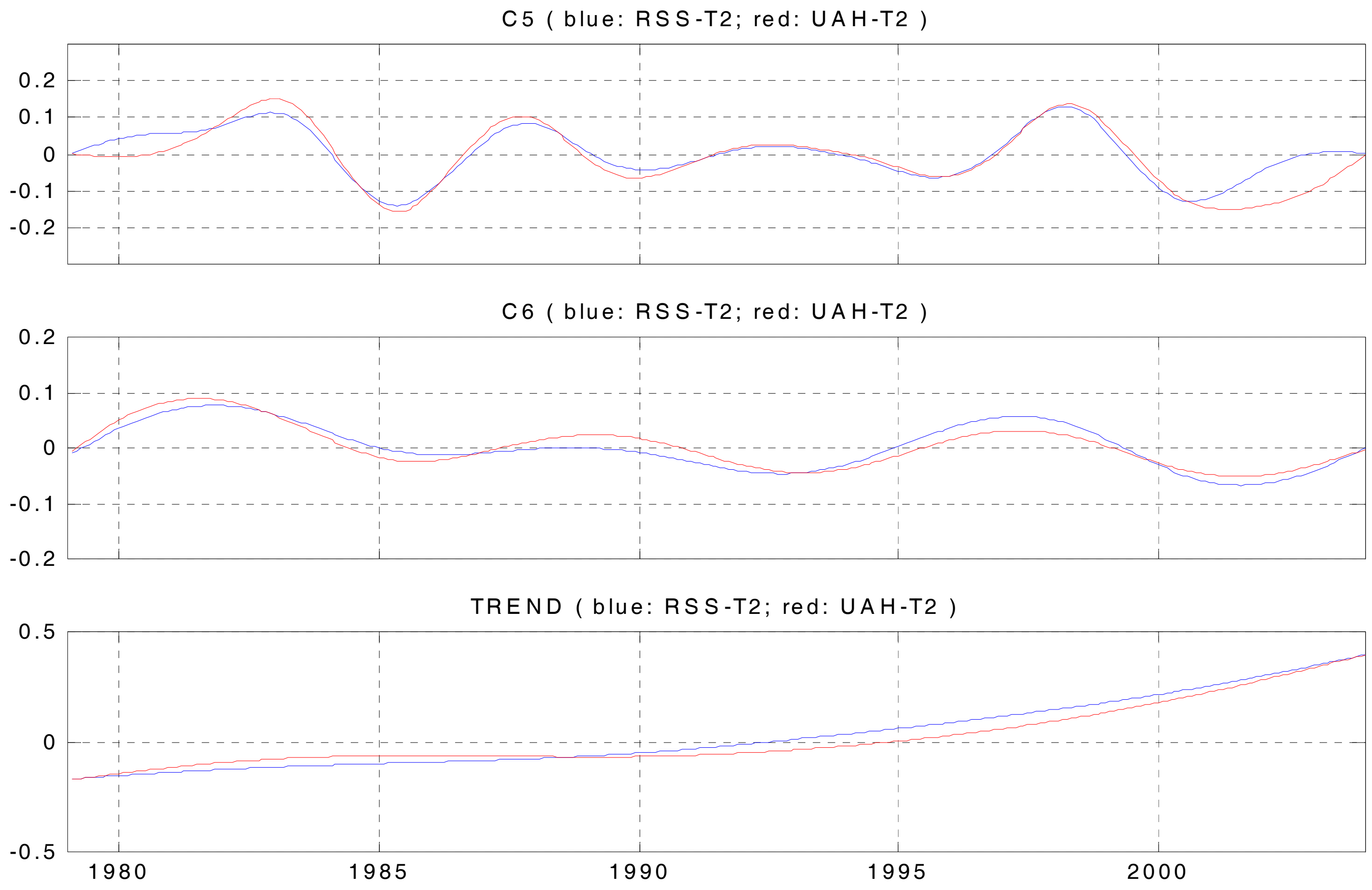
EXAMPLE I: NOISY DATA
(added noise std=0.1)



NOISY DATA DECOMPOSITION (I)
(added noise std=0.1)



NOISY DATA DECOMPOSITION (II)
(added noise std=0.1)



Definition of IMF in EEMD

The truth defined by EEMD is given by the number of the ensemble approaching infinite:

$$c_j(t) = \lim_{N \to \infty} \frac{1}{N} \sum_{k=1}^{N} \left\{ c_{j,k}(t) + \alpha \, r_k(t) \right\},$$

in which $c_{j,k}(t) + \alpha \, r_k(t)$ *is the k-th realization of the j-th IMF in the noise added signal, and* α *is the magnitude of the added noise that is not necessarily small.*

Summary

- True IMFs can be derived from adding finite amplitude of noise, rather than the case with infinitesimal noises.
- Ensemble EMD indeed enables the signals of similar scale collated together.
- No need for *a priori* criteria for intermittency.

Summary

- Less adaptive than EMD.

- As the components produced by EEMD are the averaged values of many IMFs, they might not be IMFs: some of the component might have multi-extrema. More stringent stoppage criteria and/or trials in the ensemble can improve the situation.

Some Recent Advances

- Ensemble EMD (EEMD)
- Instantaneous Frequency (Quadrature)
- Quantification of Hilbert Spectrum
- Multi-Dimensional EEMD (MDEEMD)
- Time Dependent Intrinsic Correlation (TDIC)

Advances in Adaptive Data Analysis
Vol. 1, No. 2 (2009) 177–229

ON INSTANTANEOUS FREQUENCY

NORDEN E. HUANG

Research Center for Adaptive Data Analysis
National Central University
Chungli, Taiwan 32001, Republic of China
Norden@ncu.edu.tw

ZHAOHUA WU

Department of Meteorology & Center for
Ocean-Atmospheric Prediction Studies
Florida State University
Tallahassee, FL 32306, USA

Bedrosian Theorem

Let f(x) and g(x) denotes generally complex functions in $L^2(-\infty, \infty)$ of the real variable x. If

(1) the Fourier transform F(ω) of f(x) vanished for $|\omega| > a$ and the Fourier transform G(ω) of g(x) vanishes for $|\omega| < a$, where a is an arbitrary positive constant, or

(2) f(x) and g(x) are analytic (i. e., their real and imaginary parts are Hilbert pairs),

then the Hilbert transform of the product of f(x) and g(x) is given

$$H\{ f(x)\, g(x) \} = f(x)\, H\{ g(x) \} .$$

Bedrosian, E., 1963: A Product theorem for Hilbert Transform, Proceedings of the IEEE, 51, 868-869.

Nuttall Theorem

For any function ***x(t),*** having a quadrature ***xq(t),*** and a Hilbert transform ***xh(t);*** then,

$$E = \int_{0}^{\infty} \left| xq(t) - xh(t) \right|^{2} dt$$

$$= 2 \int_{-\infty}^{0} \left| F_{q}(\omega) \right|^{2} d\omega \ ,$$

where $Fq(\omega)$ is the spectrum of ***xq(t).***

Nuttall, A. H., 1966: On the quadrature approximation to the Hilbert Transform of modulated signal, Proc. IEEE, 54, 1458

Problems with Hilbert Transform method

- If there is any amplitude change, the Fourier Spectra for the envelope and carrier are not separable. Thus, we violated the limitations stated in the Bedrosian Theorem; drastic amplitude change produce drastic deteriorating results. Nuttall theorem further reduce the applicability of Hilbert transform.

- Once we cannot separate the envelope and the carrier, the analytic signal through Hilbert Transform would not give the phase function of the carrier alone without the influence of the variation from the envelope.

- Therefore, the instantaneous frequency computed through the analytic signal ceases to have full physical meaning; it provides an approximation only.

Quadrature : Procedures

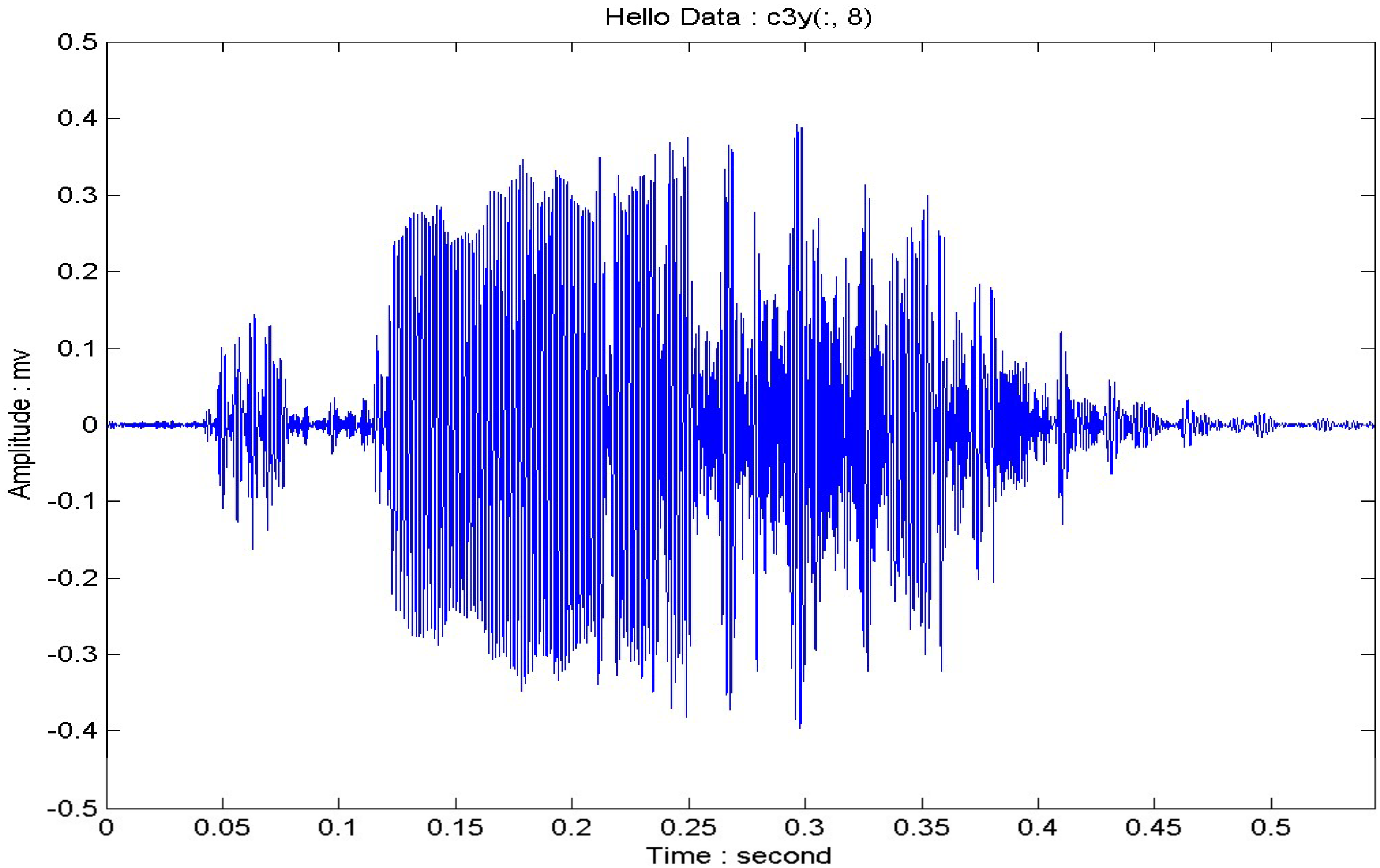
- **Normalize the IMFs as in the NHHT method.**
- **Compute IF (FM) from Quadrature of N-data as follows:**

