The Ensemble Empirical Mode Decomposition:

A Noise-Assisted Data Analysis Method

Zhaohua Wu[1], Norden E. Huang[2, 3], and Xianyao Chen[3]

[1]Department of Earth, Ocean, and Atmospheric Science
Florida State University

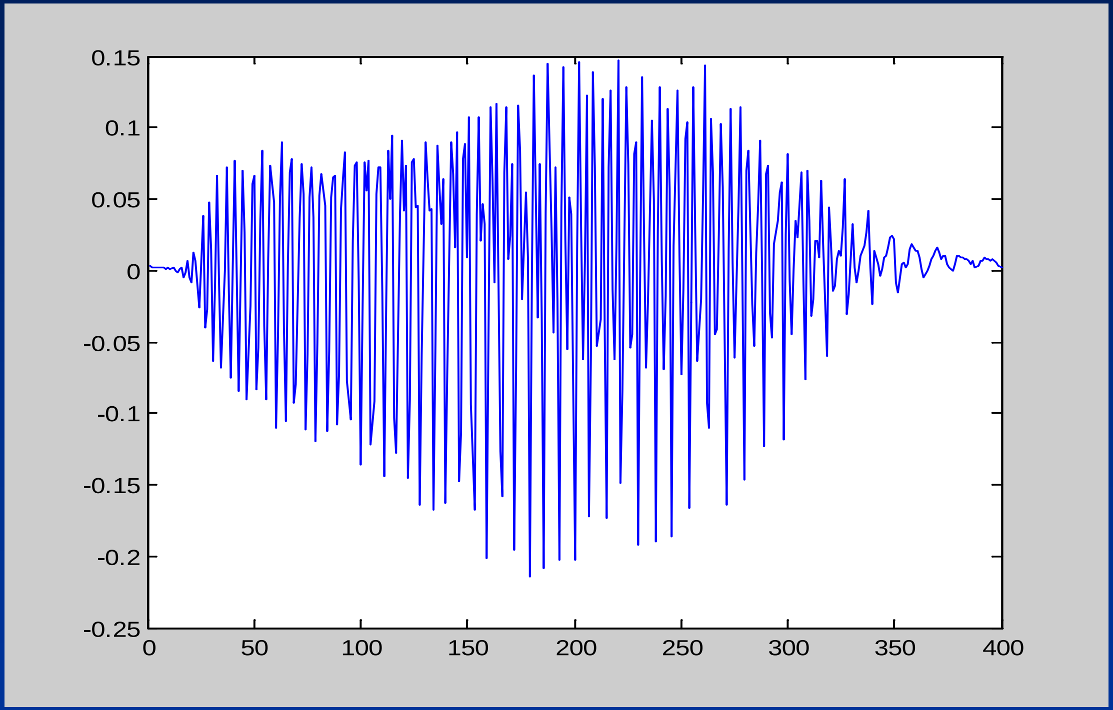
[2]Research Center for Adaptive Data Analysis
National Central University, Taiwan

[3]The First Institute of Oceanography
State Oceanic Administration, China

OUTLINE

- The Empirical Mode Decomposition (EMD)
- The Ensemble EMD (EEMD)
- Multi-dimensional EEMD

A SAMPLE INPUT



Task: To decompose this recorded vocal signal into "physically" meaningful components.

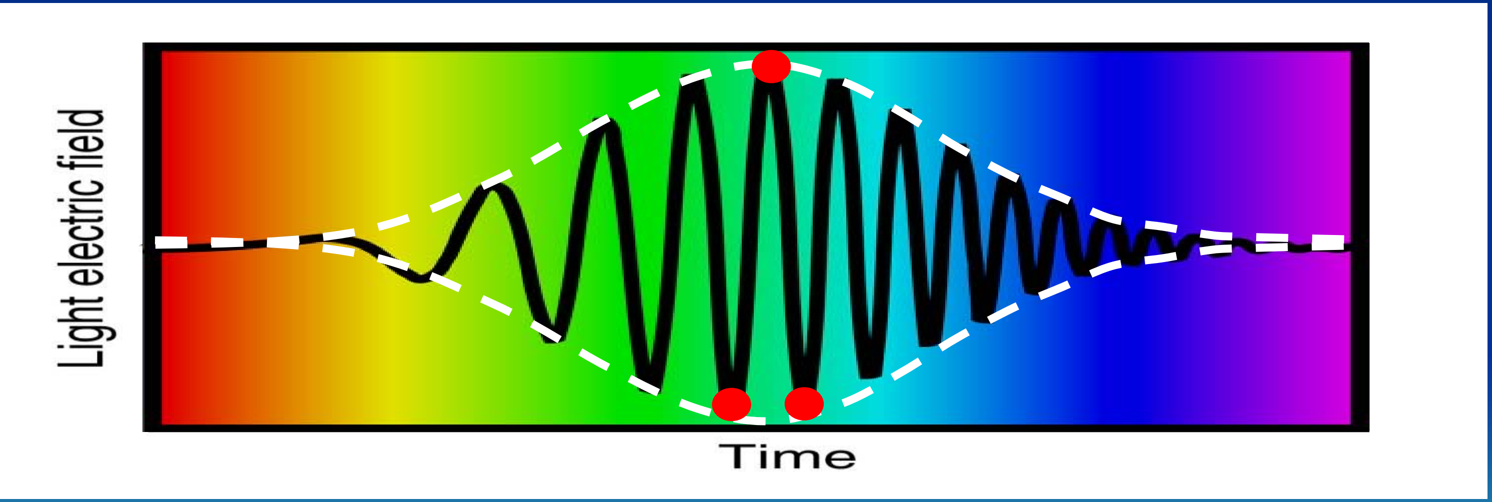
SOME CRETERIA FOR DECOMPOSITION

- From Ingrid Daubechies (Dec. 2008)

  - "natural" waveforms need not be of Fourier (or wavelet, or curvelet, ...) type
  - nevertheless, it can still be useful to decompose signals into more elementary waveforms
  - such a decomposition should be adaptive (of course)
  - should produce meaningful components
  - should also be robust to noise

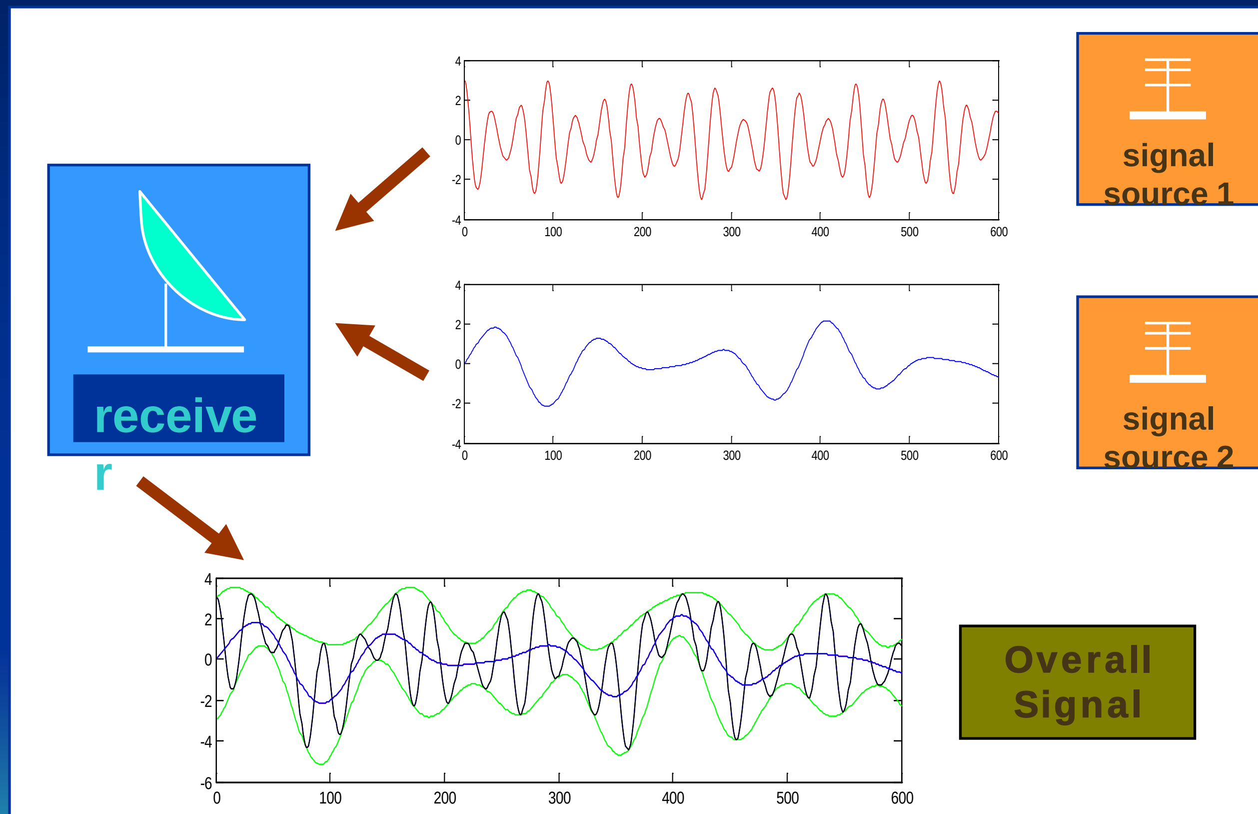
"NATURAL" WAVEFORM

$$A(t)\cdot\cos\left[\int_{t_0}^{t}\omega(\tau)\,d\tau\right]$$

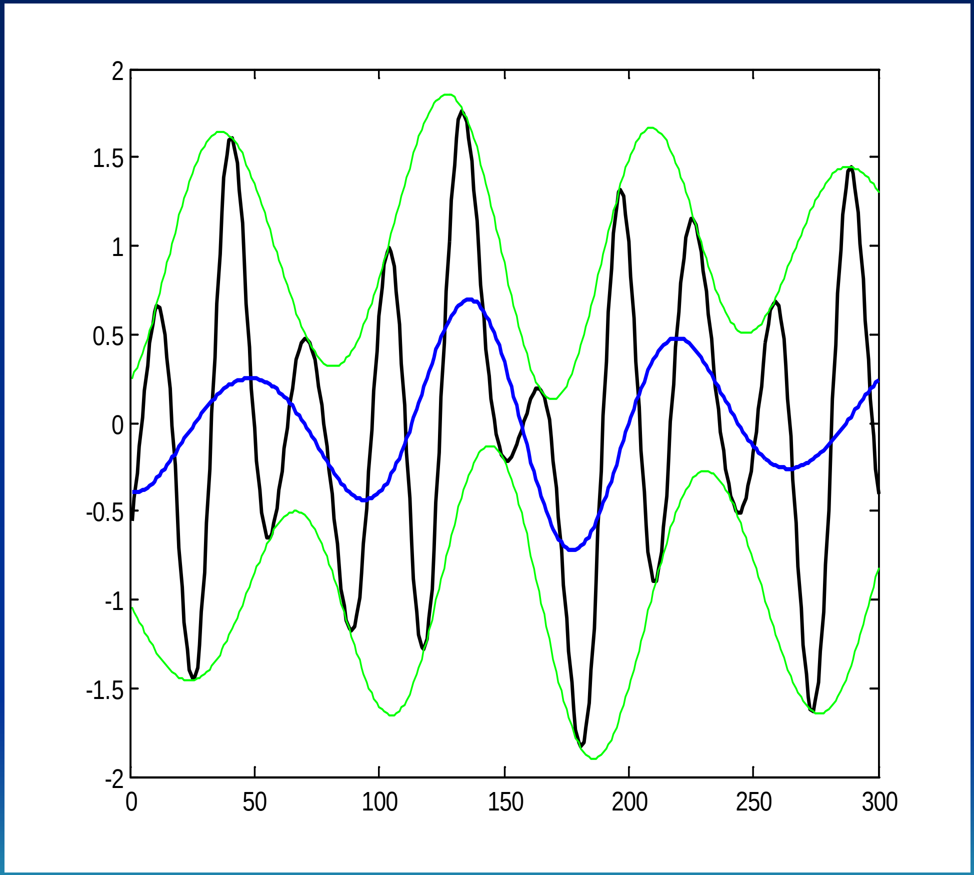


EMPIRICAL MODE DECOMP.

Norden E Huang



EMPIRICAL MODE DECOMPOSITION



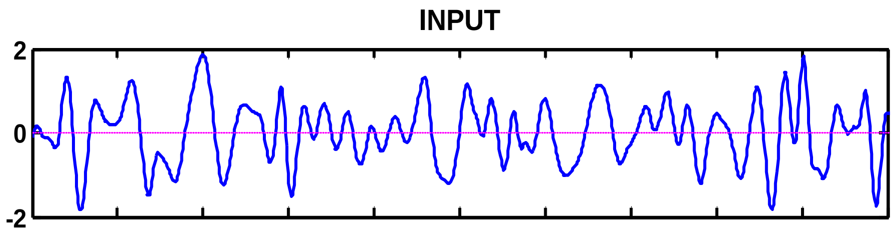
$$x(t) - m_1 = h_1 ,$$
$$h_1 - m_2 = h_2 ,$$
$$\cdots\cdots$$
$$\cdots\cdots$$
$$h_{k-1} - m_k = h_k .$$

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

$$x(t) - c_1 = r_1 ,$$
$$r_1 - c_2 = r_2 ,$$
$$\cdots$$
$$r_{n-1} - c_n = r_n .$$

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

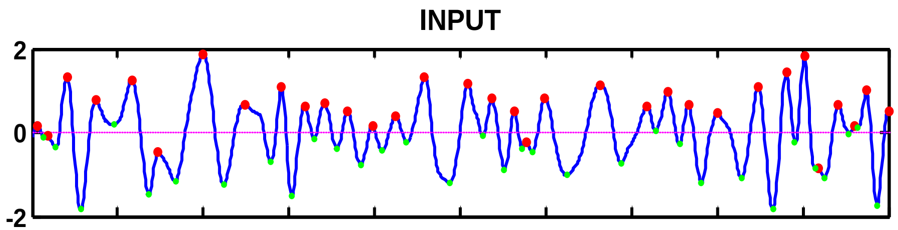
SIFTING



$$x(t) - \boxed{m_1} = h_1 \ ,$$
$$h_1 - m_2 = h_2 \ ,$$
$$\cdots\cdots$$
$$\cdots\cdots$$
$$h_{k-1} - m_k = h_k \ .$$

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

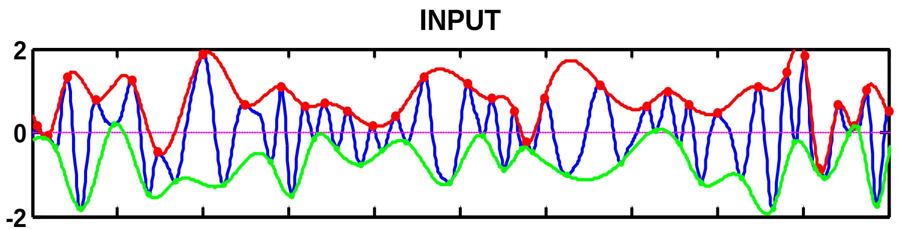
SIFTING



$$x(t) - \boxed{m_1} = h_1 \ ,$$
$$h_1 - m_2 = h_2 \ ,$$
$$\cdots\cdots$$
$$\cdots\cdots$$
$$h_{k-1} - m_k = h_k \ .$$

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

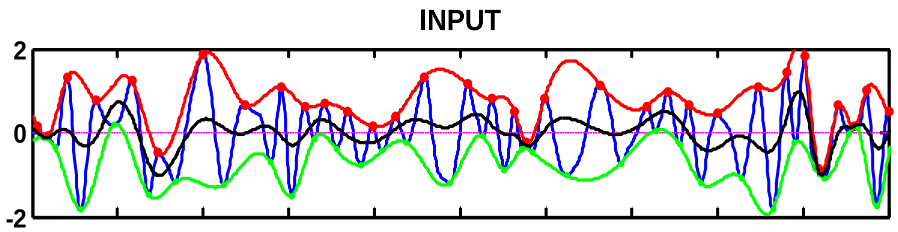
SIFTING



$$x(t) - \boxed{m_1} = h_1 \,,$$
$$h_1 - m_2 = h_2 \,,$$
$$\cdots\cdots$$
$$\cdots\cdots$$
$$h_{k-1} - m_k = h_k \,.$$

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

SIFTING



Blue line

Black line

$$x(t) - \boxed{m_1} = h_1 ,$$

$$h_1 - m_2 = h_2 ,$$

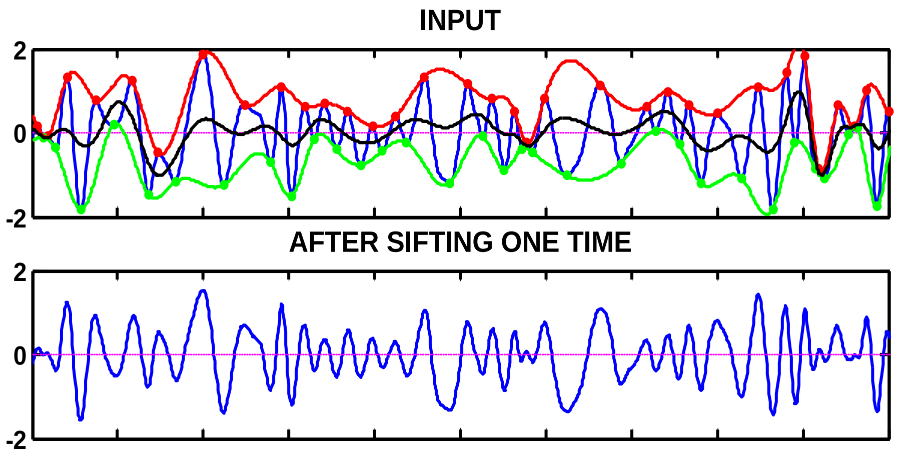
.....