Quadrature method

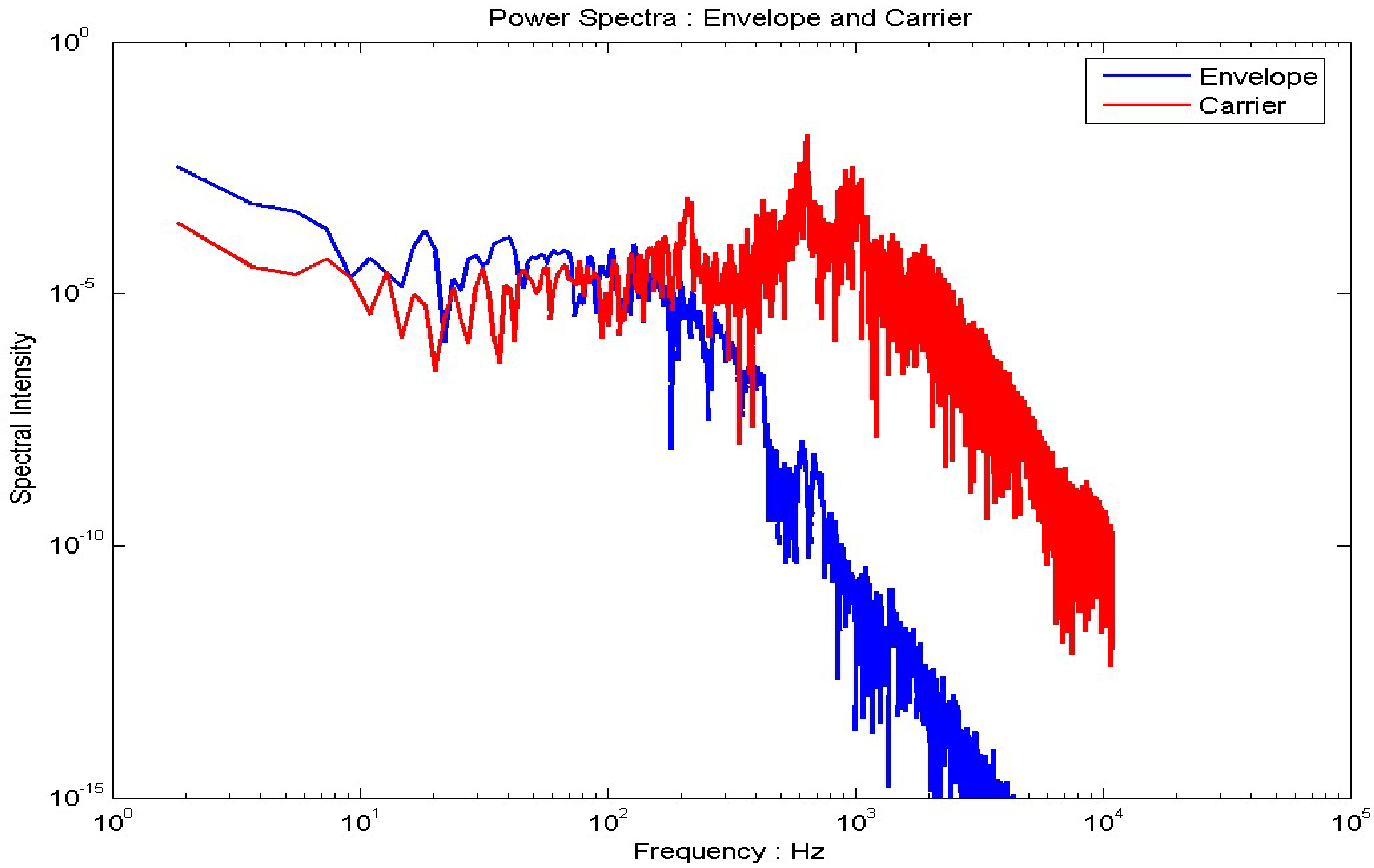
$$Normalized\ x\ ;\quad y = \sqrt{1 - x^2}\quad ;\quad \theta(t) = \tan^{-1}\frac{y}{x}$$

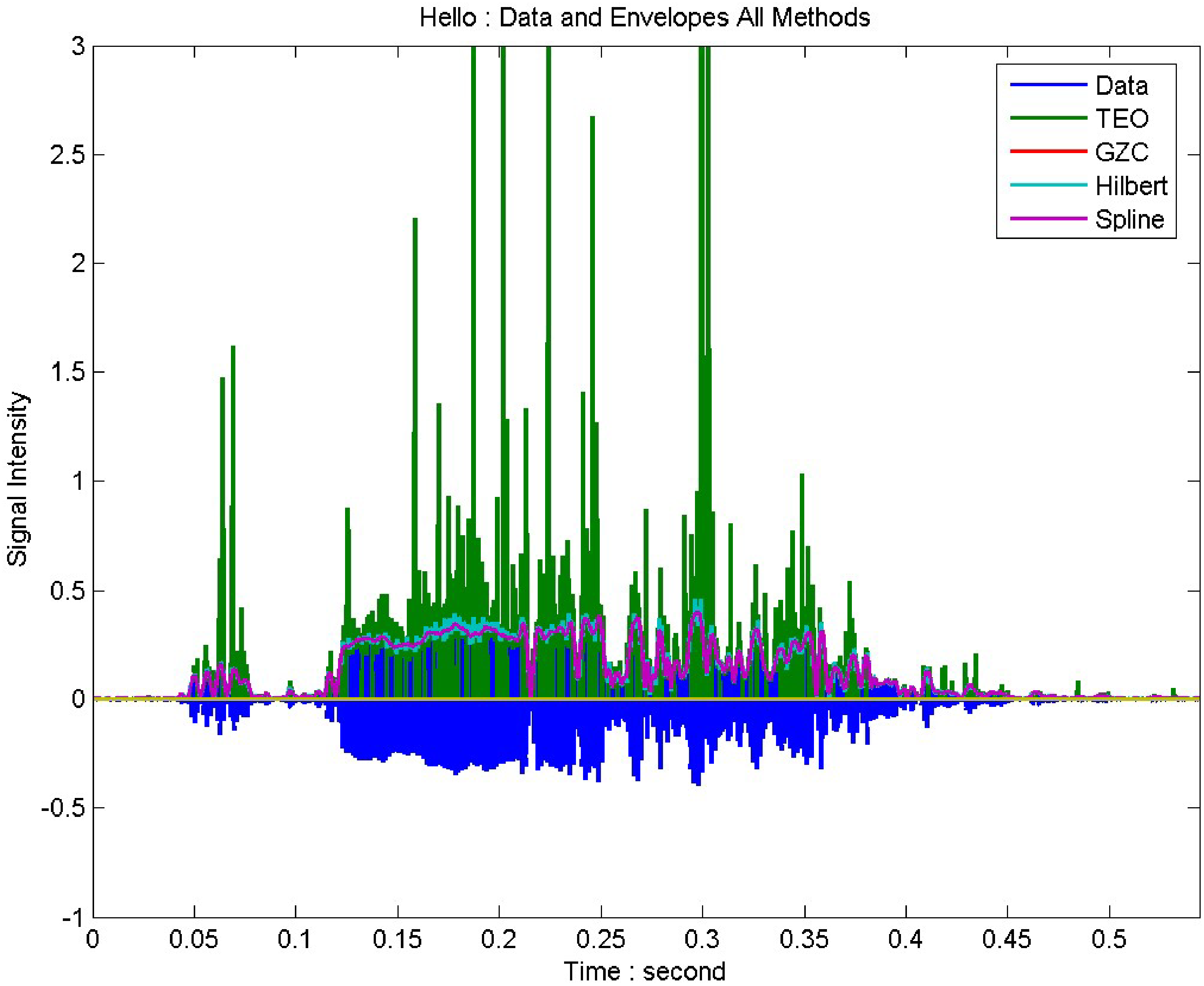
$$\omega(t) = \frac{d\theta(t)}{dt}\ .$$

Hello : Data c3y(8)

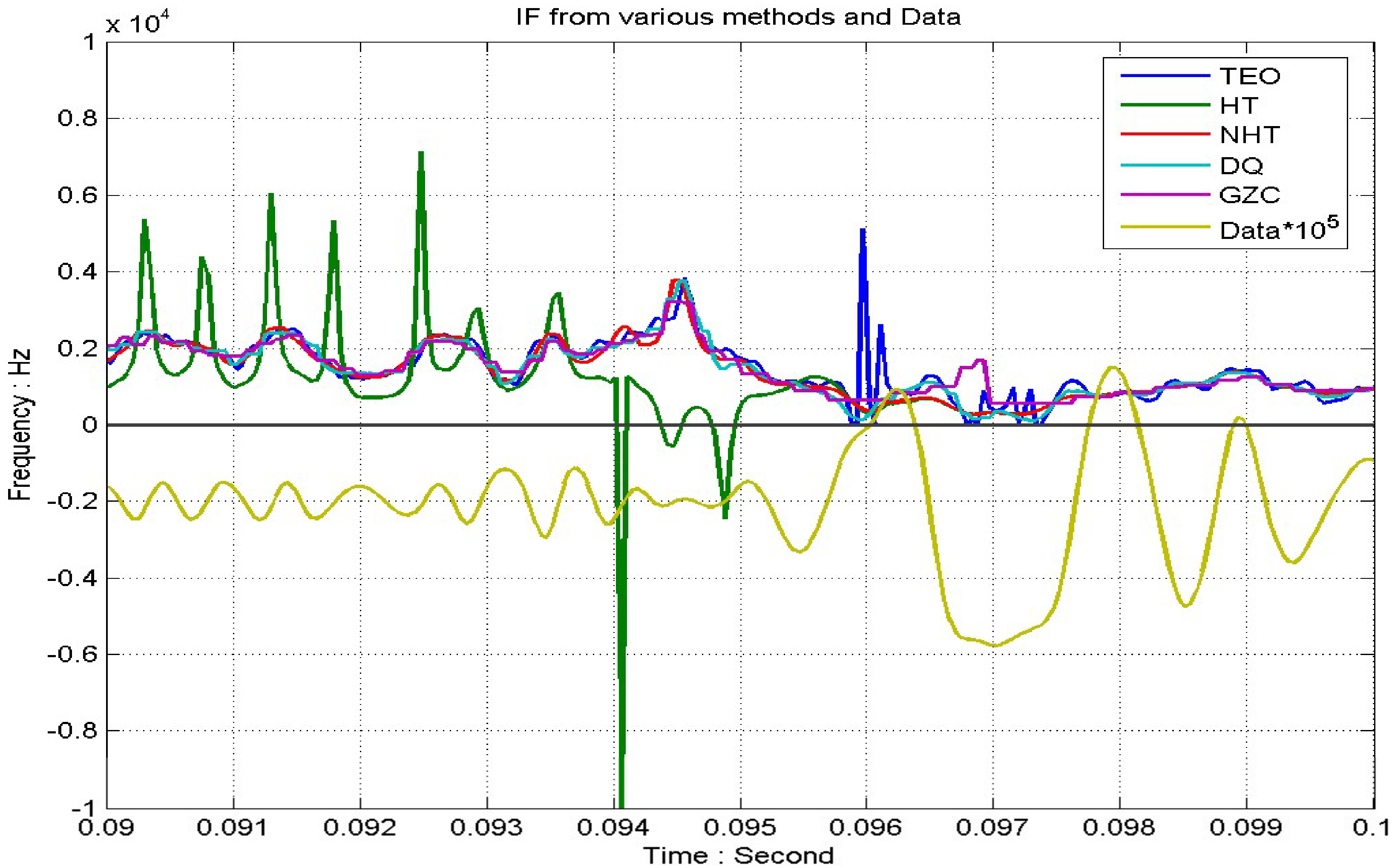


Hello : Check Bedrosian Theorem

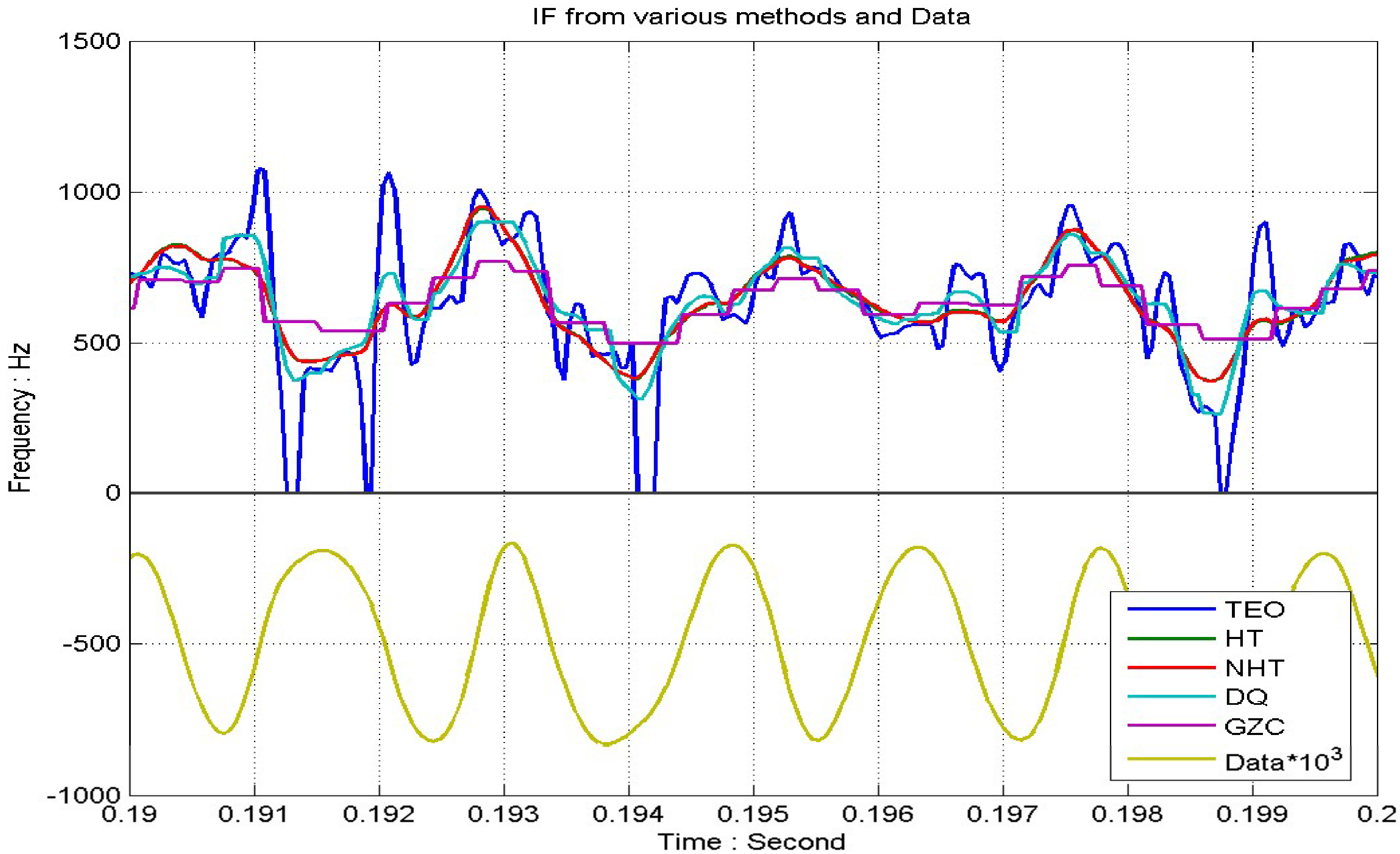



Hello : Data and Envelopes All Methods
Data
TEO
GZC
Hilbert
Spline
Signal Intensity
Time : second
3
2.5
2
1.5
1
0.5
0
-0.5
-1
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0.45
0.5

Hello : Instantaneous Frequency & data c3y(8)



Hello : Instantaneous Frequency & data Details c3y(8)



Summary

- Instantaneous Frequency could be calculated routinely from the normalized IMFs through quadrature (for high data density) or Hilbert Transform (for low data density).
- For any signal, there might be more than one IF value at any given time.
- For data from nonlinear processes, there has to be intra-wave frequency modulations; therefore, the Instantaneous Frequency could be highly variable.

Some Recent Advances

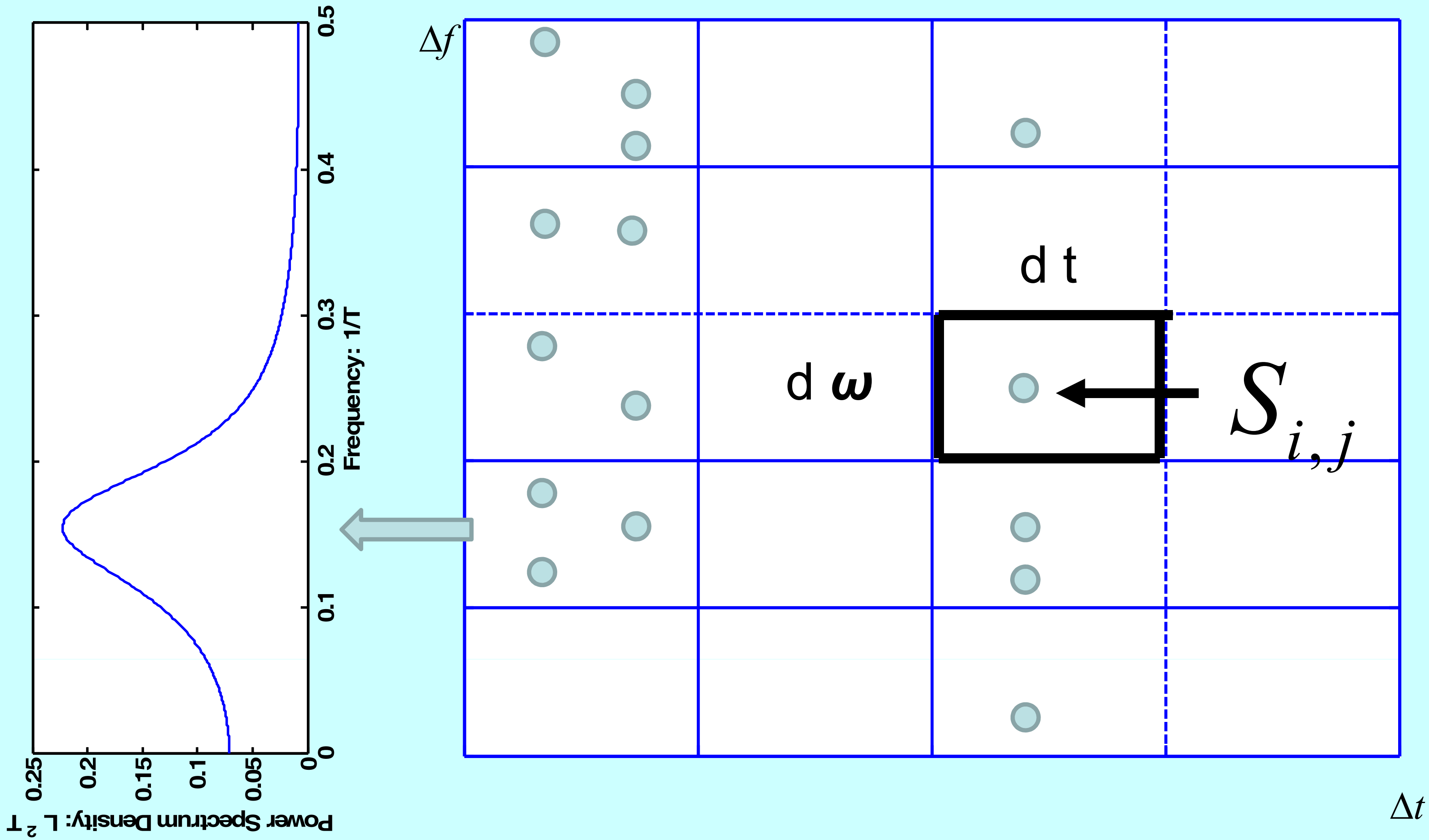
- Ensemble EMD (EEMD)
- Instantaneous Frequency (Quadrature)
- Quantification of Hilbert Spectrum
- Multi-Dimensional EEMD (MDEEMD)
- Time Dependent Intrinsic Correlation (TDIC)

Definition of Hilbert Spectra

The Hilbert Energy Spectrum is defined as the energy density distribution in a time-frequency space divided into equal size bins of $\Delta t \times \Delta\omega$ with the value in each bin designated as $a^2(t)$ at the proper time, t, and the proper instantaneous frequency, ω.