.....

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

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

SIFTING



Blue line

Black line

$$x(t) - \boxed{m_1} = \boxed{h_1},$$

$$h_1 - m_2 = h_2,$$

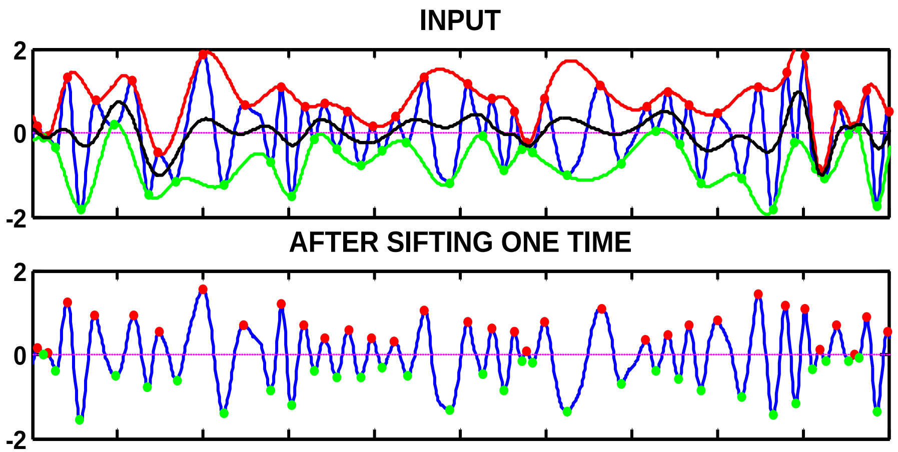
.....

.....

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

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

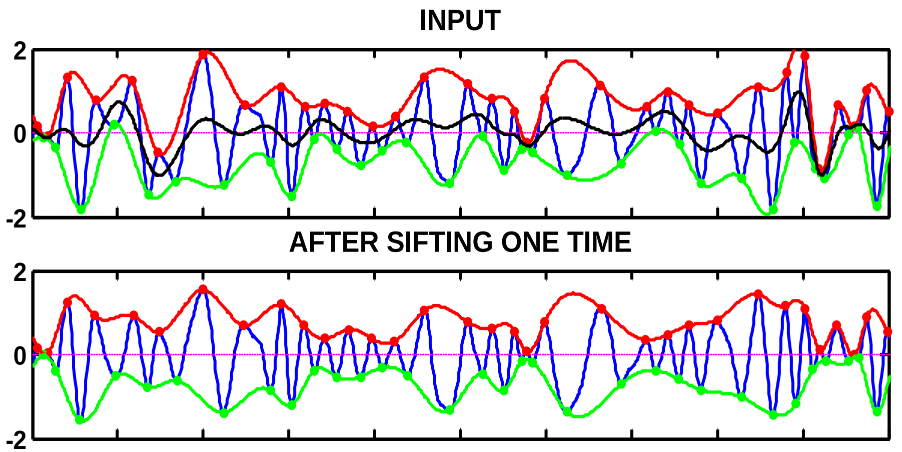
SIFTING



$$x(t) - m_1 = h_1 \,,$$
$$\boxed{h_1} - m_2 = h_2 \,,$$
$$\cdots\cdots$$
$$\cdots\cdots$$
$$h_{k-1} - m_k = h_k \,.$$

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

SIFTING



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

$$\boxed{h_1} - m_2 = h_2 \,,$$

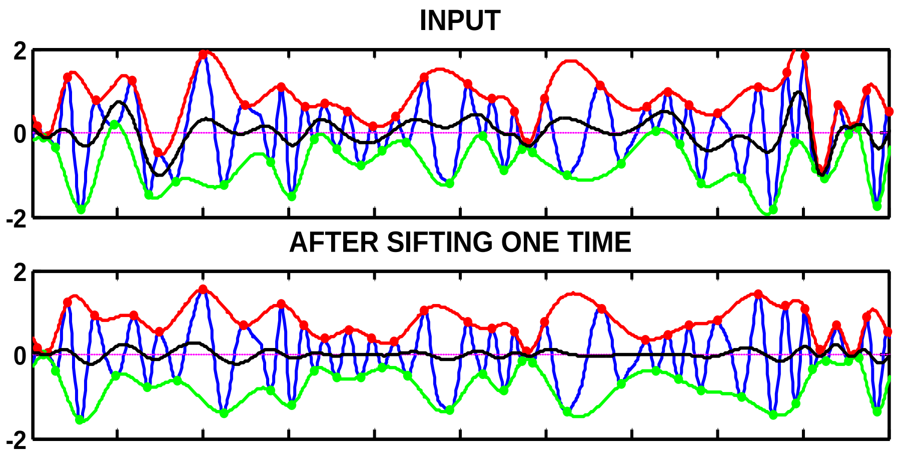
.....

.....

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

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

SIFTING



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

$$\boxed{h_1 - m_2 = h_2 \,,}$$

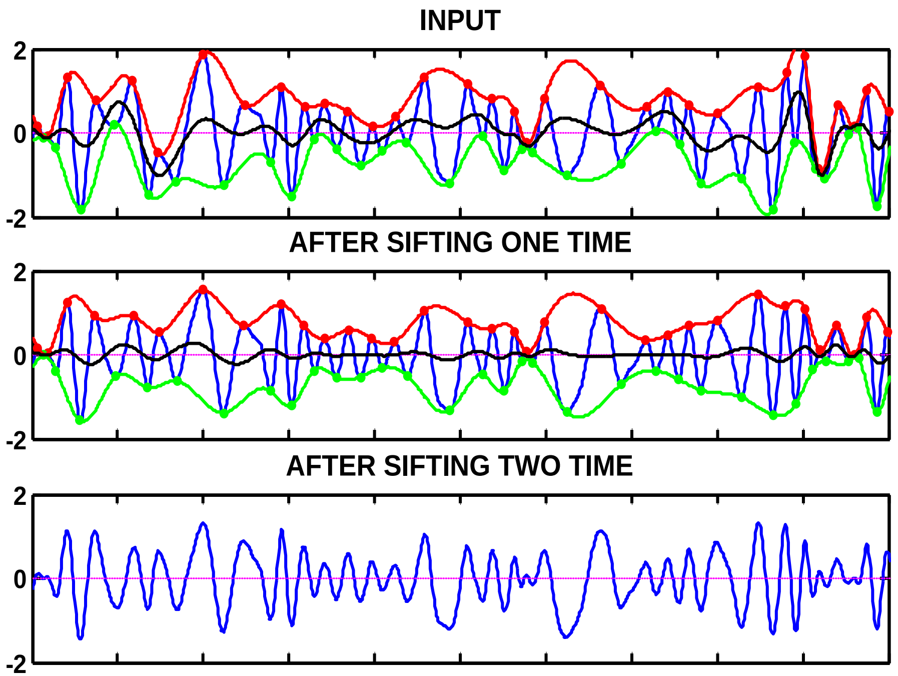
.....

.....

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

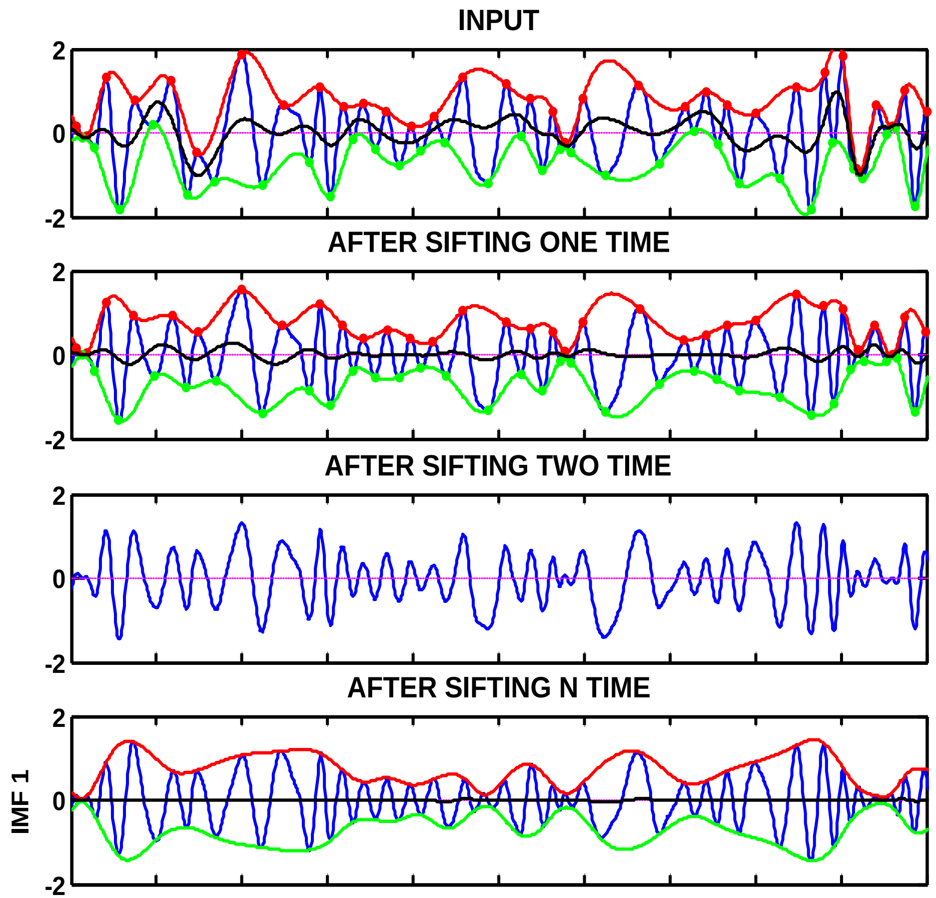
$$\Rightarrow \quad h_k = c_1 \,.$$

SIFTING



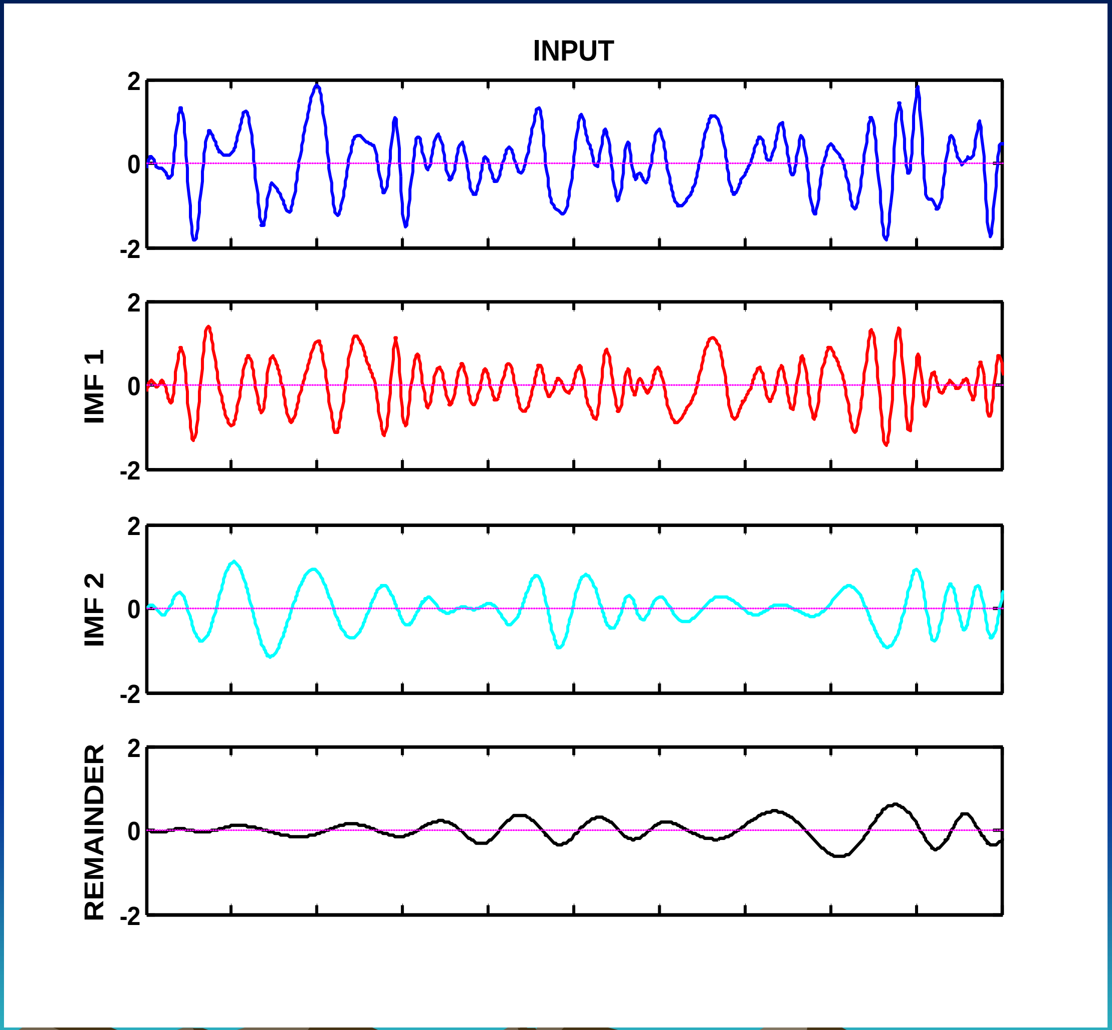
$$x(t) - m_1 = h_1\,,$$
$$h_1 - m_2 = \boxed{h_2}\,,$$
$$\cdots\cdots$$
$$\cdots\cdots$$
$$h_{k-1} - m_k = h_k\,.$$
$$\Rightarrow \quad h_k = c_1\,.$$

SIFTING



$$x(t) - m_1 = h_1 ,$$
$$h_1 - m_2 = h_2 ,$$
$$\cdots\cdots$$
$$\cdots\cdots$$
$$h_{k-1} - m_k = \boxed{h_k} .$$
$$\Rightarrow \quad h_k = c_1 .$$