The Hilbert Amplitude Spectrum is defined as the amplitude density distribution in a time-frequency space divided into equal size bins of $\Delta t \times \Delta\omega$ with the value in each bin designated as $a(t)$ at the proper time, t, and the proper instantaneous frequency, ω.

Schematic of Hilbert Spectrum



$S_{i,j}$ can be amplitude or the square of amplitude (energy).

Hilbert Spectra

Currently, the spectrum is not defined in terms of density.
The value is simply the energy value at the particuler bin.

To define the spectra in terms of density would facilitate comparison with the Fourier spectra which is defined in terms of density.

Therefore, the value in each bin should be

$$\frac{a_{i,j}}{\Delta t \times \Delta \omega} \quad \textit{for amplitude spectra}$$

$$\frac{a_{i,j}^2}{\Delta t \times \Delta \omega} \quad \textit{for energy spectra}$$

Definition of the Marginal Hilbert Spectrum

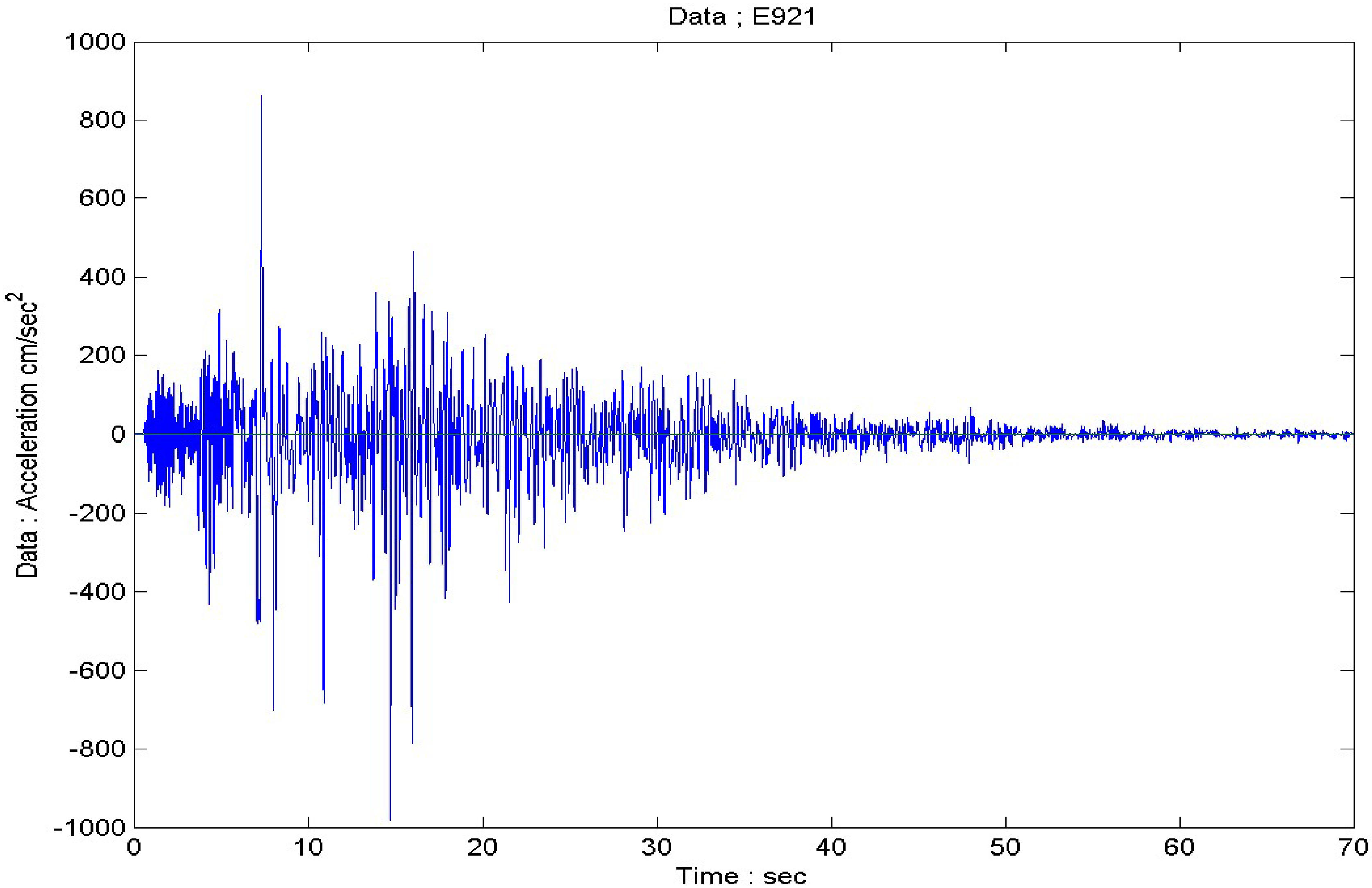
Given the Hilbert Spectrum as $H(\omega,t)$, the Marginal Spectrum is defined simply as

$$h(\omega) = \int_0^T H(\omega,t)\,dt = \sum_{i=1}^{N} H(\omega, t_i)\ .$$

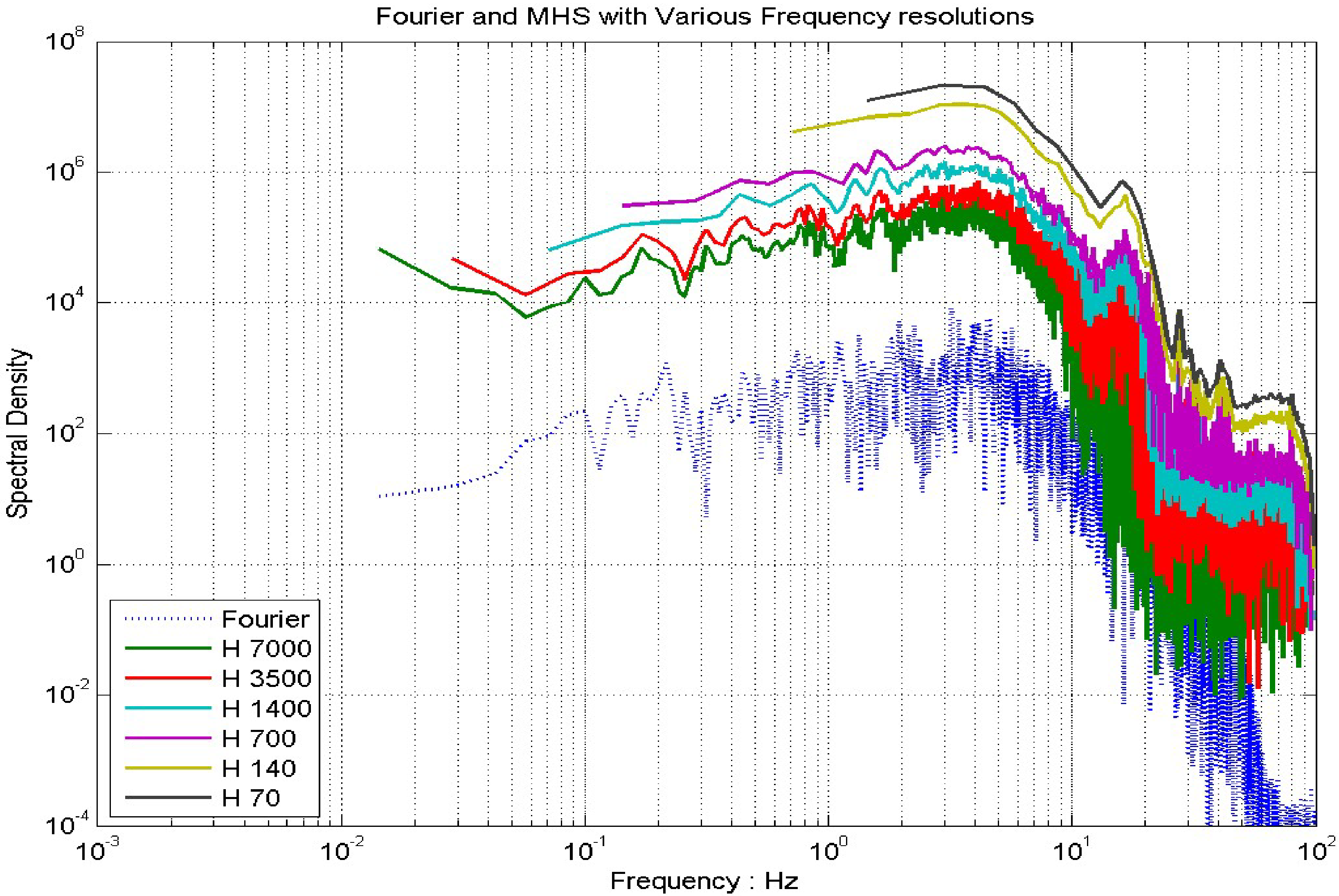
Simple as it seems, the actual computation and evaluation is more involved. The main reason is that, with the adaptive basis, we do not have the rigid limitation on frequency resolution dictated by the total data length and the uncertainty principle.

The freedom on our choice of frequency-time resolution; however, makes the marginal frequency evaluation much more complicated. We need to define it rigorously for detailed comparisons with other forms of spectrum.