SIFTING



$$\boxed{x(t)} - m_1 = h_1 \, ,$$

$$h_1 - m_2 = h_2 \, ,$$

$$\cdots\cdots$$

$$\cdots\cdots$$

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

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

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

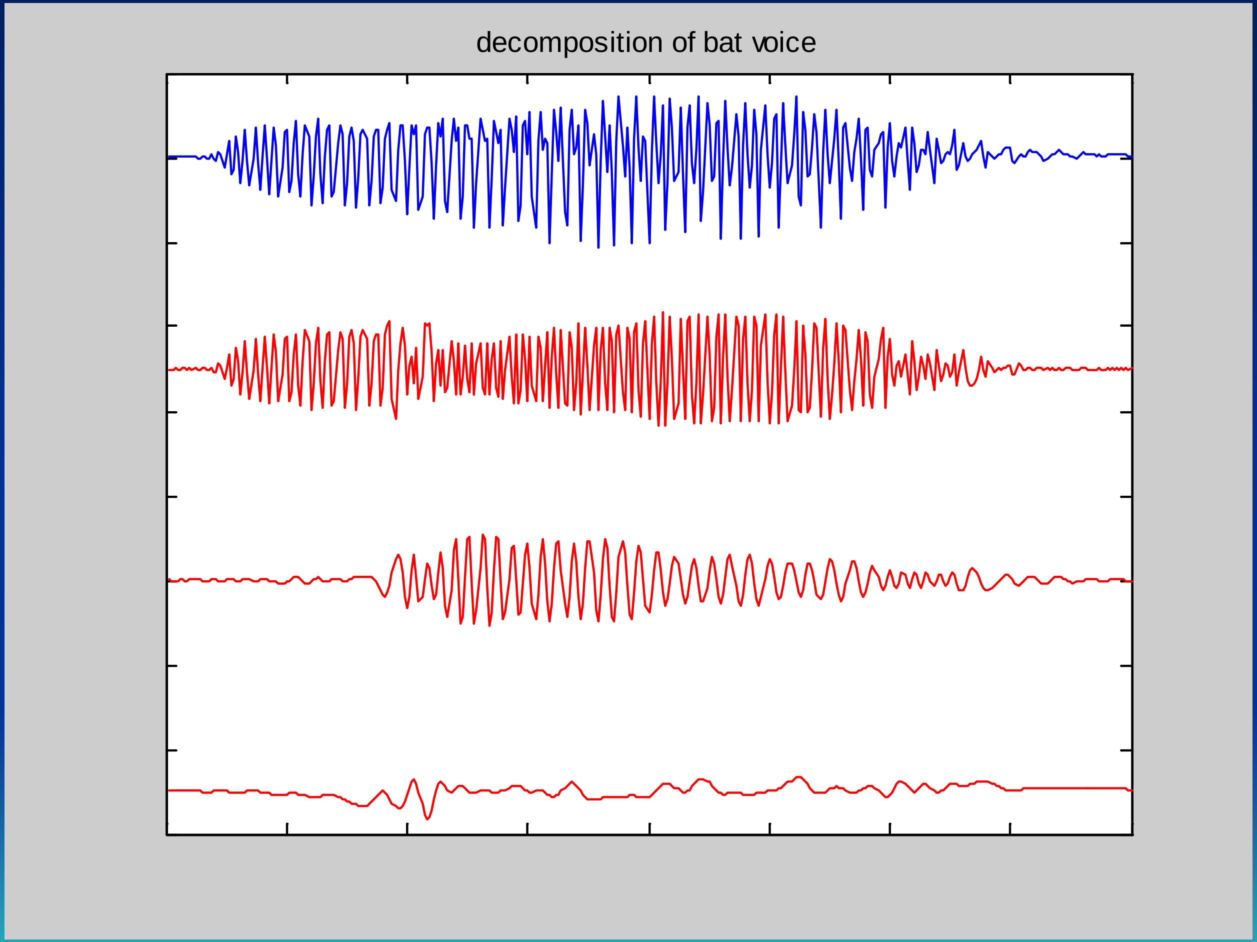
$$r_1 - c_2 = r_2 \, ,$$

$$\cdot \; \cdot \; \cdot$$

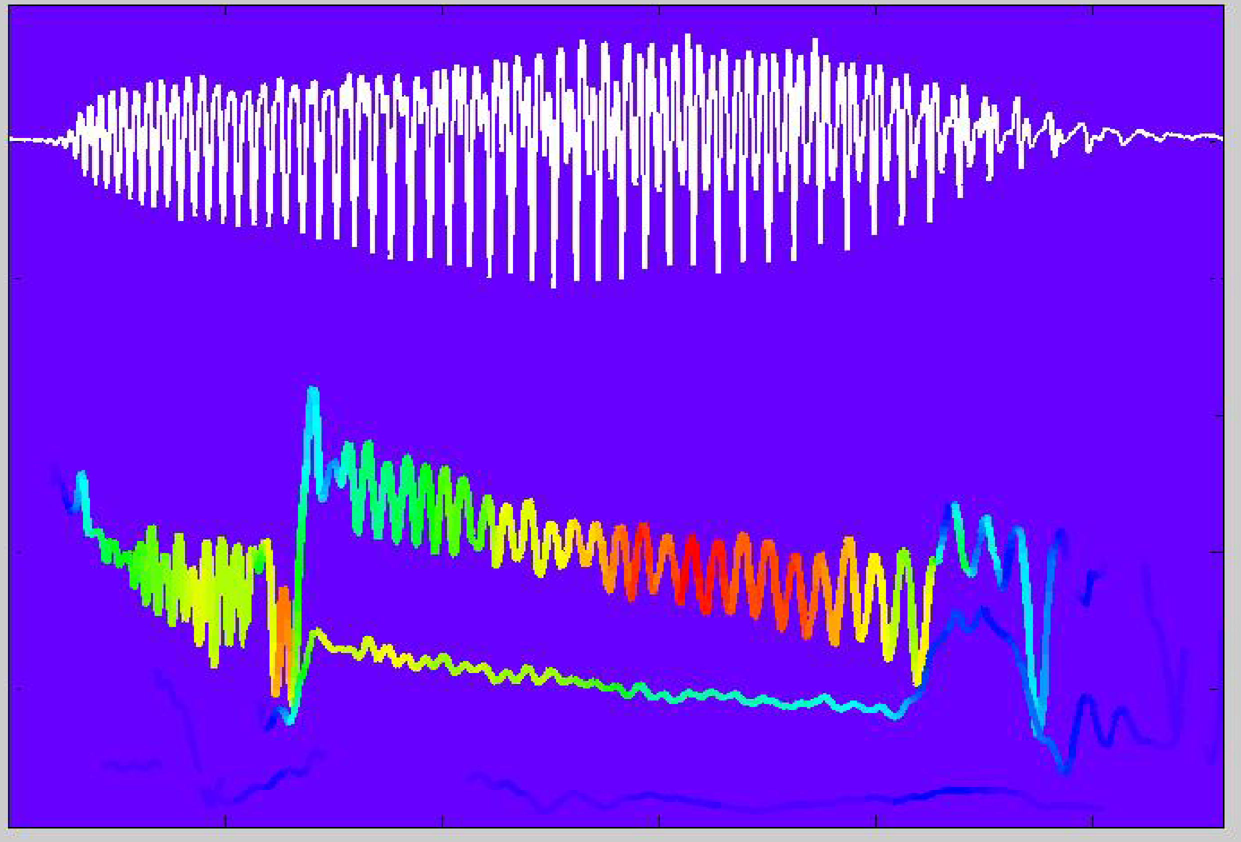
$$r_{n-1} - c_n = r_n \, .$$

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

DECOMPOSITION OF BAT VOICE



TIME-FREQUENCY-AMPLITUDE DIAGRAM OF BAT VOICE



SOME CRETERIA FOR DECOMPOSITION

- From Ingrid Daubechies (Dec. 2008)

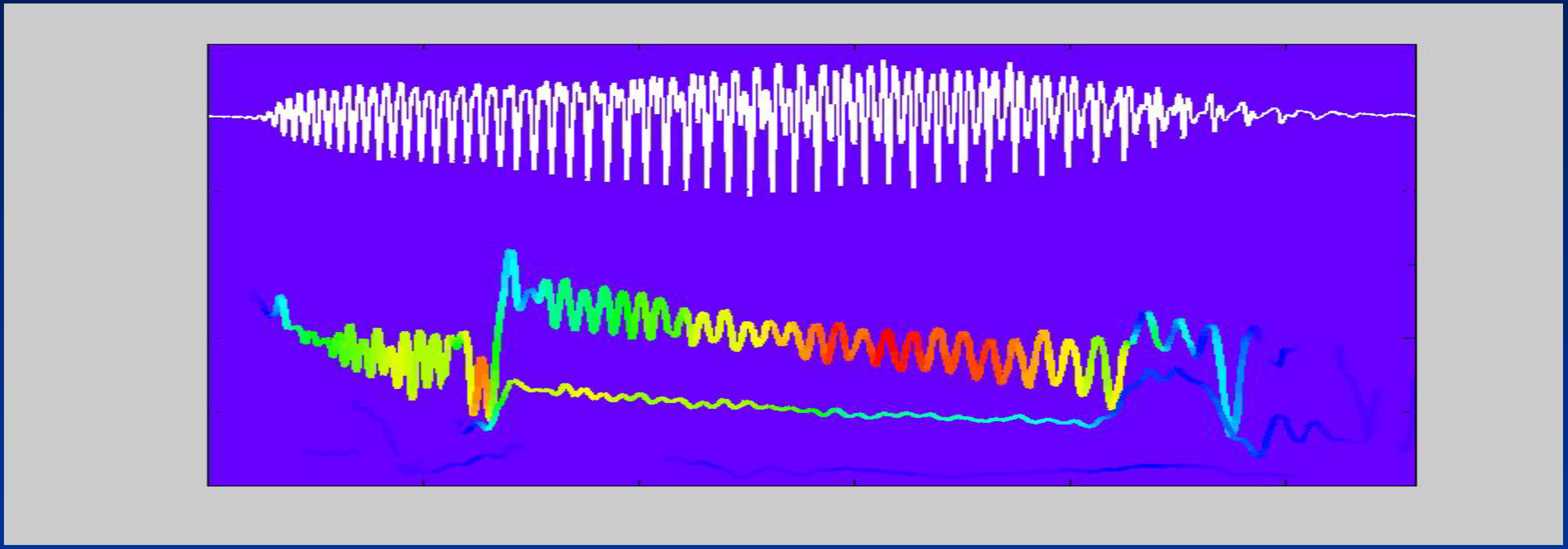
  - "natural" waveforms need not be of Fourier (or wavelet, or curvelet, ...) type
  - nevertheless, it can still be useful to decompose signals into more elementary waveforms
  - such a decomposition should be adaptive (of course)
  - should produce meaningful components
  - should also be robust to noise

SOME CRETERIA FOR DECOMPOSITION

- From Ingrid Daubechies (Dec. 2008)

- "natural" waveforms need not be of Fourier (or wavelet, or curvelet, ...) type
- nevertheless, it can still be useful to decompose signals into more elementary waveforms
- such a decomposition should be adaptive (of course)
- should produce meaningful components
- should also be robust to noise

PHYSICAL MEANING



Imagined Story (physical hypothesis to be tested):

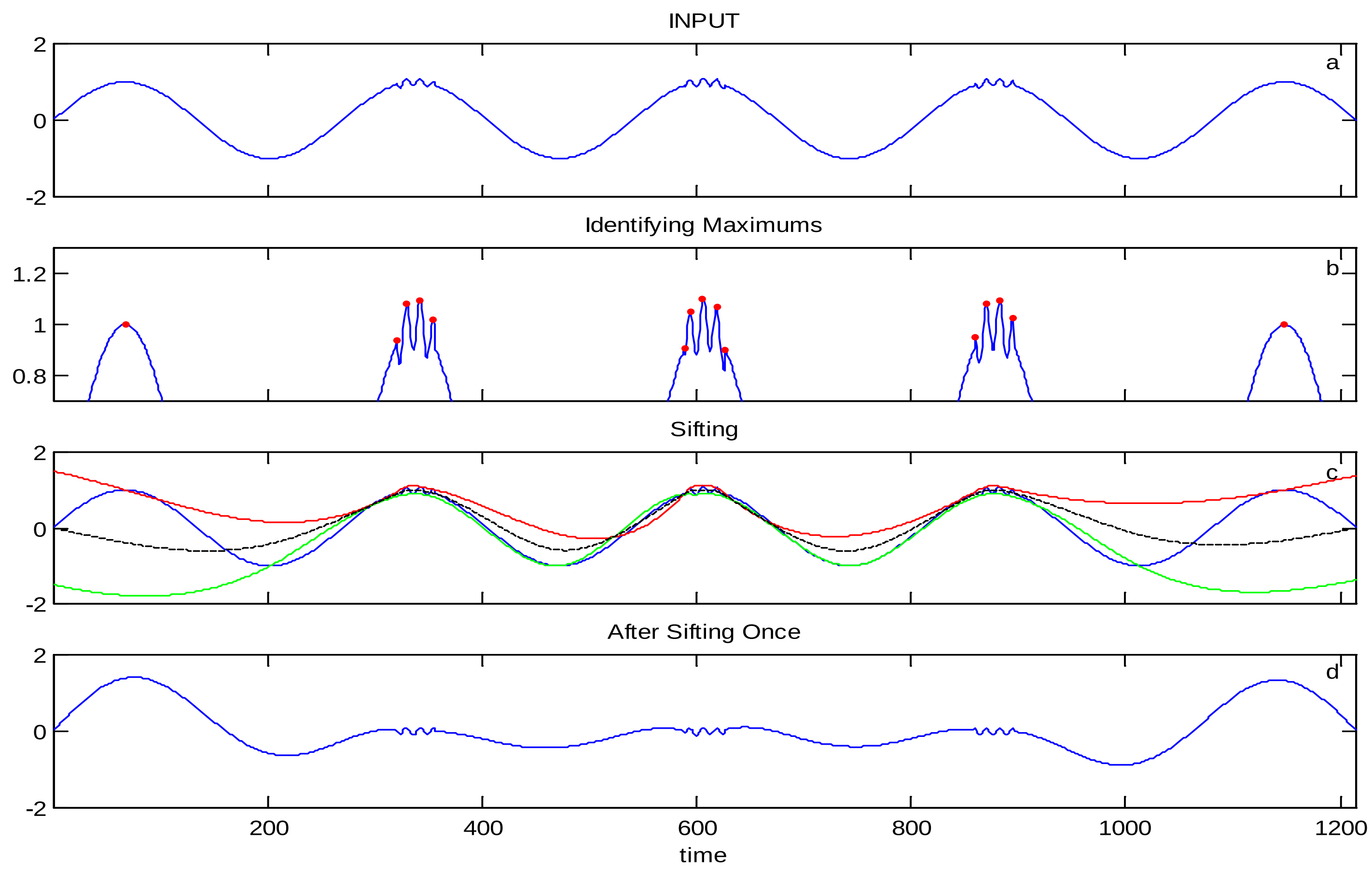
The brown bat first stretched tightly its vocal cord, and then let it relax. After some time, an extra "finger" touched the cord, like a violinist plays his/her music, and a second tone came. The vocal cord continued to relax and both tones shifted to lower frequency.

SOME CRETERIA FOR DECOMPOSITION

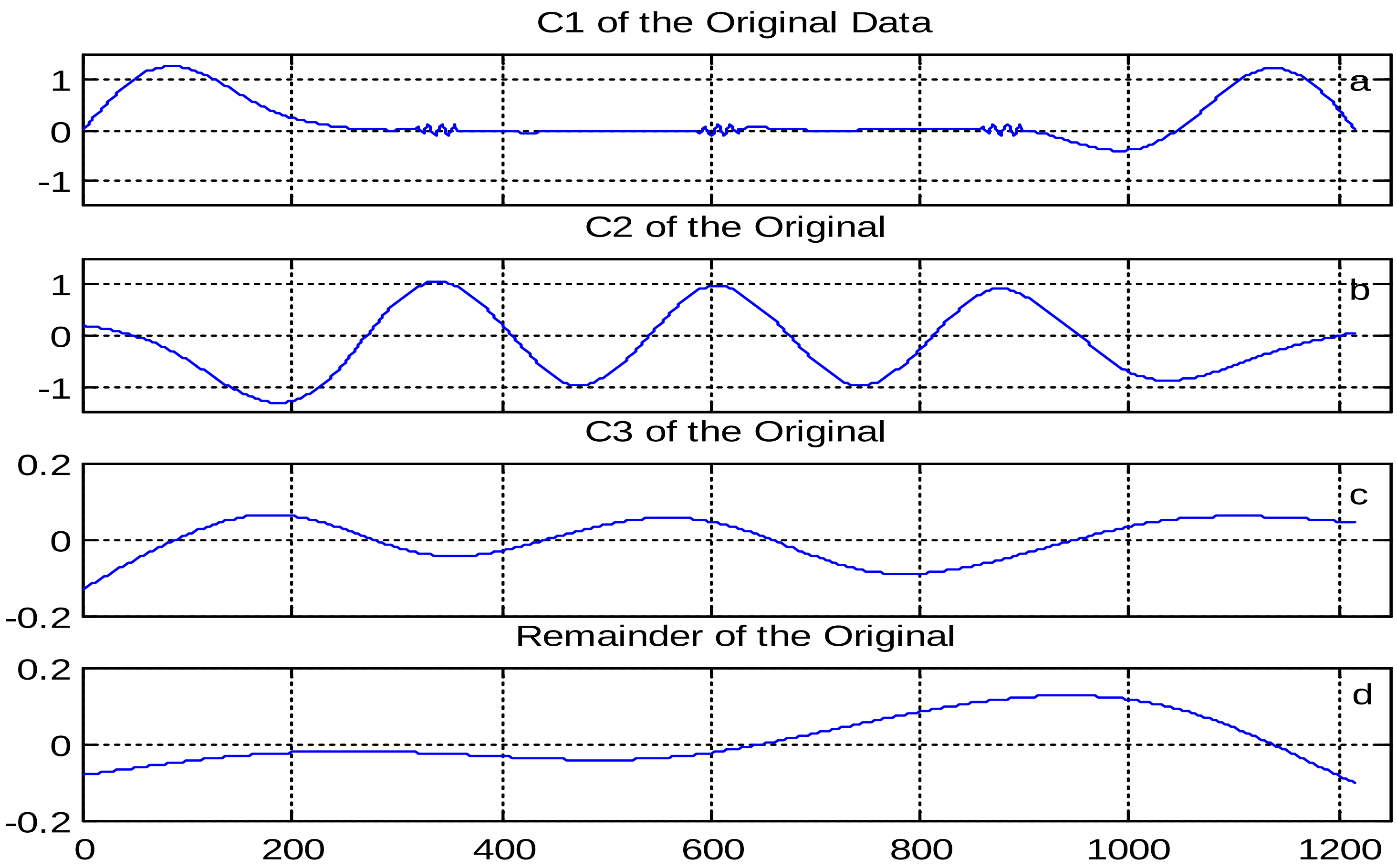
- From Ingrid Daubechies (Dec. 2008)

  - "natural" waveforms need not be of Fourier (or wavelet, or curvelet, ...) type
  - nevertheless, it can still be useful to decompose signals into more elementary waveforms
  - such a decomposition should be adaptive (of course)
  - should produce meaningful components
  - should also be robust to noise

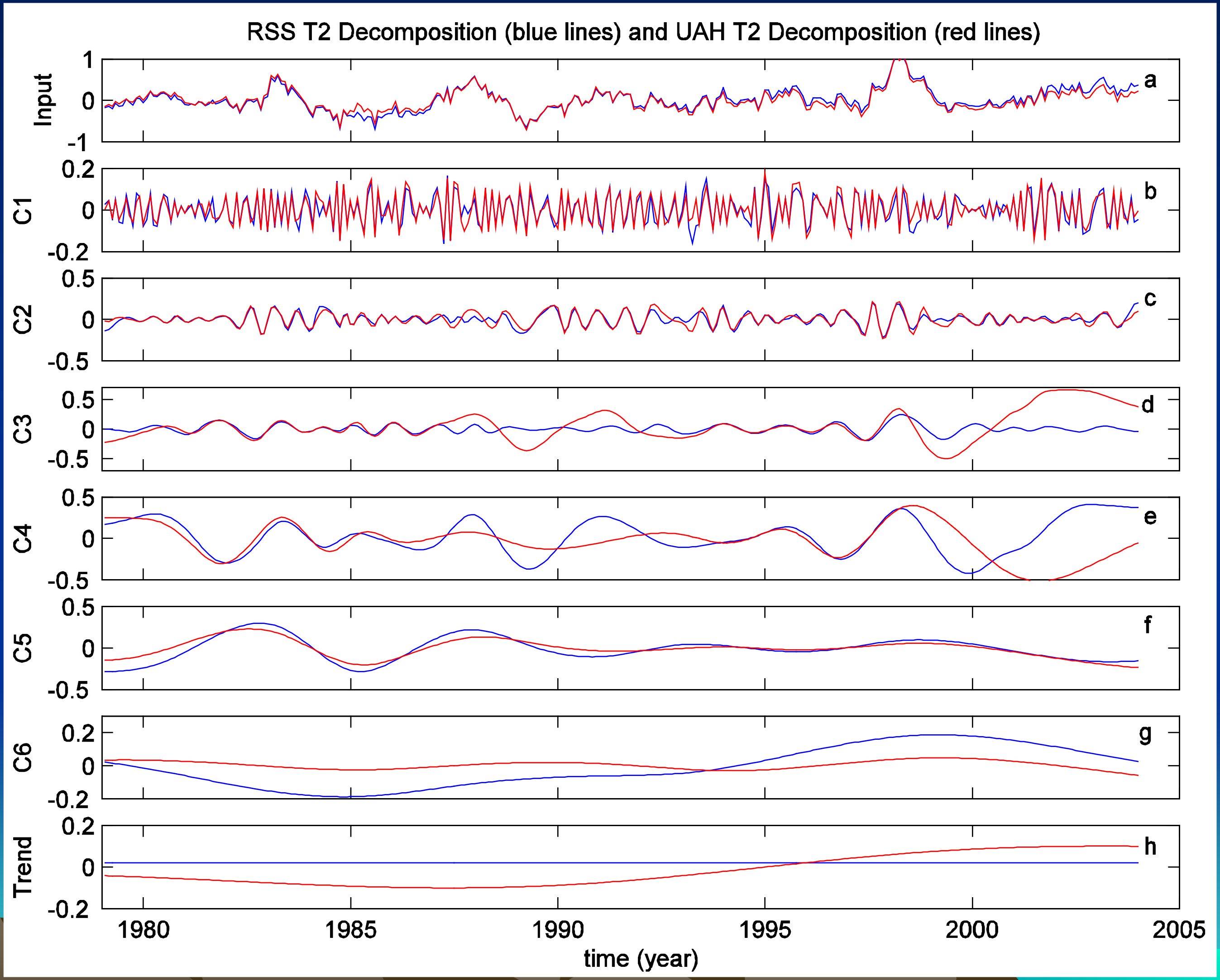
SCALE MIXING PROBLEM



SCALE MIXING PROBLEM



IMPLICATION OF SCALE MIXING



PHYSICAL UNIQUENESS

- The Physical Uniqueness (P-U)

  the decompositions of a data set and of the same data set with added noise perturbation of small but not infinitesimal amplitude bear little quantitative and no qualitative change

- Does P-U Matter in Data Analysis?
  – Yes, since a data set from real world always contains random noise

- Does a method currently available satisfy P-U
  – Fourier Transform does
  – Wavelet decomposition does
  – EMD often does not, and is not stable and hard to interpret

AGAIN, WHAT IS DATA ?

- Definition
  - A _collection_ or _representation_ of _facts_, concepts, or instructions in a manner suitable for communication, interpretation, analysis, or processing

data = _facts_ + _distortion_

$$X(t) = S(t) + N(t)$$

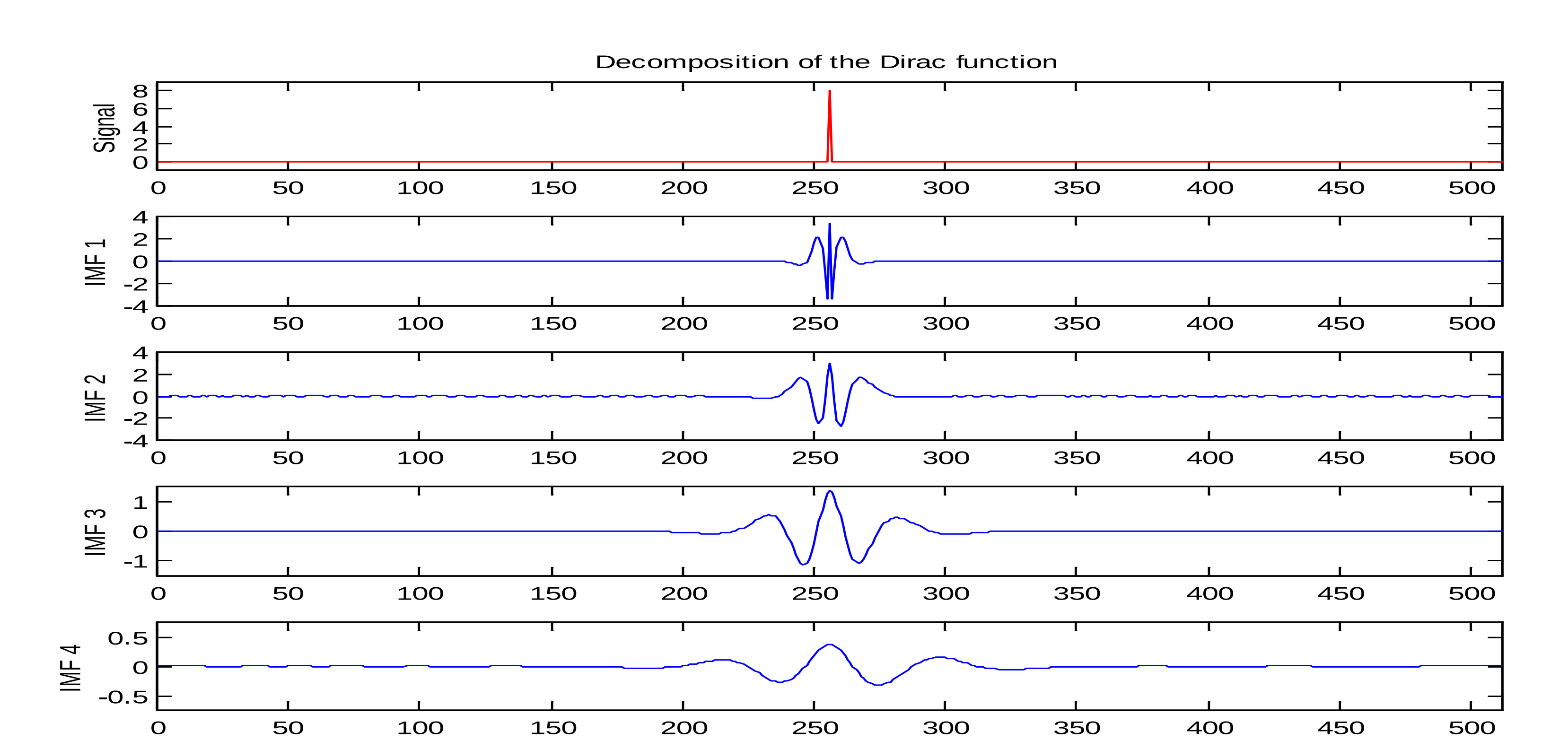
- Observations
  - Observation I

$$X_1(t) = X(t) + N_1(t)$$

  - Observation II

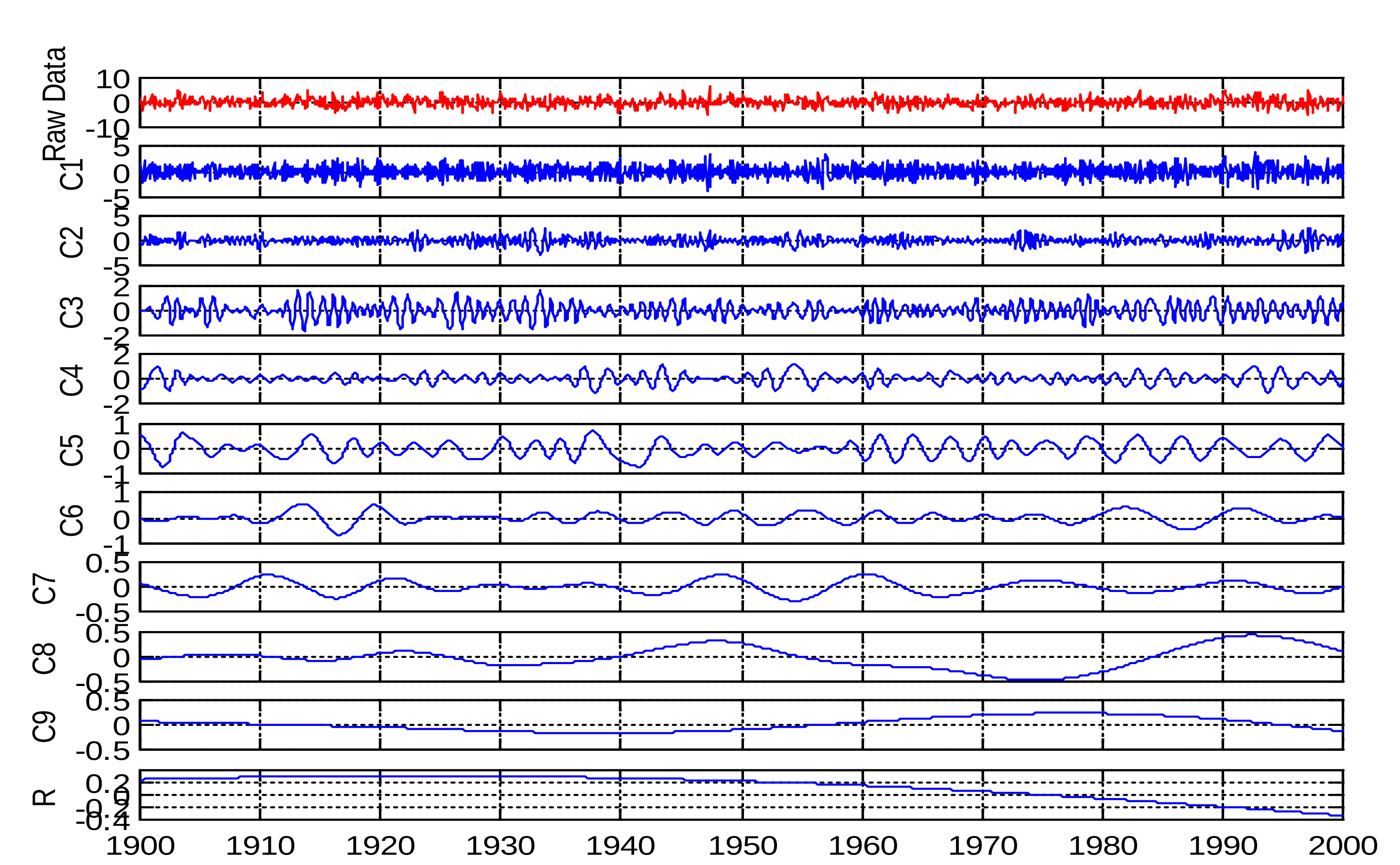
$$X_2(t) = X(t) + N_2(t)$$

DECOMPOSITION OF DIRAC DELTA DUNCTION



EMD is, in this case, an adaptive wavelet.

DECOMPOSITION OF NOISE



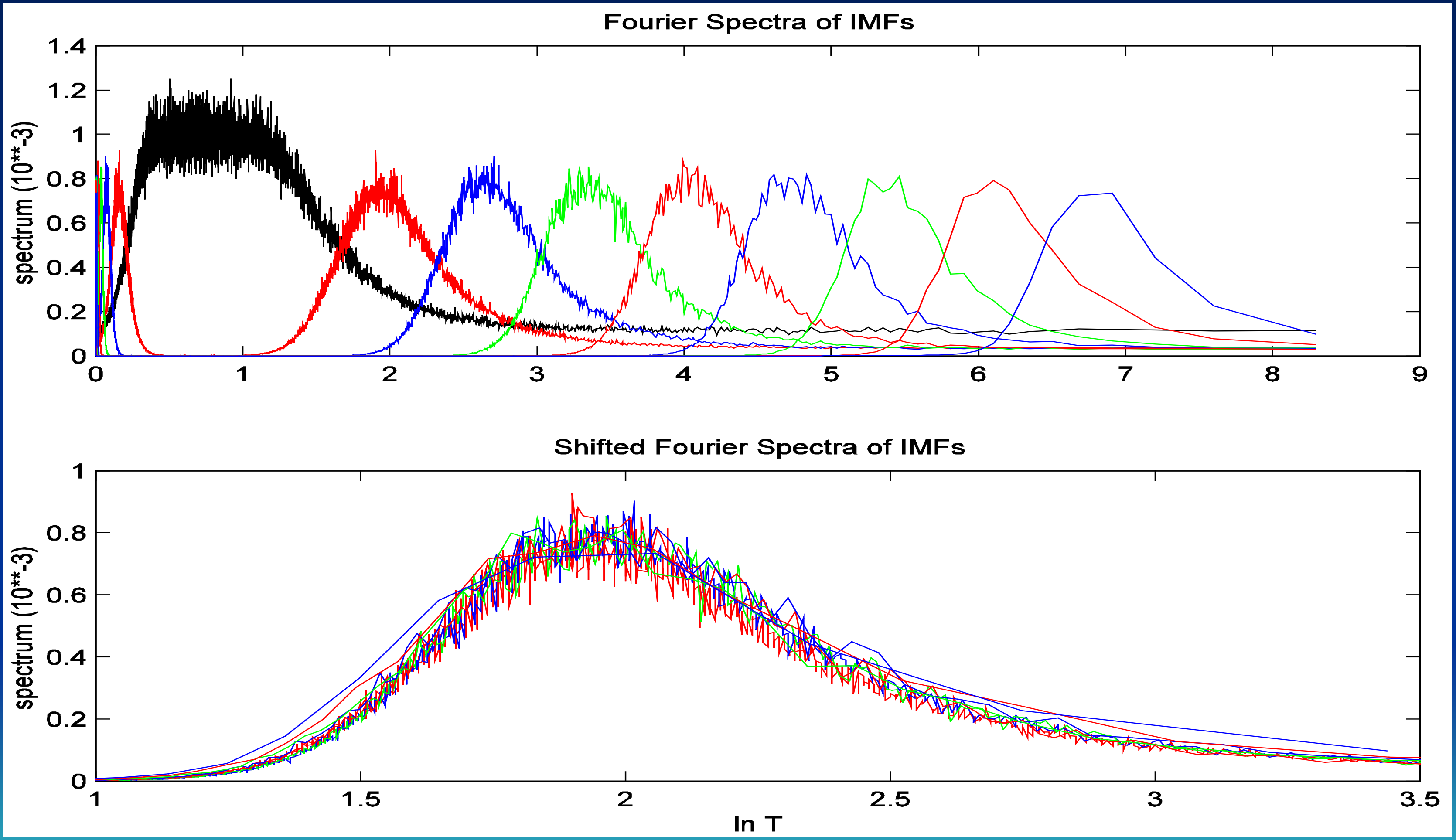
PERIODS OF WHITE NOISE COMPONENTS

- A Million Data Points

Mode		1	2	3	4	5	6	7	8	9
White Noise	# peaks	347042	168176	83456	41632	20877	10471	5290	2658	1348
	period	2.881	5.946	11.98	24.02	47.90	95.50	189.0	376.2	741.8

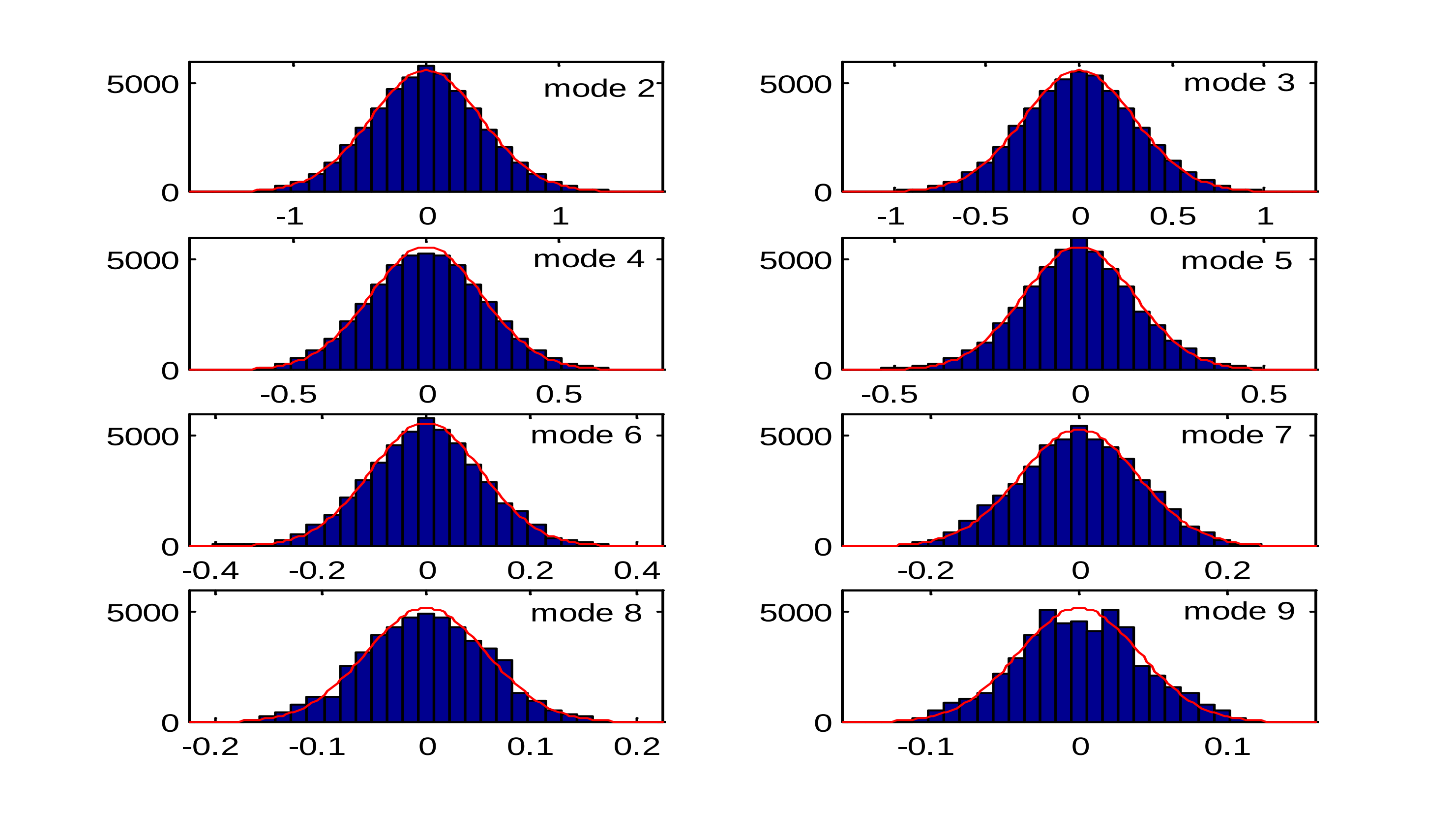
Period Doubling !!!!!