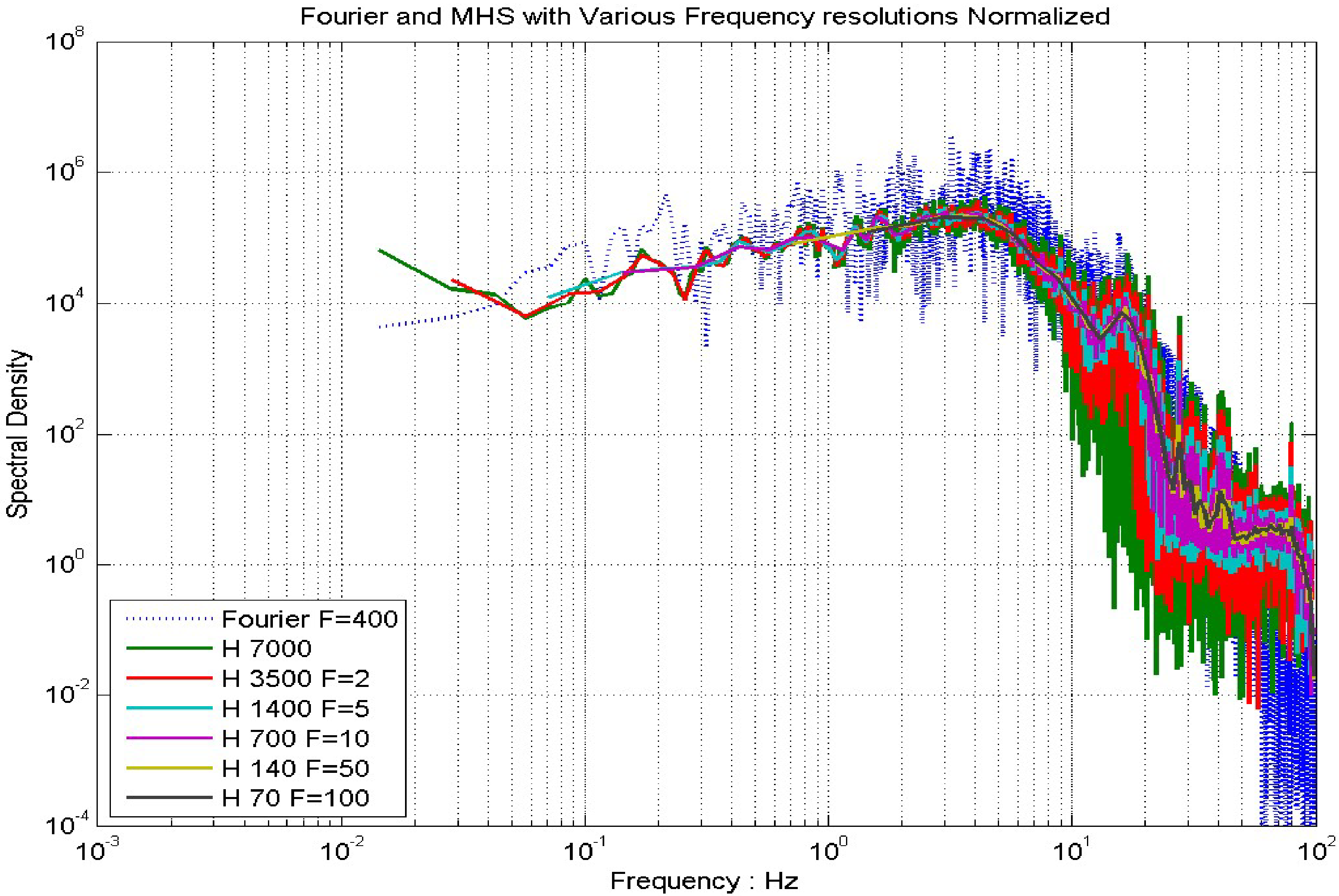
Earthquake data E921



MHS Different Frequency Resolutions



MHS Different Resolutions Normalized



Some Recent Advances

- Ensemble EMD (EEMD)
- Instantaneous Frequency (Quadrature)
- Quantification of Hilbert Spectrum
- Multi-Dimensional EEMD (MDEEMD)
- Time Dependent Intrinsic Correlation (TDIC)

Advances in Adaptive Data Analysis
Vol. 1, No. 3 (2009) 339–372

THE MULTI-DIMENSIONAL ENSEMBLE EMPIRICAL MODE DECOMPOSITION METHOD

ZHAOHUA WU

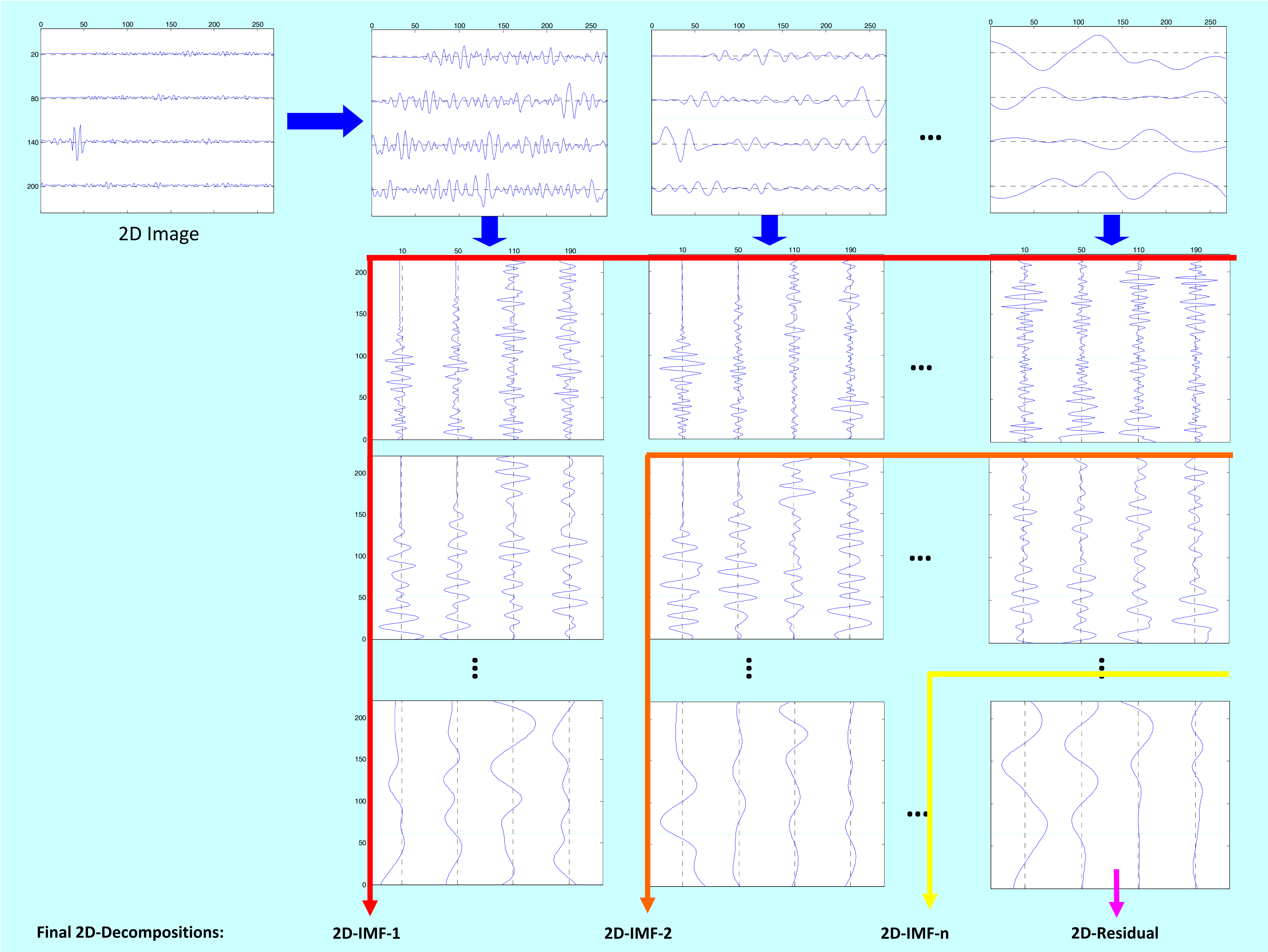
Department of Meteorology and
Center for Ocean-Atmospheric Prediction Studies
Florida State University
Tallahassee, FL 32306, USA

NORDEN E. HUANG

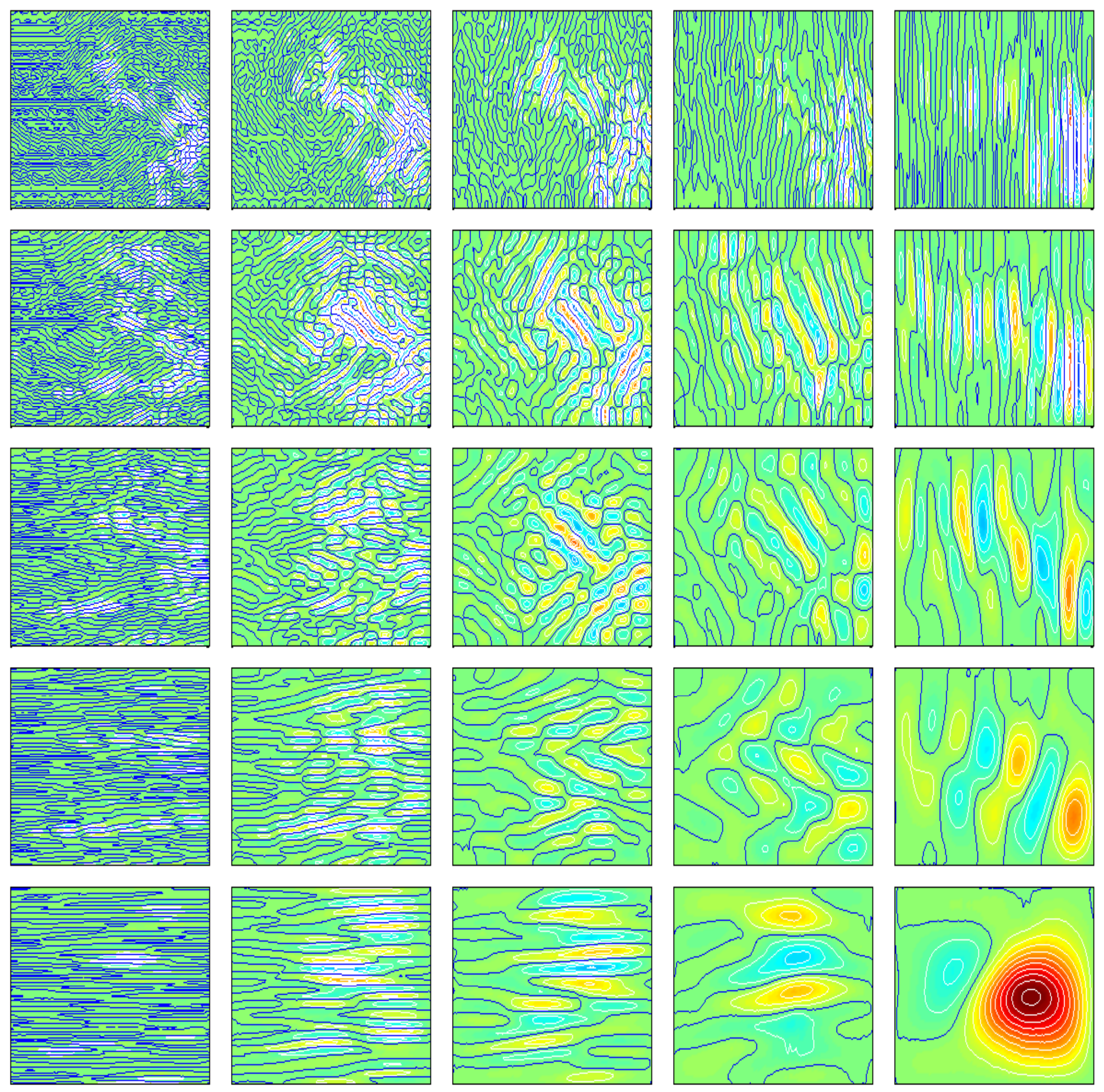
Research Center for Adaptive Data Analysis
National Central University
Chungli, Taiwan 32001, ROC

XIANYAO CHEN

The First Institute of Oceanography, SOA
Qingdao 266061, People's Republic of China



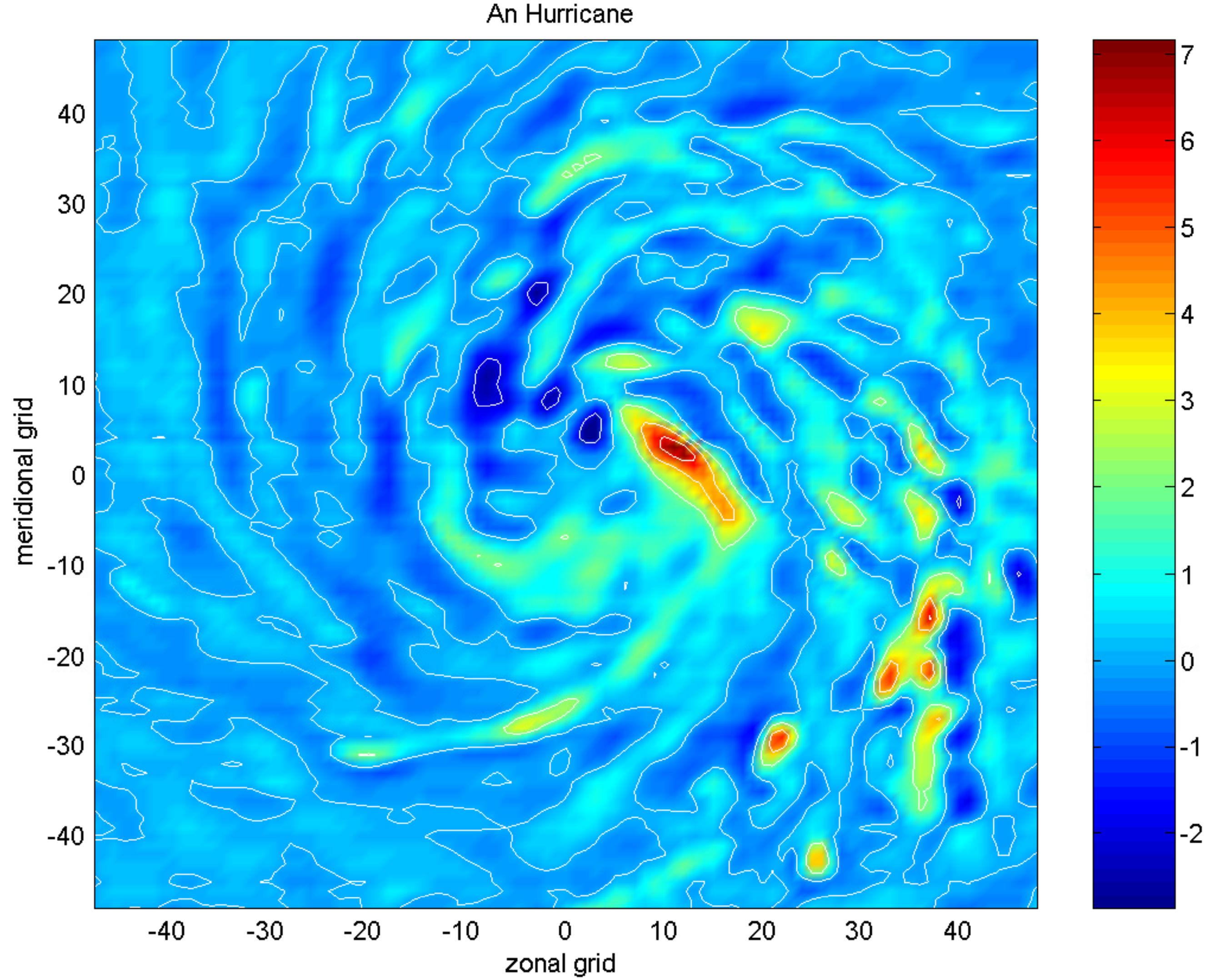
2D Image
Final 2D-Decompositions:
2D-IMF-1
2D-IMF-2
2D-IMF-n
2D-Residual



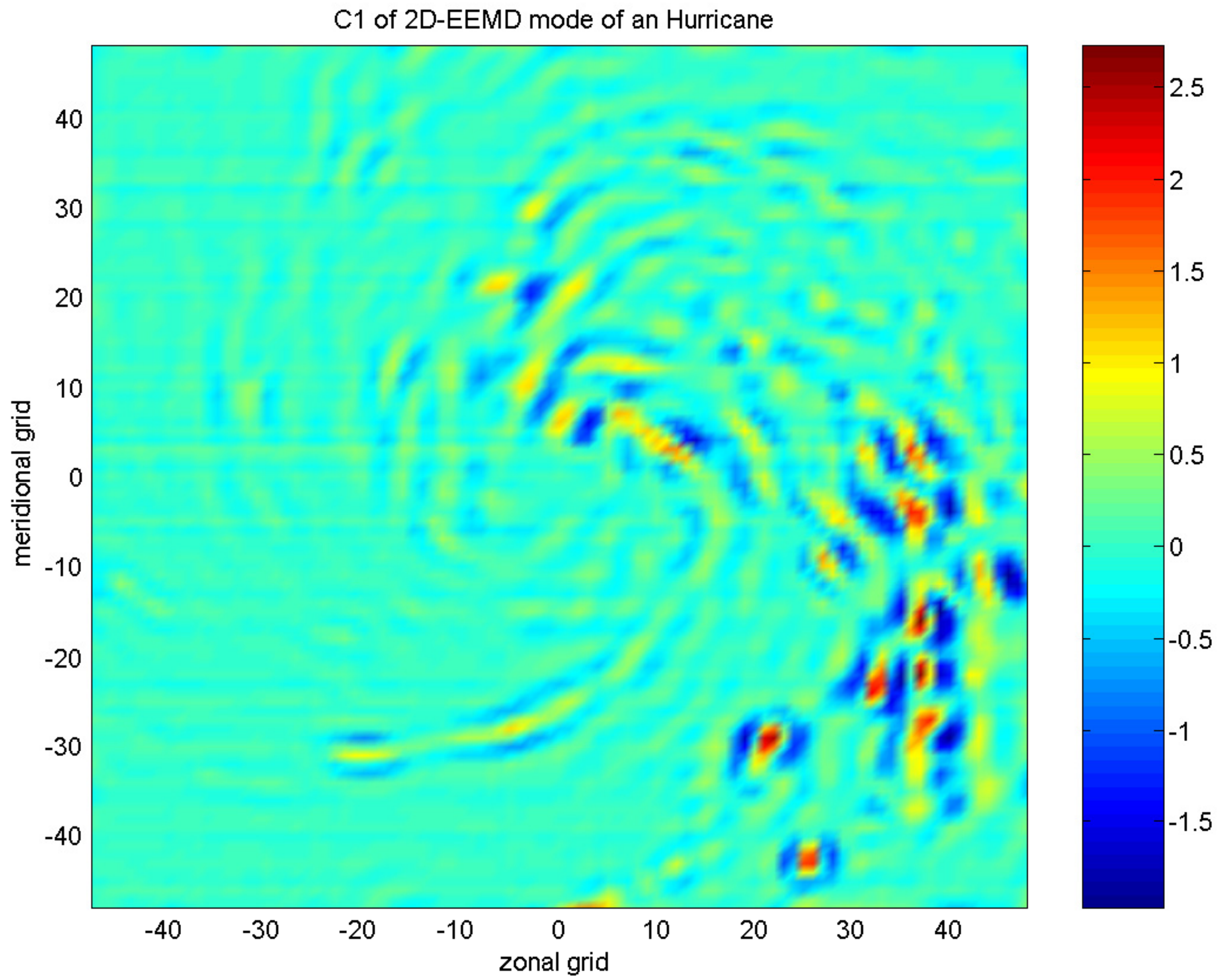
The EEMD components of the vertical velocity.

In each panel, the color scales are different, the blue lines corresponding to zeros.