FOURIER SPECTRA OF IMFs



DISTRIBUTIONS OF IMFs

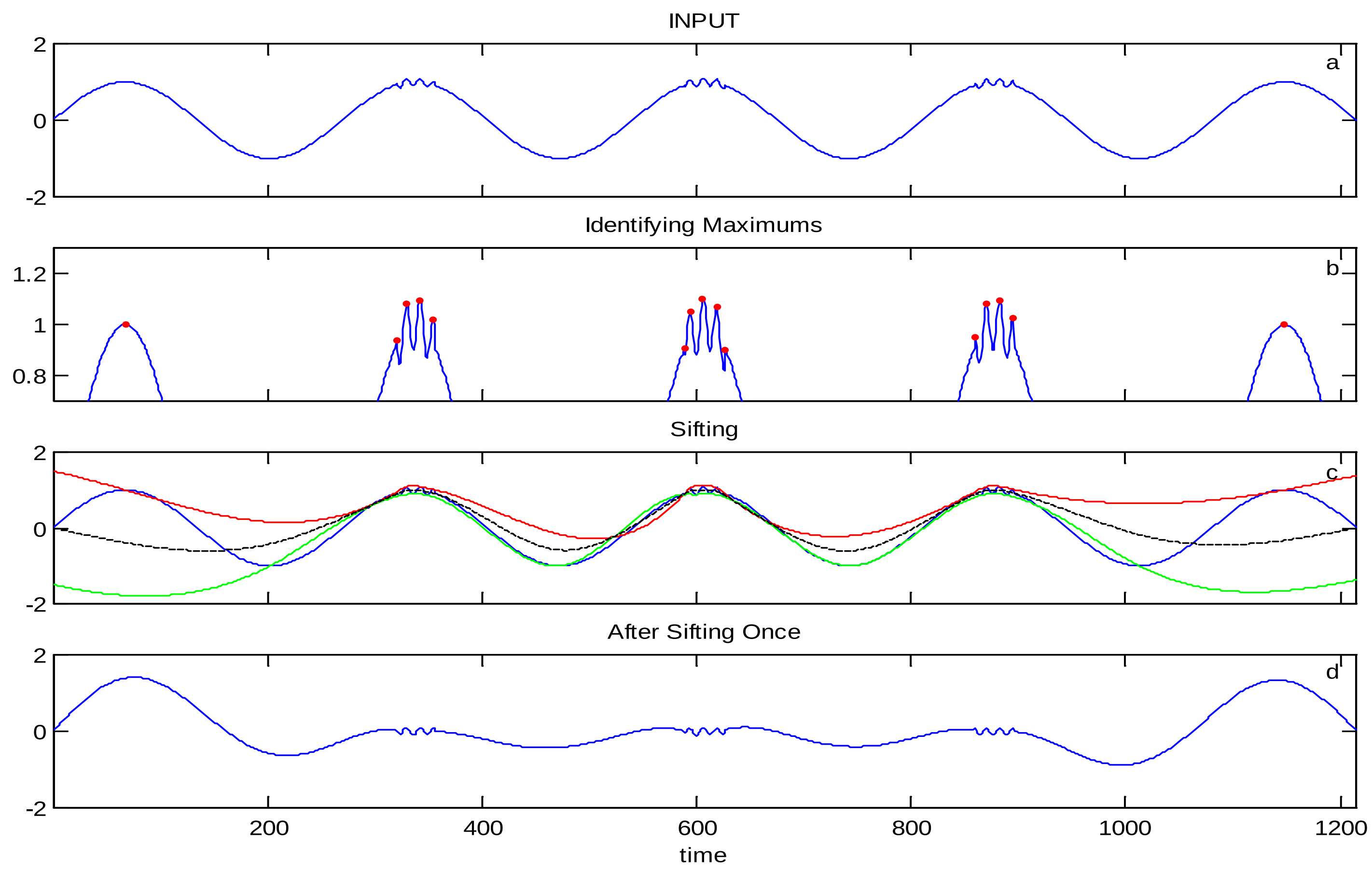
- Normal Distribution



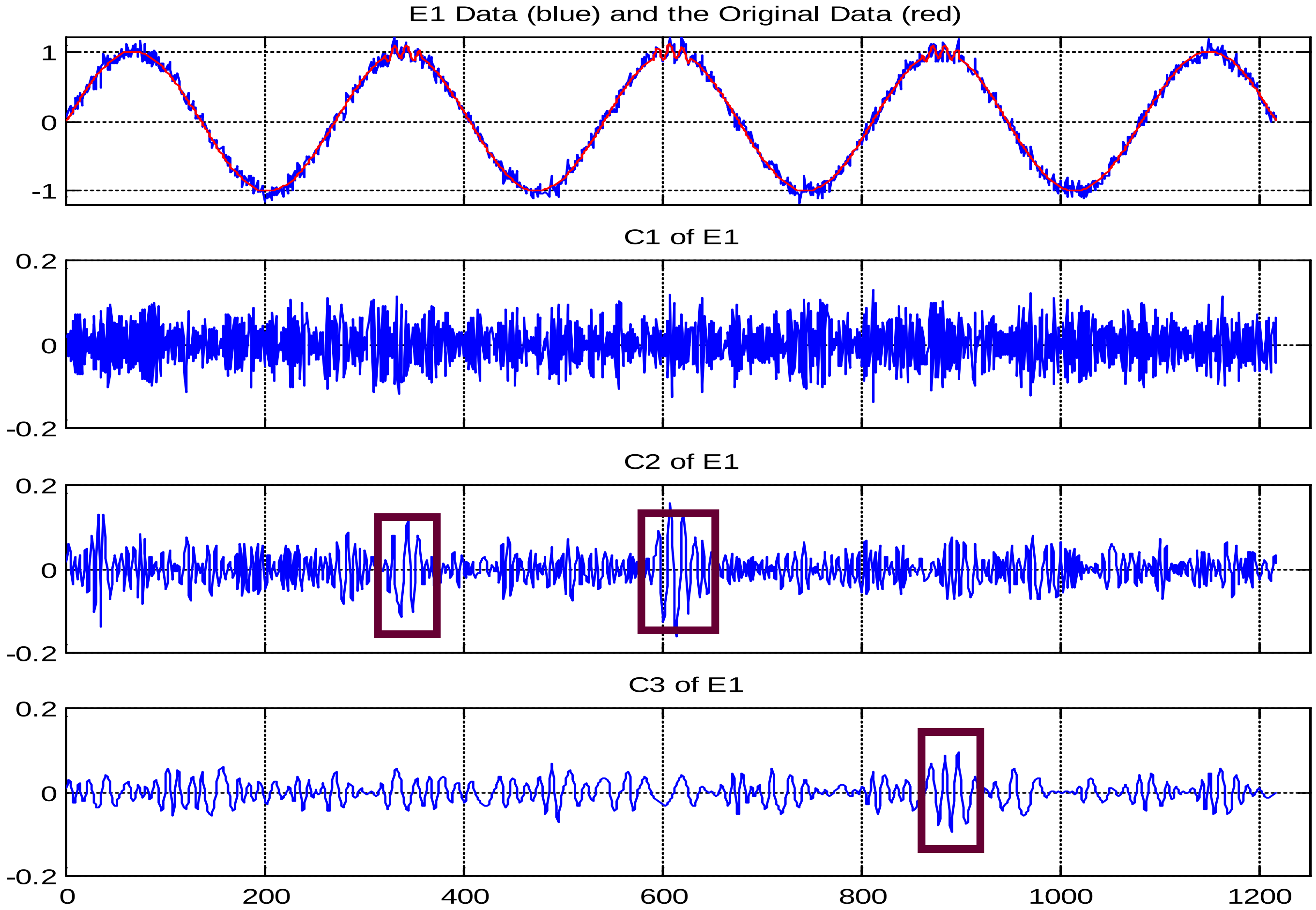
USE NOISE TO ASSISTS DECOMPOSITION

- Two qualities
  - The true signals in data should not be affected by the observations
  - White noise, as a dyadic filter bank in EMD, should provide some control of the width of spectral window of real data decomposition, consequently robust decomposition
- One wishful thinking
  - *adding noise* to the targeted data during data analysis could be helpful — Noise-Assisted Data Analysis (NADA) **?!**

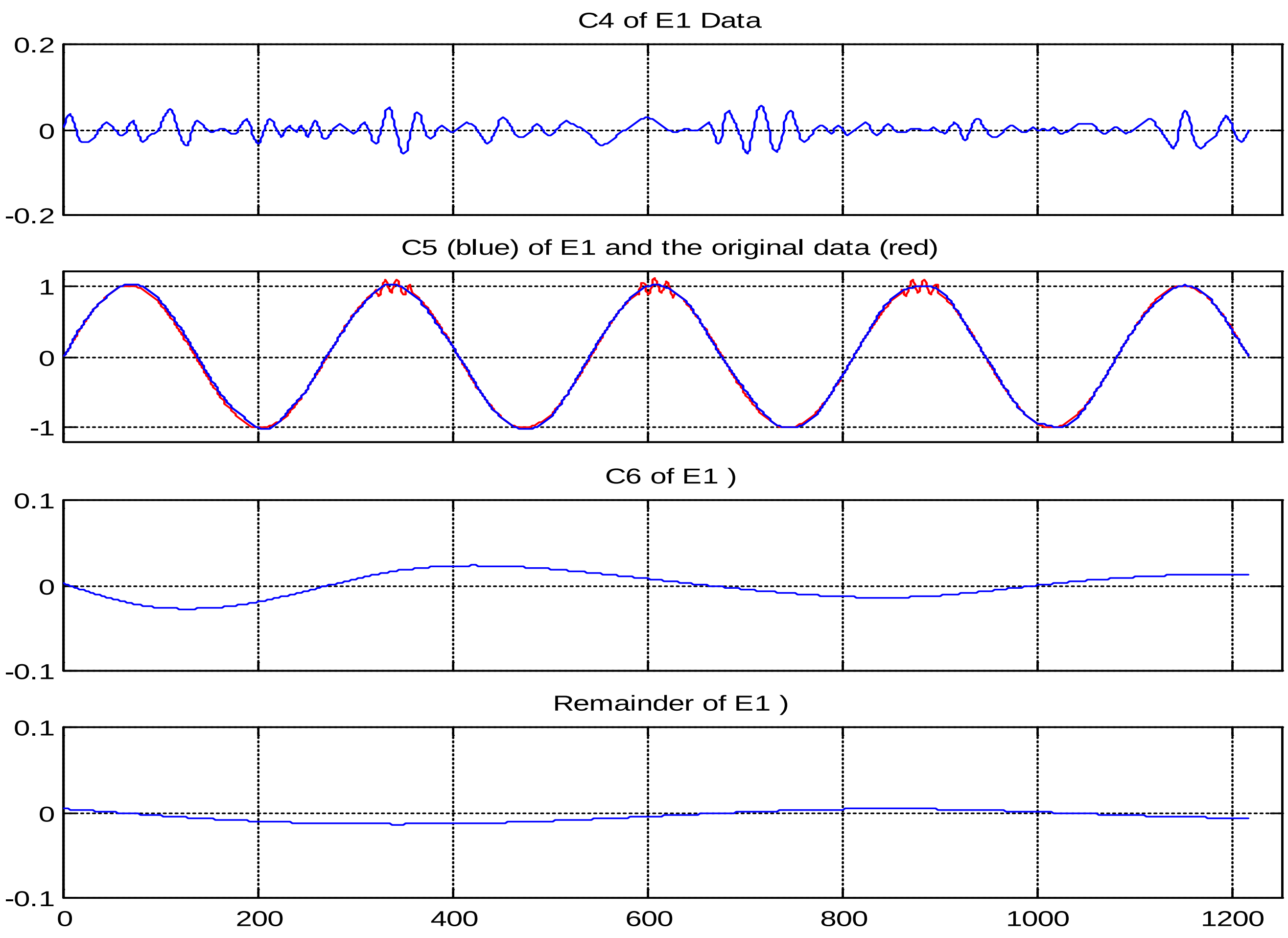
SCALE MIXING PROBLEM



NADA — PRELIMINARY TEST (I)



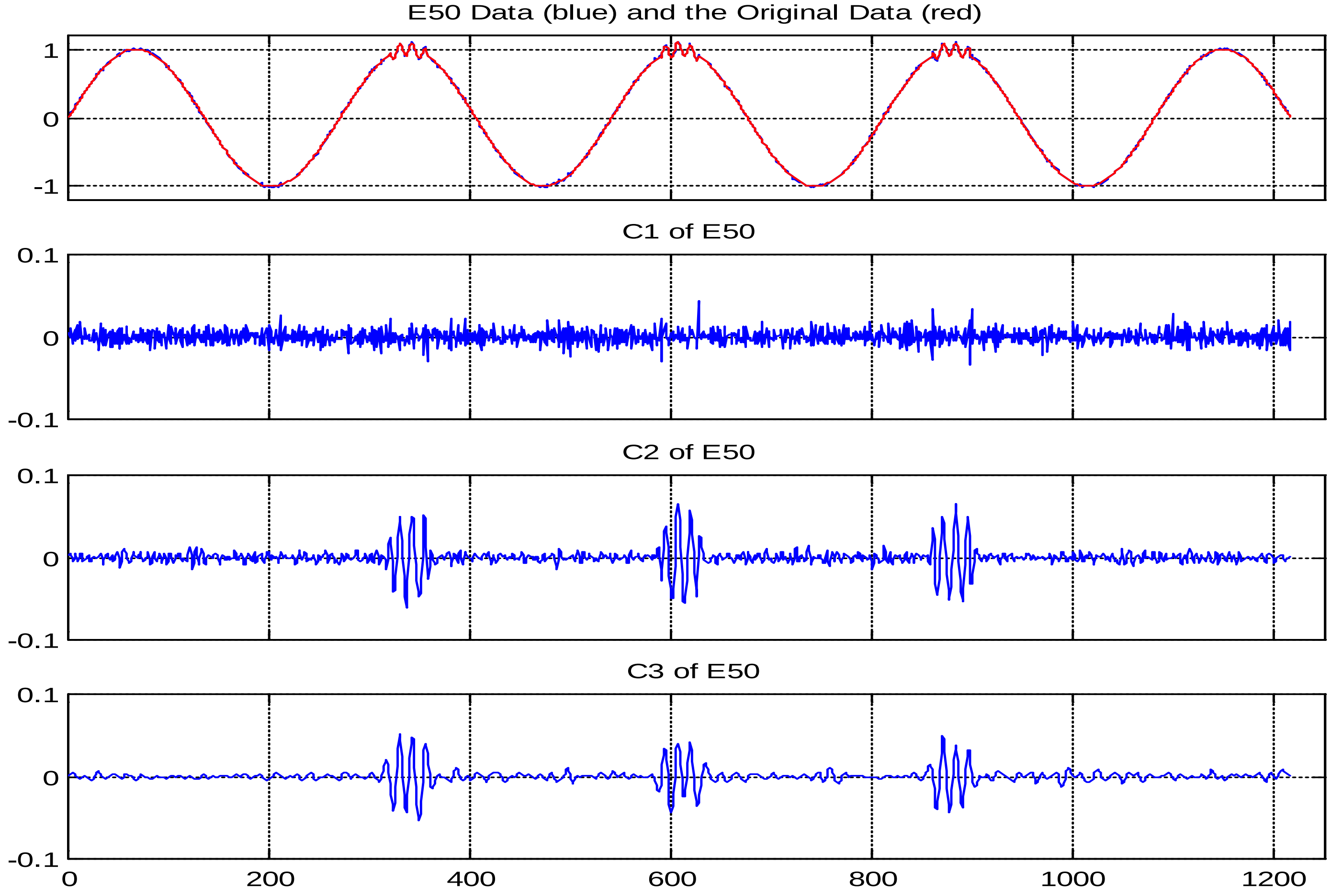
NADA — PRELIMINARY TEST (II)