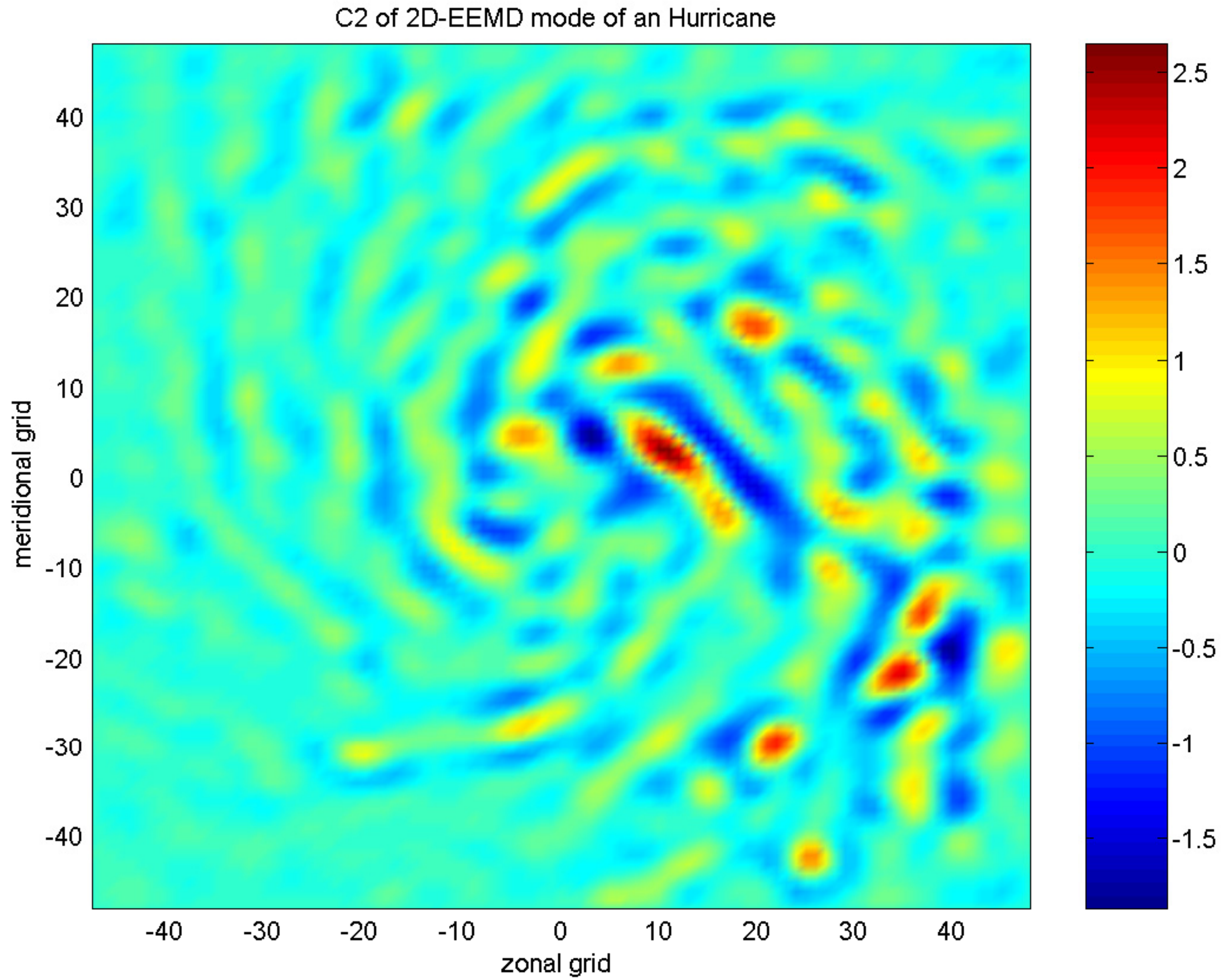
HURRICANE (w): DATA



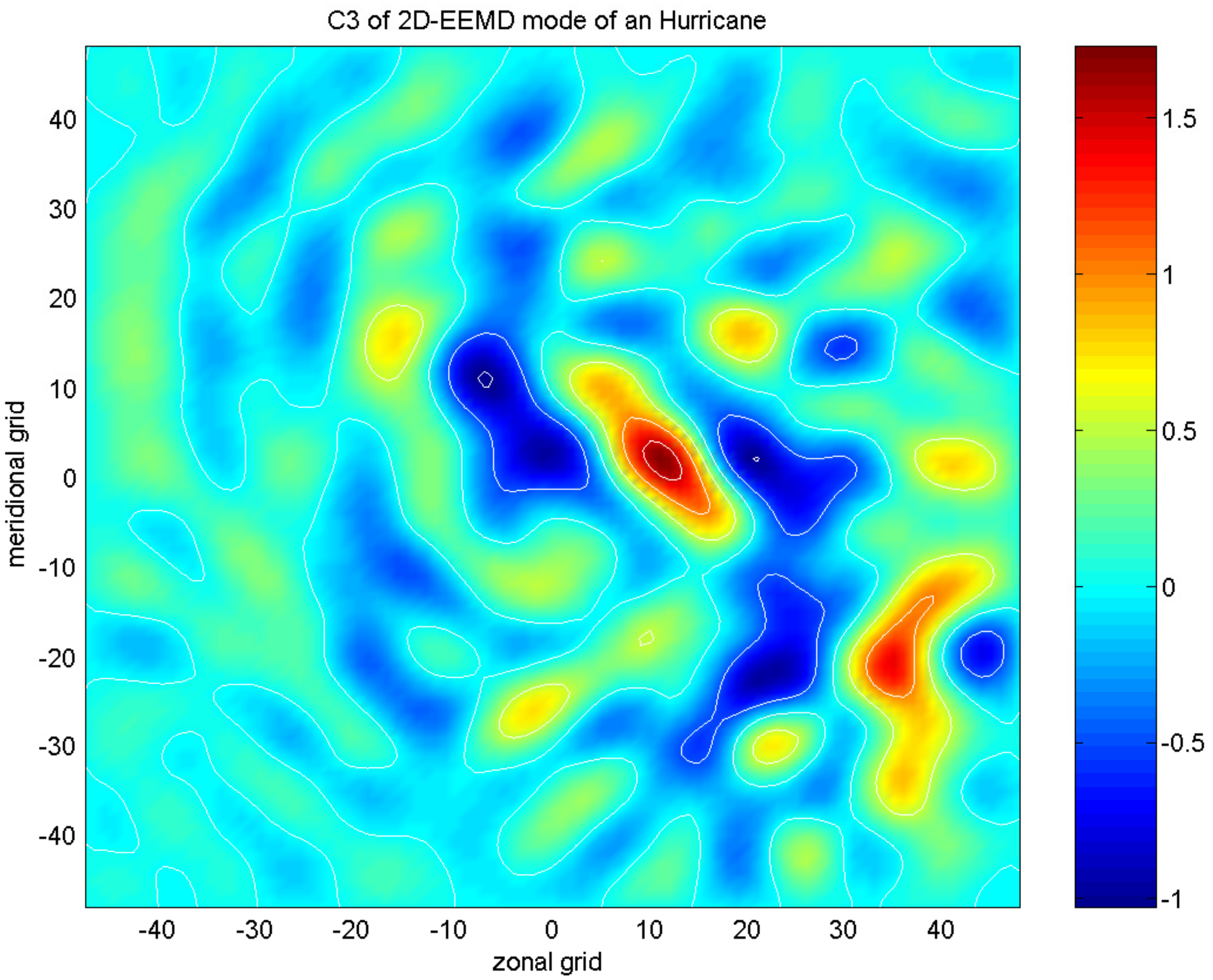
HURRICANE (w): C1



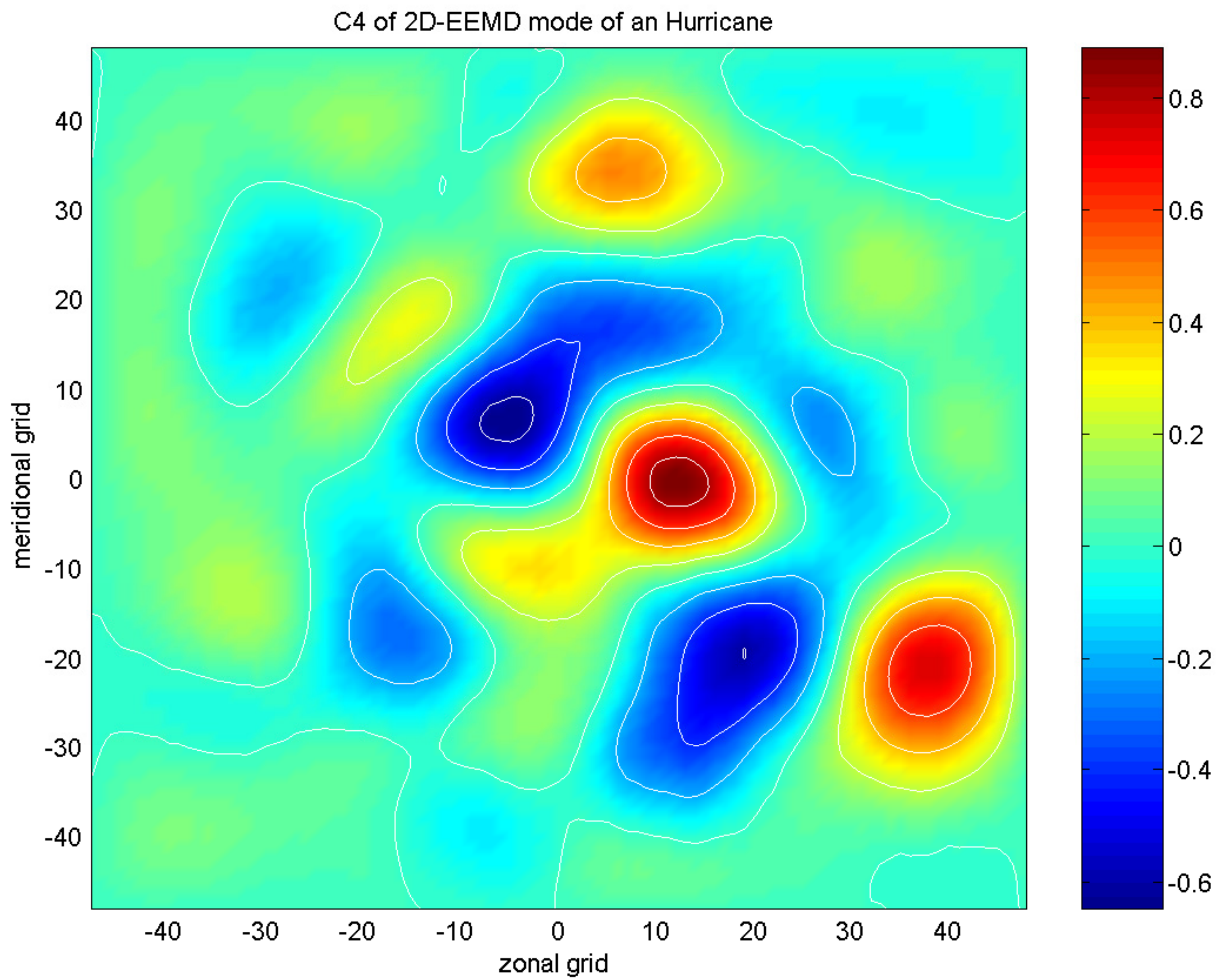
HURRICANE (w): C2



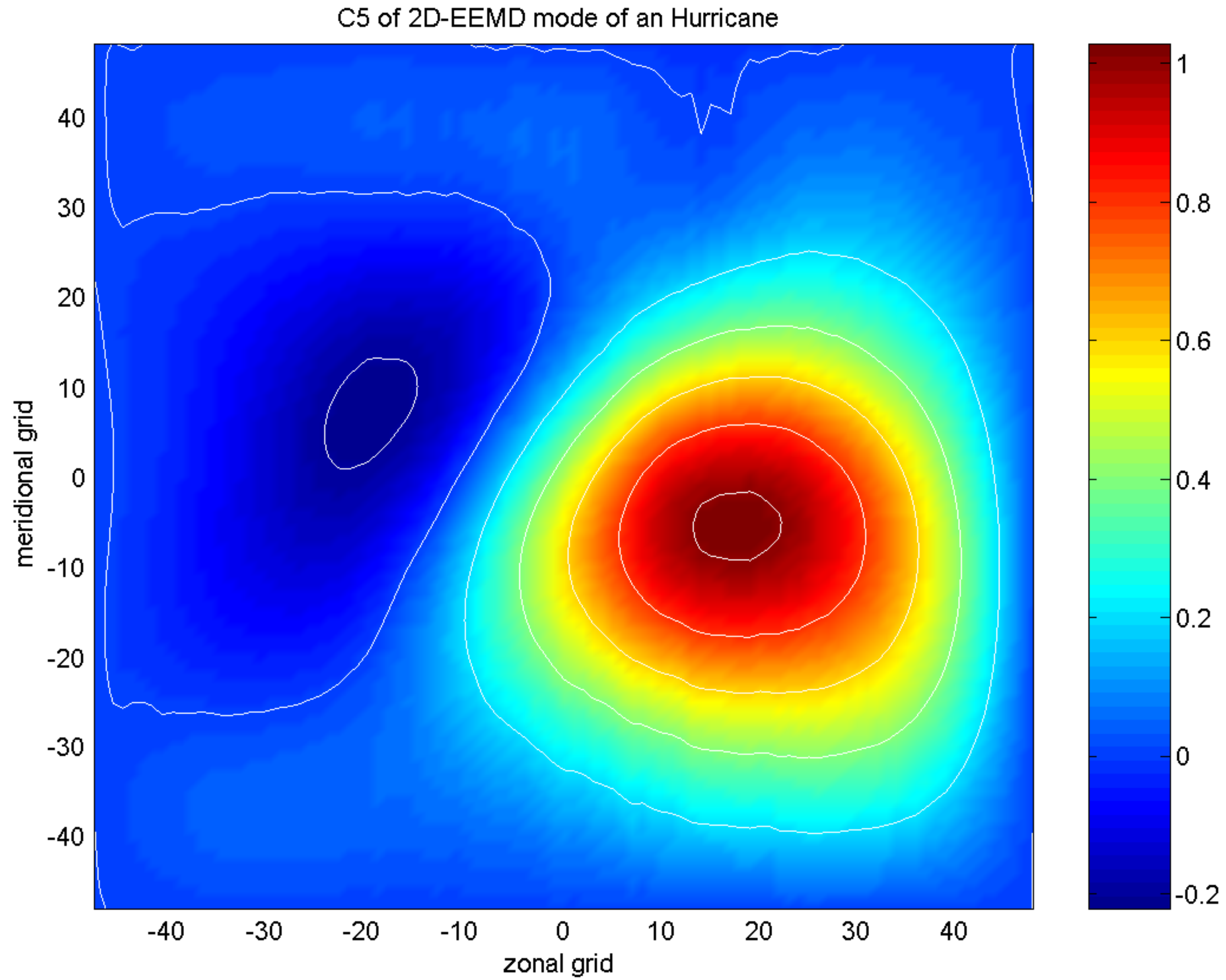
HURRICANE (w): C3



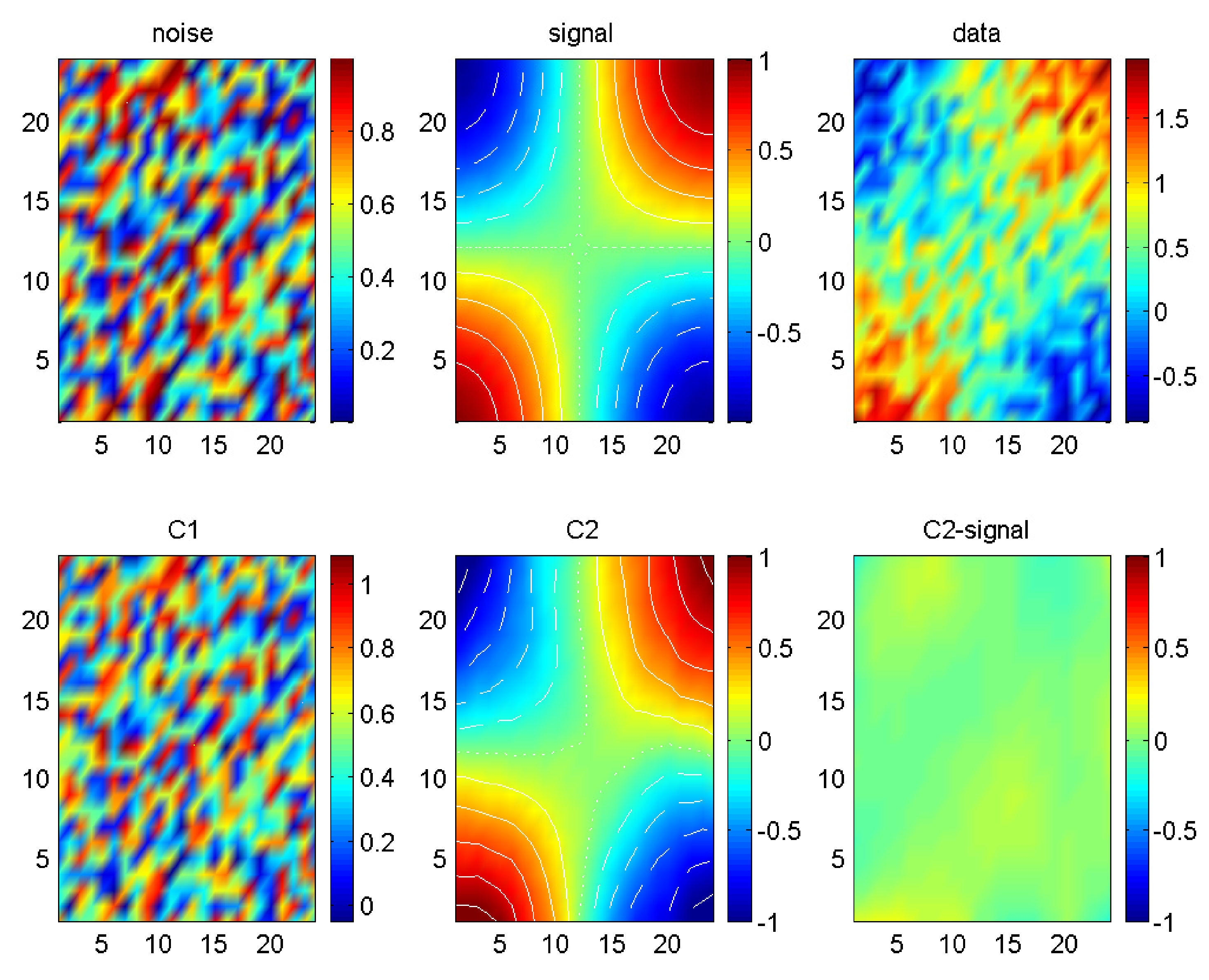
HURRICANE (w): C4



HURRICANE (w): C5



Two dimensional HHT



Conclusion

Adaptive method is the only scientifically meaningful way to analyze data.

It is the only way to find out the underlying physical processes; therefore, it is indispensable in scientific research.

It is physical, direct, and simple.

But, we have only started and what we have done is only a scratch of the surface.

Outstanding Mathematical Problems

1. Adaptive data analysis methodology in general

2. Nonlinear system identification methods

3. Prediction problem for nonstationary processes (end effects)

4. Optimization problem (the best IMF selection and uniqueness. Is there a unique solution?)

5. Spline problem (best spline implement of HHT, convergence and 2-D)

http://rcada.ncu.edu.tw/research1.htm

At this website you will find all the programs used in HHT and many references.

Some Recent Advances

- Ensemble EMD (EEMD)
- Instantaneous Frequency (Quadrature)
- Quantification of Hilbert Spectrum
- Multi-Dimensional EEMD (MDEEMD)