NOISE-ASSISTED DATA ANALYSIS

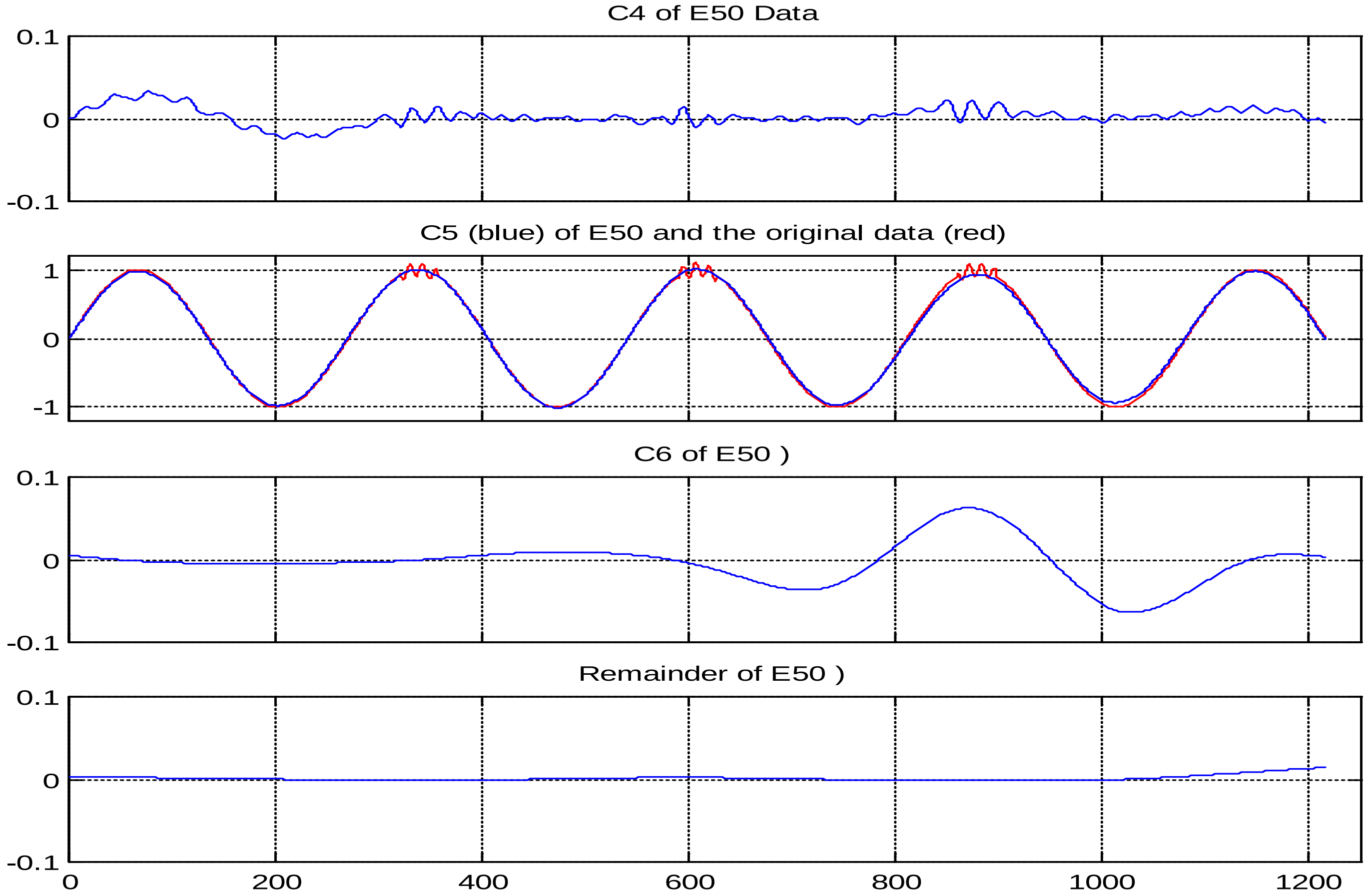
- **Ensemble EMD**
  - STEP 1: add a noise series to the targeted data
  - STEP 2: decompose the data with added noise into IMFs
  - STEP 3: *repeat STEP 1 and STEP 2 again and again*, but with different noise series each time
  - STEP 4: obtain the (ensemble) means of corresponding IMFs of the decompositions as the final result
- **Effects**
  - In the mean IMFs, the added noise canceled with each other
  - The mean IMFs stays within the natural filter period windows (significantly reducing the chance of scale mixing and preserving dyadic property)

EEMD — NADA (I)



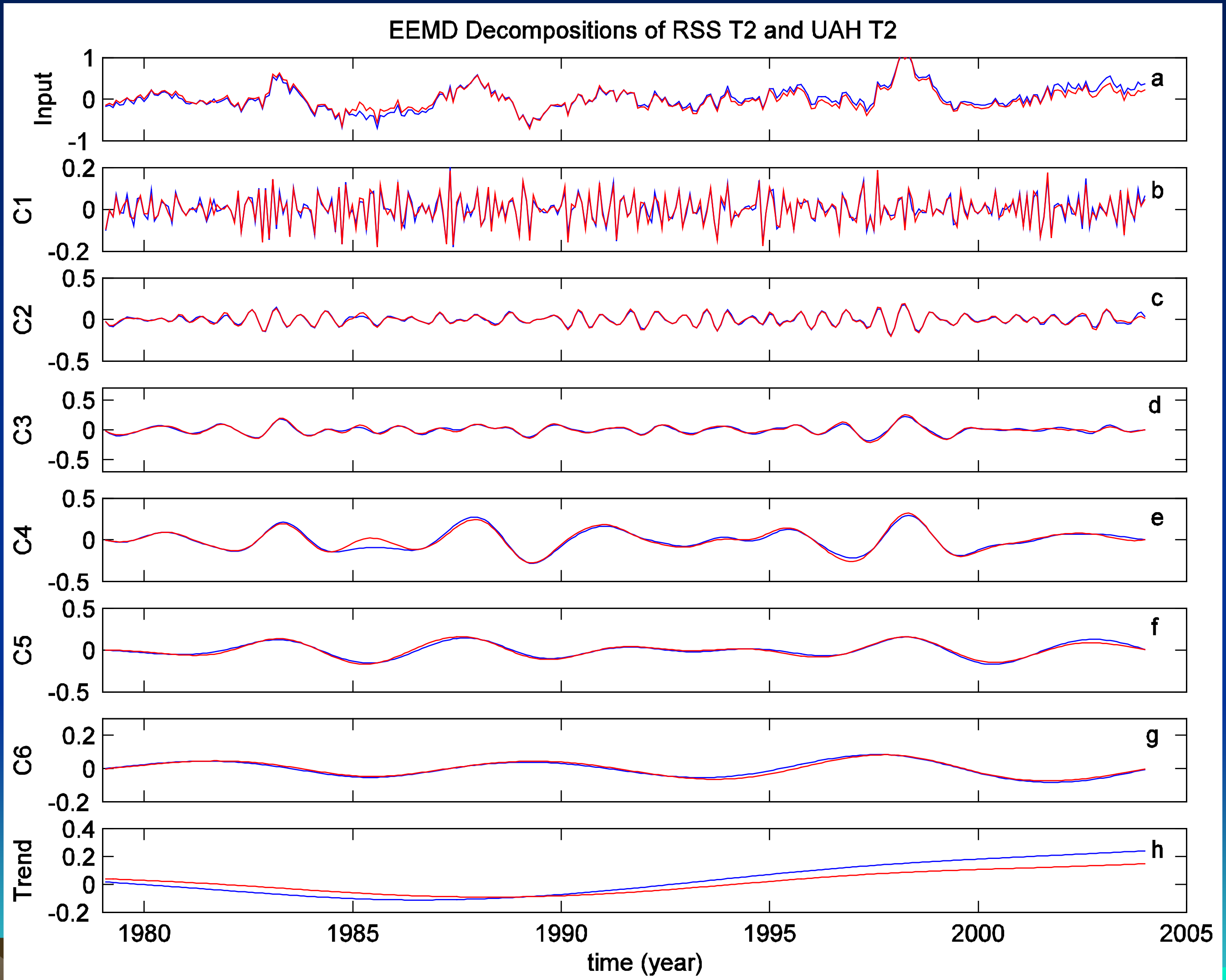
E50 Data (blue) and the Original Data (red)
1
0
-1
C1 of E50
0.1
0
-0.1
C2 of E50
0.1
0
-0.1
C3 of E50
0.1
0
-0.1
0
200
400
600
800
1000
1200

EEMD — NADA (II)

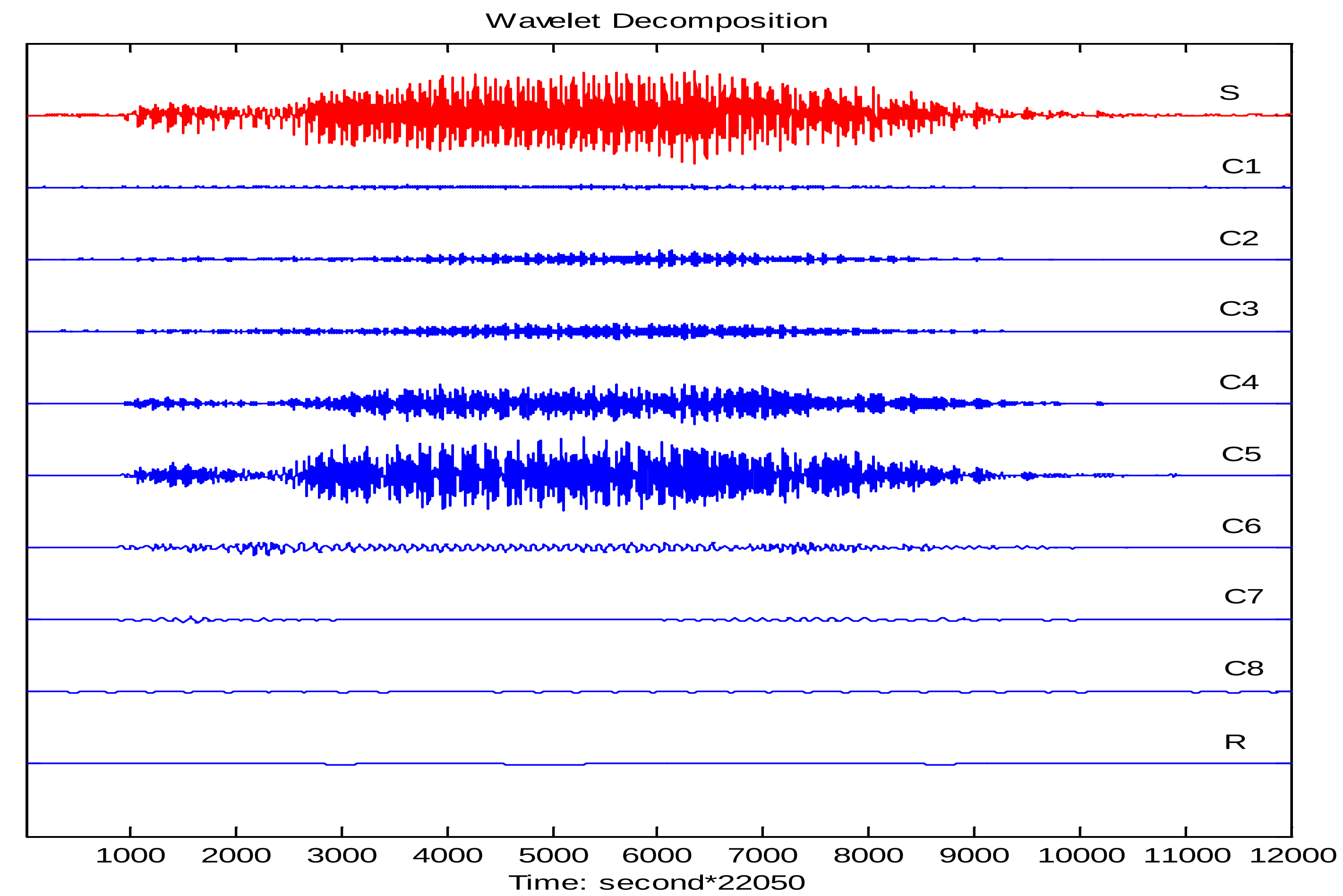


C4 of E50 Data
0.1
0
-0.1
C5 (blue) of E50 and the original data (red)
1
0
-1
C6 of E50)
0.1
0
-0.1
Remainder of E50)
0.1
0
-0.1
0
200
400
600
800
1000
1200

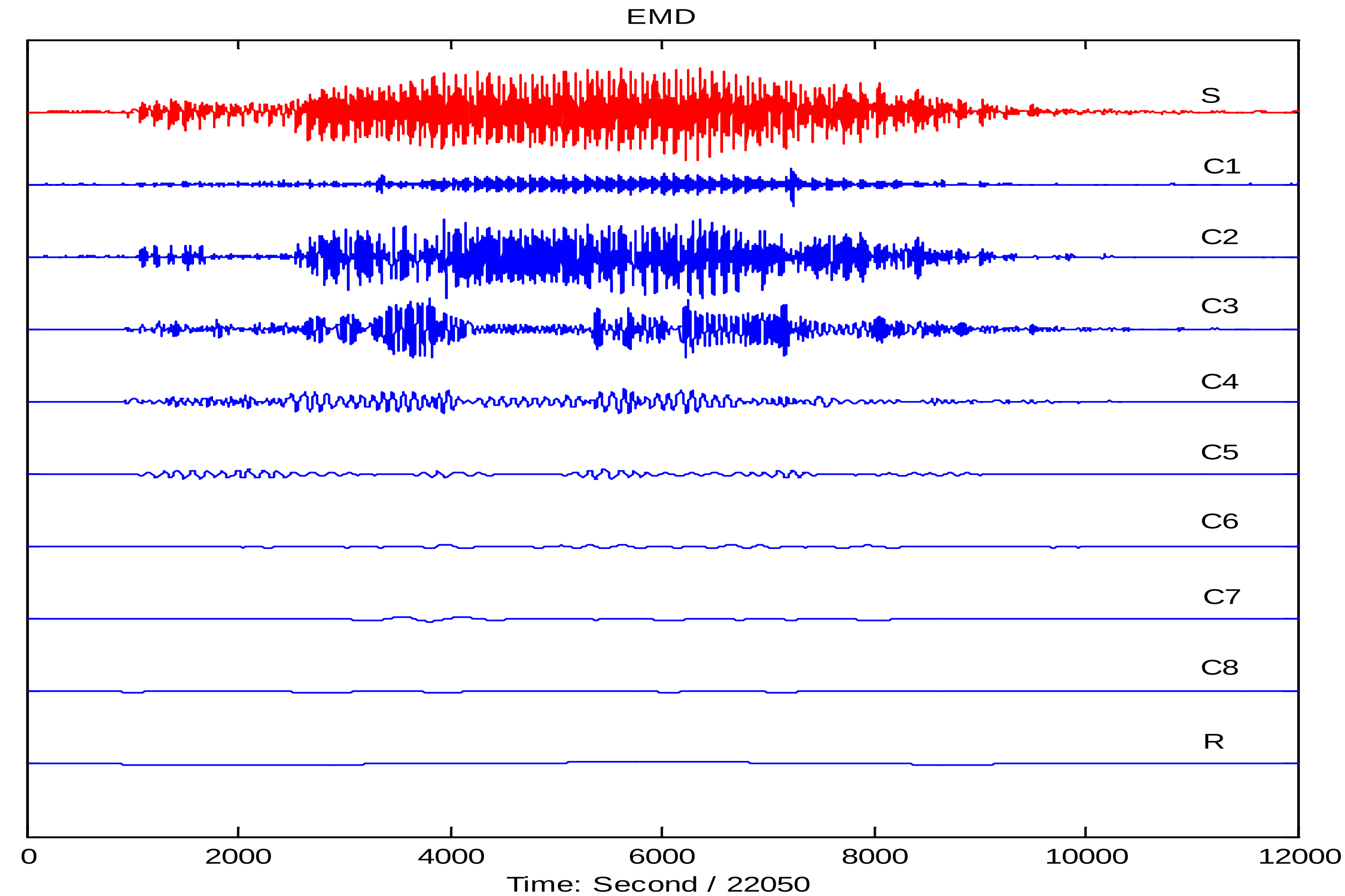
DEMONSTRATION OF STABILITY



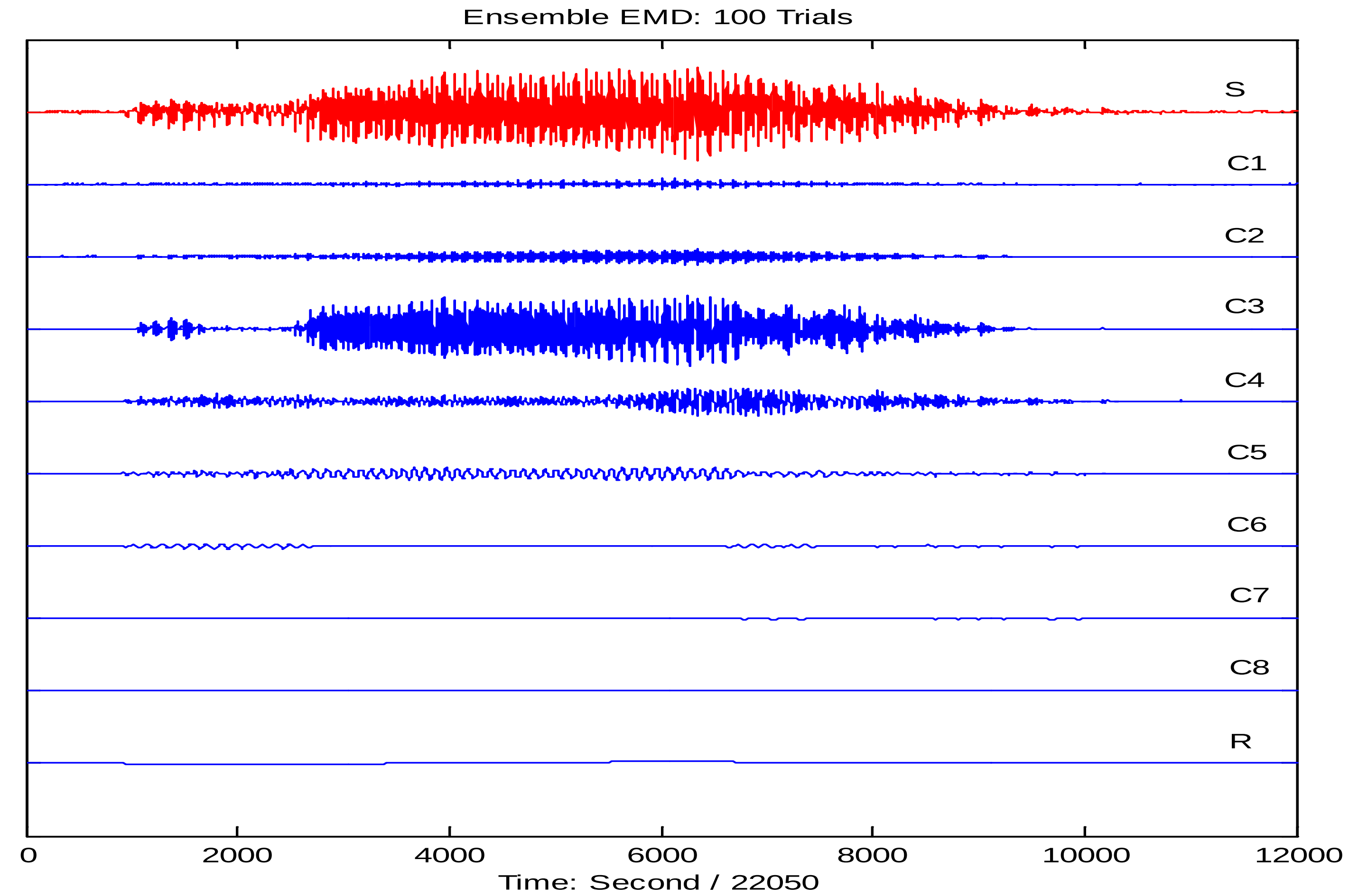
VOICE: WPD


Wavelet Decomposition
S
C1
C2
C3
C4
C5
C6
C7
C8
R
1000 2000 3000 4000 5000 6000 7000 8000 9000 10000 11000 12000
Time: second*22050

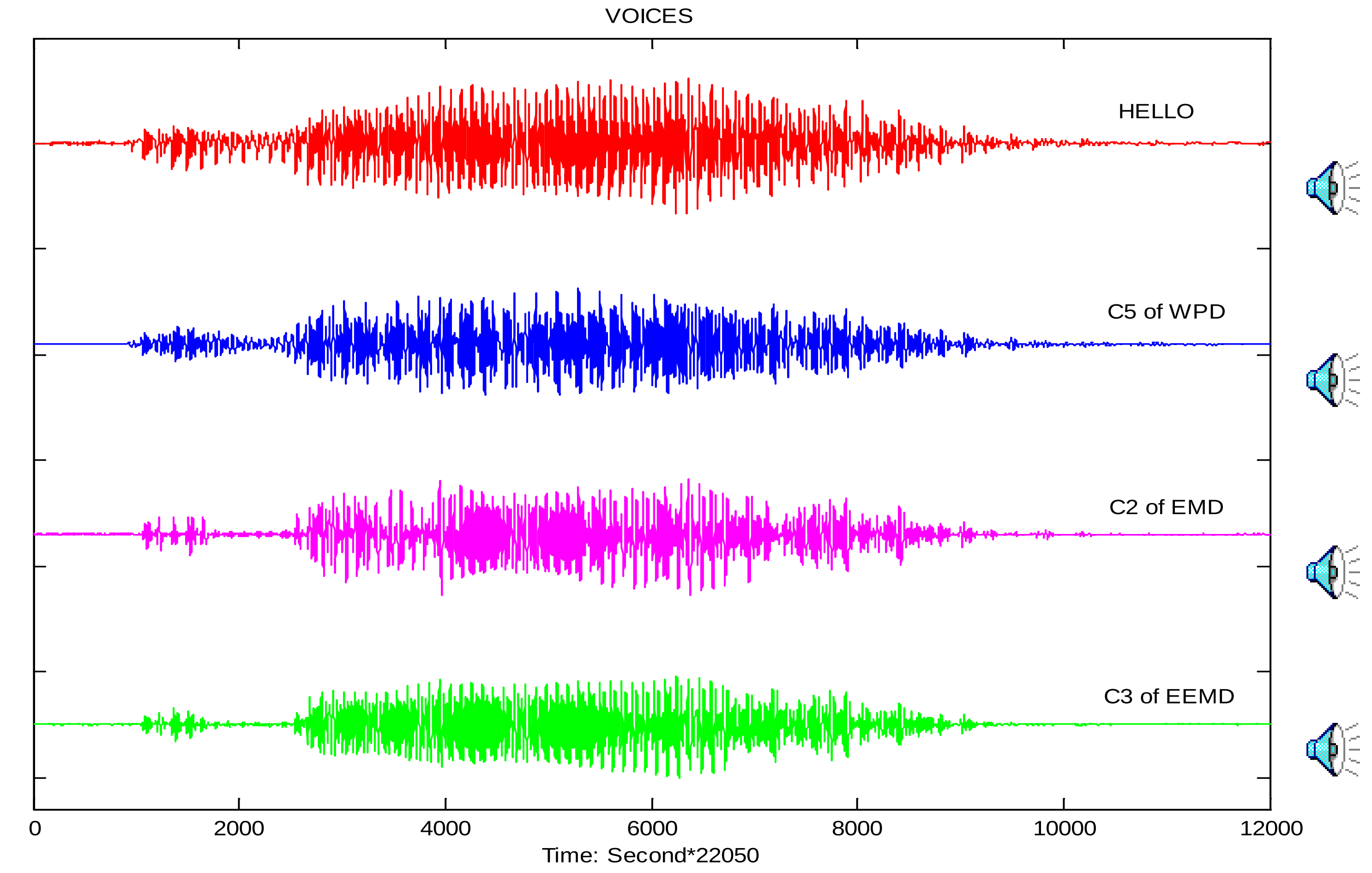
VOICE: EMD


EMD
S
C1
C2
C3
C4
C5
C6
C7
C8
R
0
2000
4000
6000
8000
10000
12000
Time: Second / 22050

VOICE: EEMD

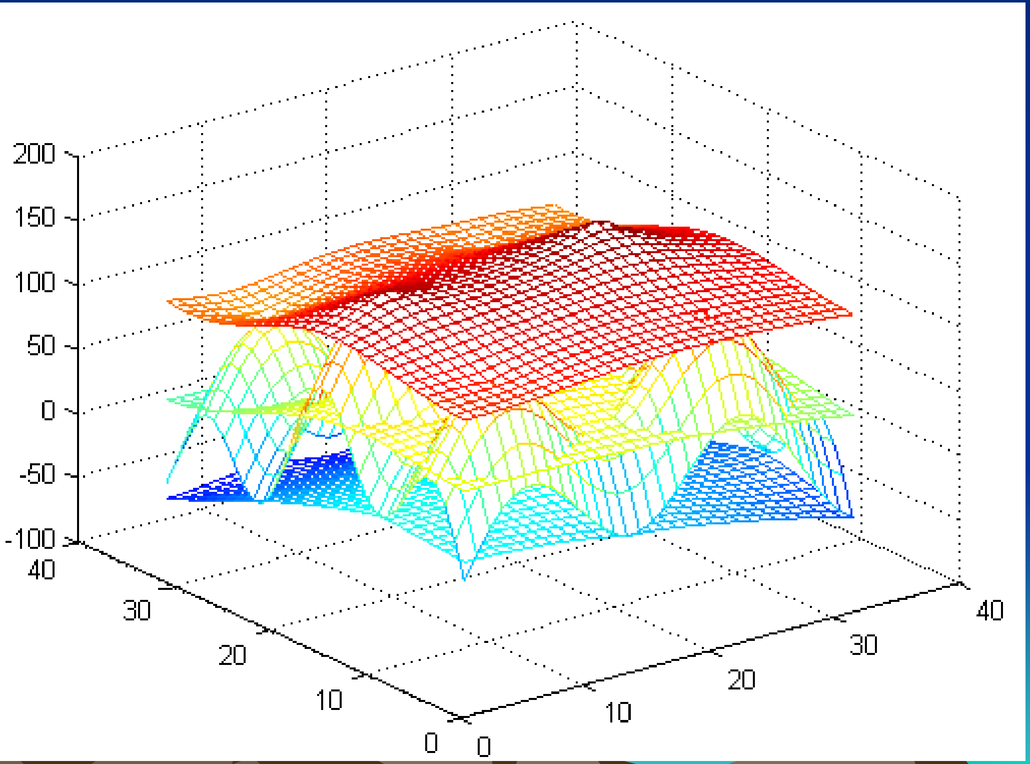

Ensemble EMD: 100 Trials
S
C1
C2
C3
C4
C5
C6
C7
C8
R
0
2000
4000
6000
8000
10000
12000
Time: Second / 22050

VOICES


VOICES
HELLO
C5 of WPD
C2 of EMD
C3 of EEMD
0
2000
4000
6000
8000
10000
12000
Time: Second*22050

EXTENSION TO 2D

- Find extrema in 2D
- Replace 1D envelopes with 2D membranes



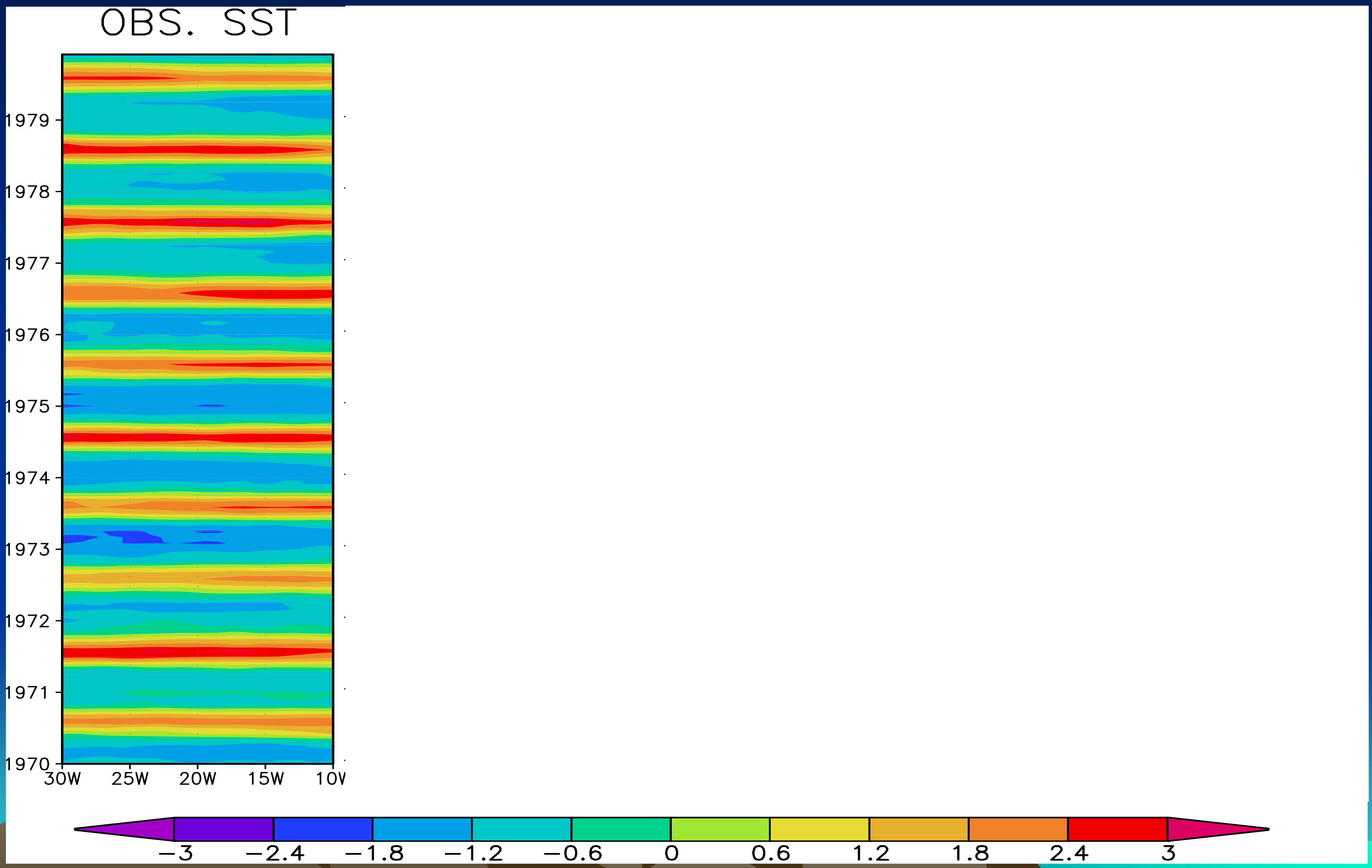
DIFFICULTIES IN DIRECT EXTENSION

- 2D: extrema to membranes
  - Is a saddle point a maximum or minimum?
  - Should the ridge (trough) be considered a line of maximum (minimum)
  - Scale mixing problem
  - Scale unconvertible in temporal-spatial data
- 3D: …
  - No known surface fitting

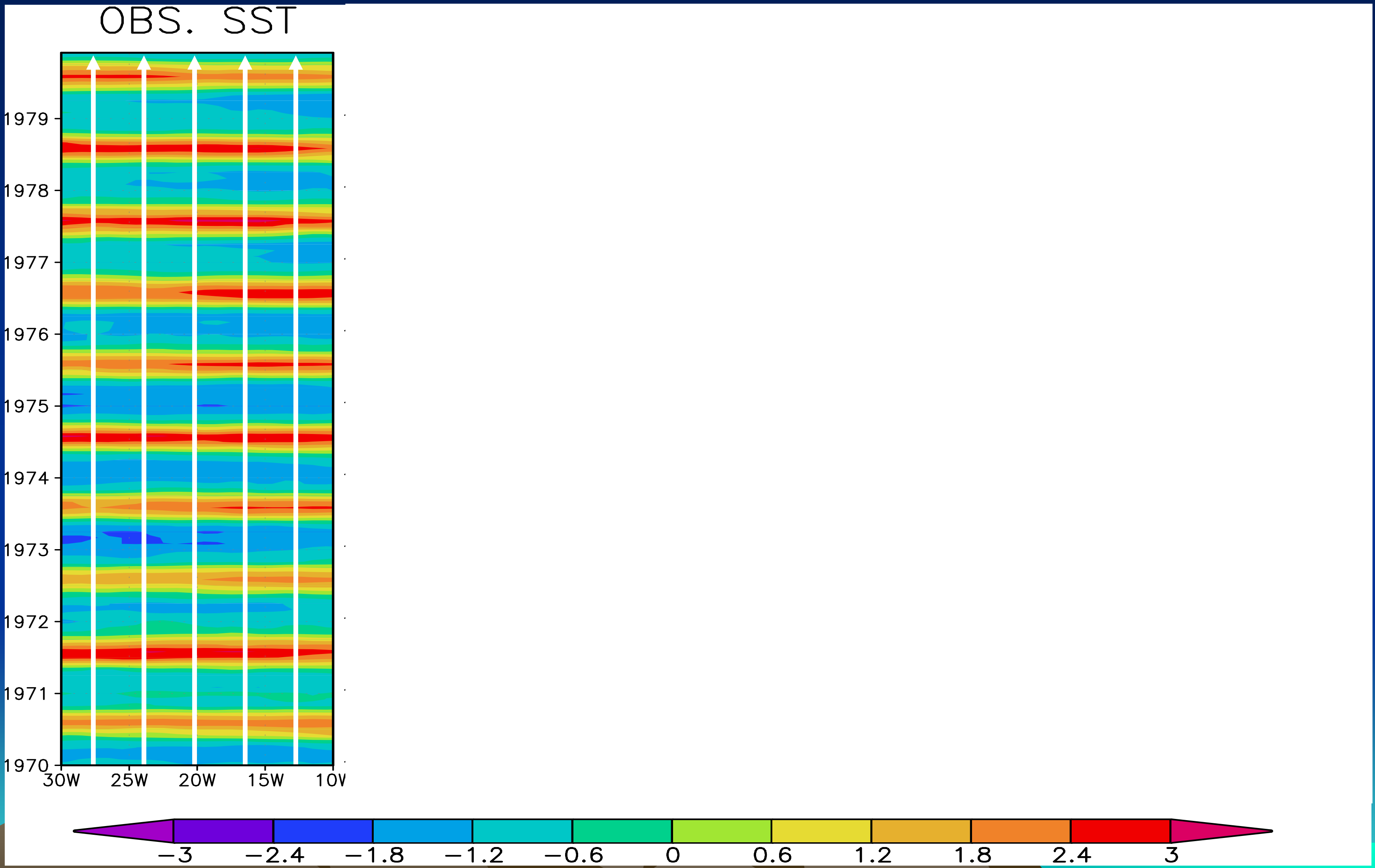
DIRECT 2D EEMD

- 2D EEMD
  - Add noise to 2D data, and decompose noise added 2D data
  - Repeat the processes many times
  - Taking ensemble mean
- Problems
  - Computationally demanding
  - Extrema definition difficulty remains
  - Extension to multi-dimensional EMD unimaginable