- Time Dependent Intrinsic Correlation (TDIC)

Advances in Adaptive Data Analysis
Vol. 1, No. 1 (2009) 1–41

ENSEMBLE EMPIRICAL MODE DECOMPOSITION: A NOISE-ASSISTED DATA ANALYSIS METHOD

ZHAOHUA WU* and NORDEN E. HUANG†

*Center for Ocean–Land–Atmosphere Studies
4041 Powder Mill Road, Suite 302
Calverton, MD 20705, USA

†Research Center for Adaptive Data Analysis
National Central University
300 Jhongda Road, Chungli, Taiwan 32001

Advances in Adaptive Data Analysis
Vol. 1, No. 2 (2009) 177–229

World Scientific
www.worldscientific.com

ON INSTANTANEOUS FREQUENCY

NORDEN E. HUANG

Research Center for Adaptive Data Analysis
National Central University
Chungli, Taiwan 32001, Republic of China
Norden@ncu.edu.tw

ZHAOHUA WU

Department of Meteorology & Center for
Ocean-Atmospheric Prediction Studies
Florida State University
Tallahassee, FL 32306, USA

Advances in Adaptive Data Analysis
Vol. 1, No. 3 (2009) 339–372

THE MULTI-DIMENSIONAL ENSEMBLE EMPIRICAL MODE DECOMPOSITION METHOD

ZHAOHUA WU

Department of Meteorology and
Center for Ocean-Atmospheric Prediction Studies
Florida State University
Tallahassee, FL 32306, USA

NORDEN E. HUANG

Research Center for Adaptive Data Analysis
National Central University
Chungli, Taiwan 32001, ROC

XIANYAO CHEN

The First Institute of Oceanography, SOA
Qingdao 266061, People's Republic of China