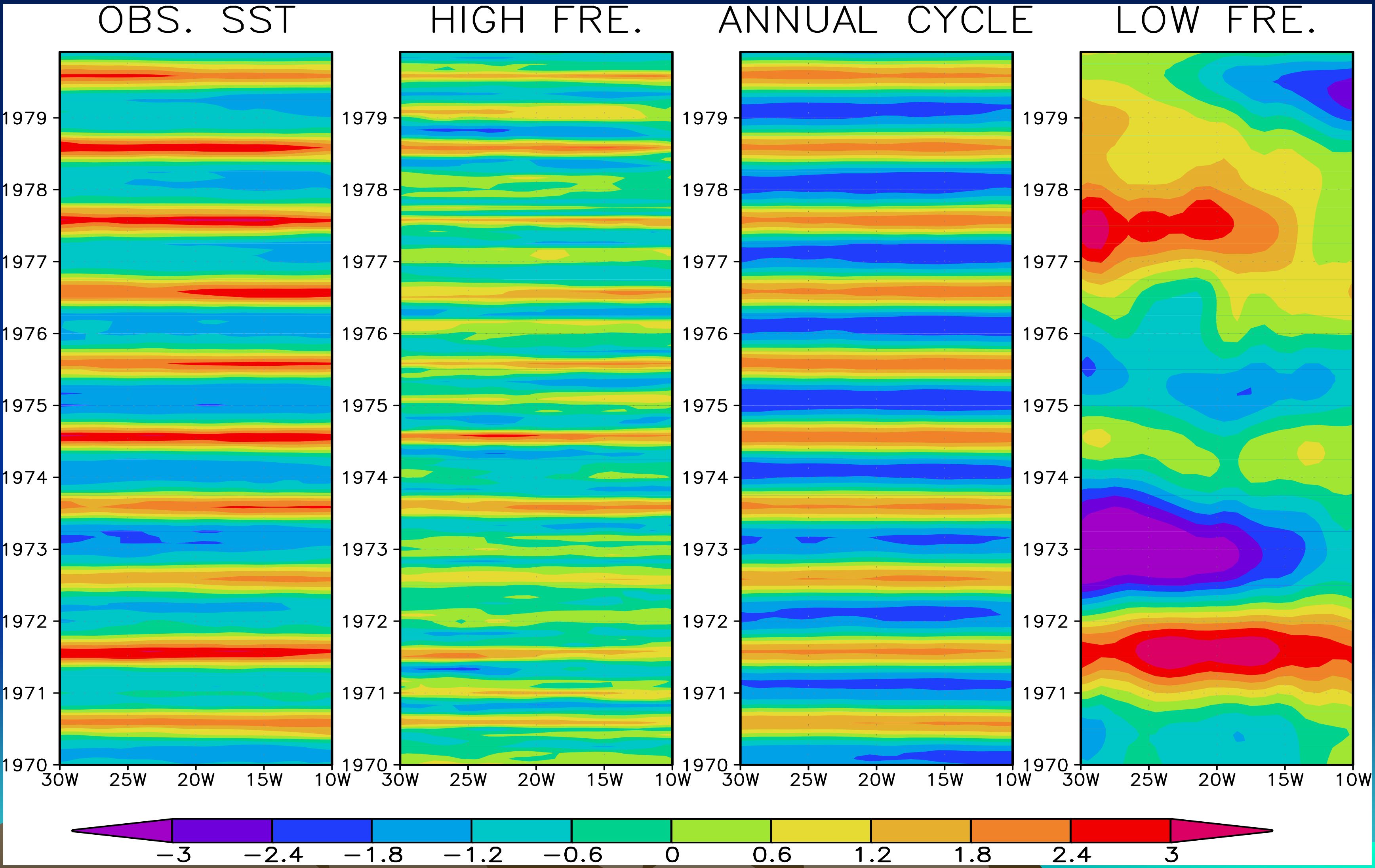
MULTI-DIMENSIONAL EMD/EEMD TEMPORAL-SPATIAL 2D



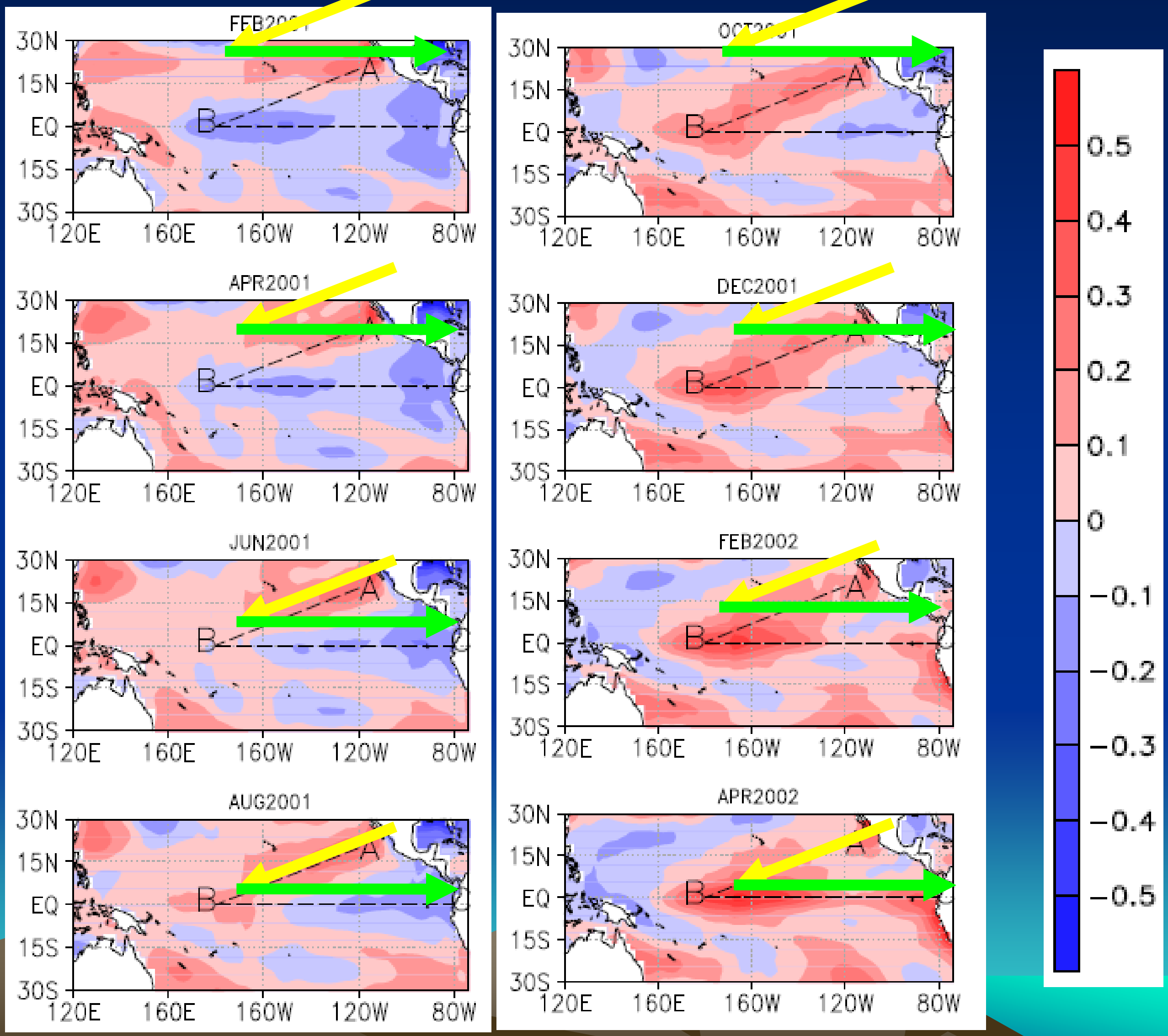
MULTI-DIMENSIONAL EMD/EEMD TEMPORAL-SPATIAL 2D

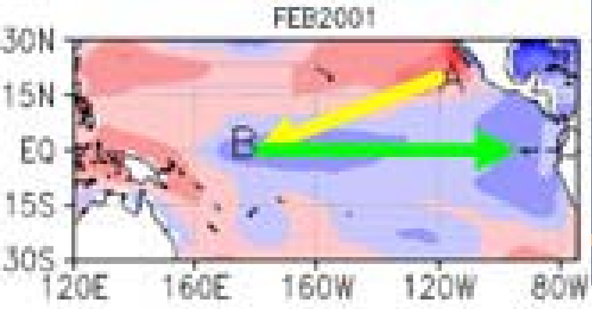


MULTI-DIMENSIONAL EMD/EEMD TEMPORAL-SPATIAL 2D



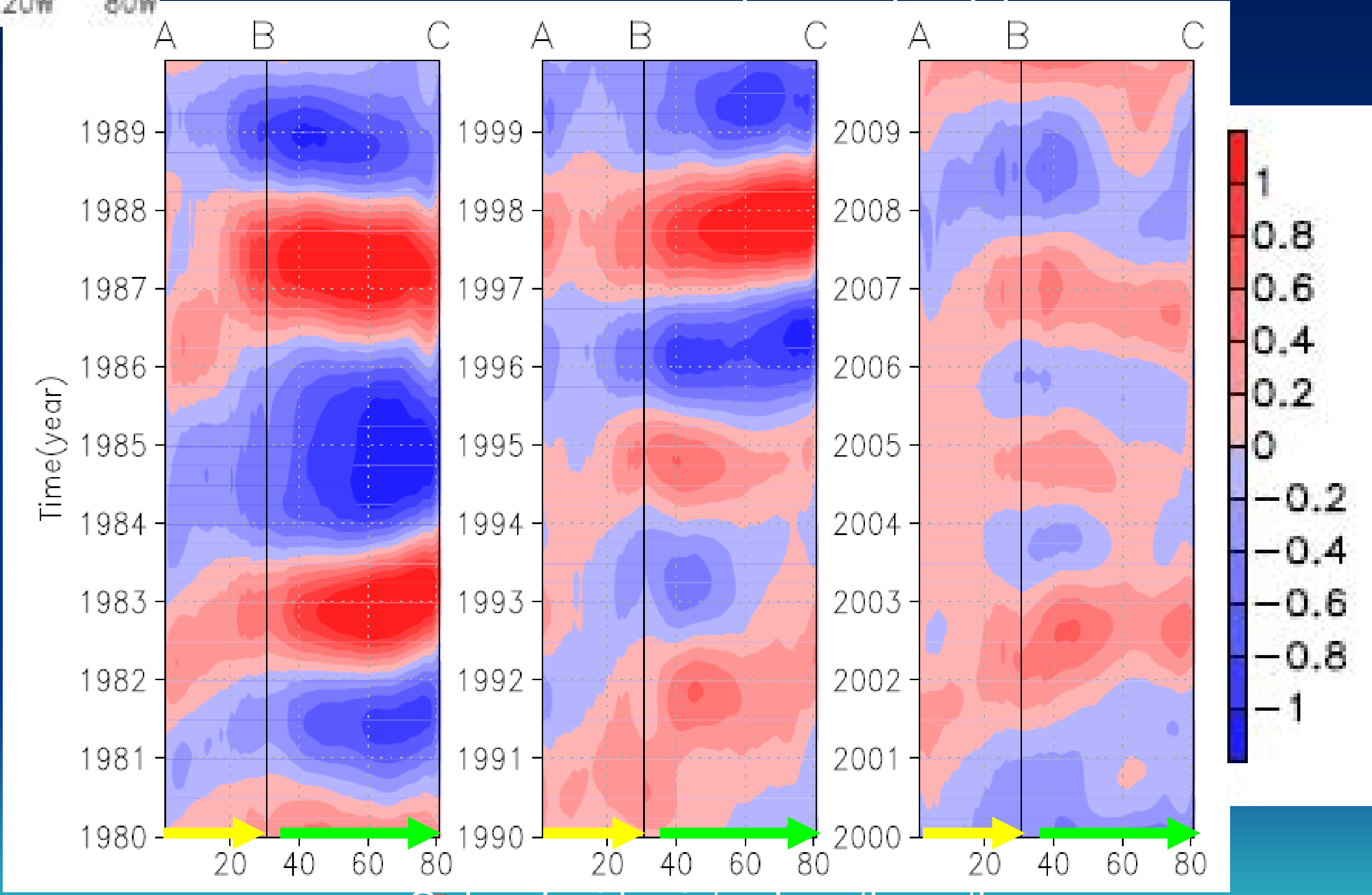
EVOLUTION OF ENSO





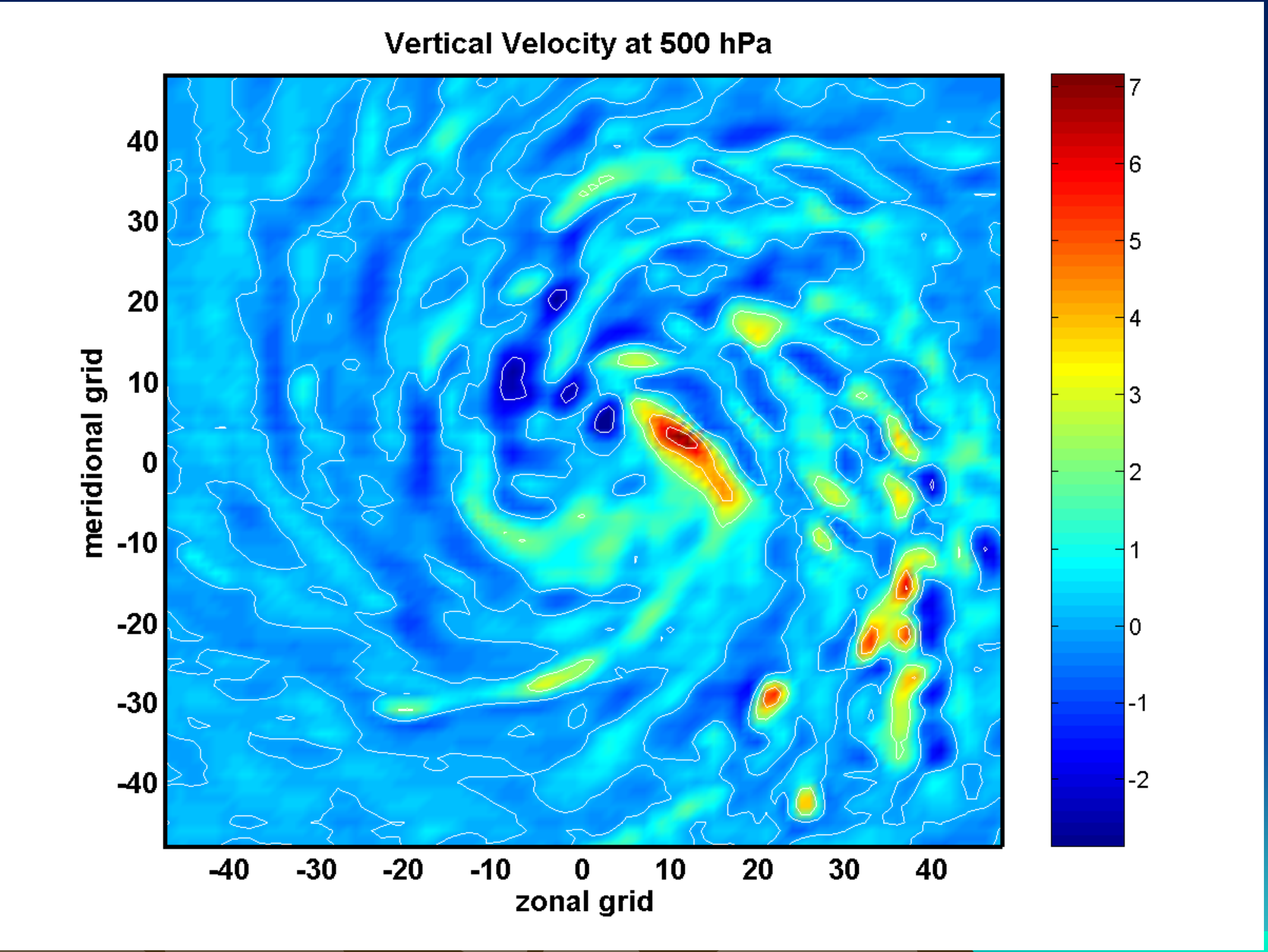
TRIGGER OF ENSO

First Interannual Component (2-3yr)

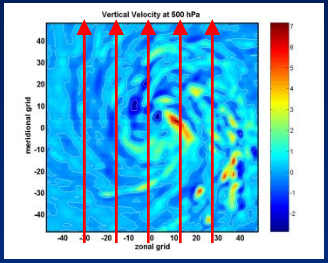


Order of grid points along the path

SPATIAL 2D EMD



SCHEMATIC OF DECOMPOSITION

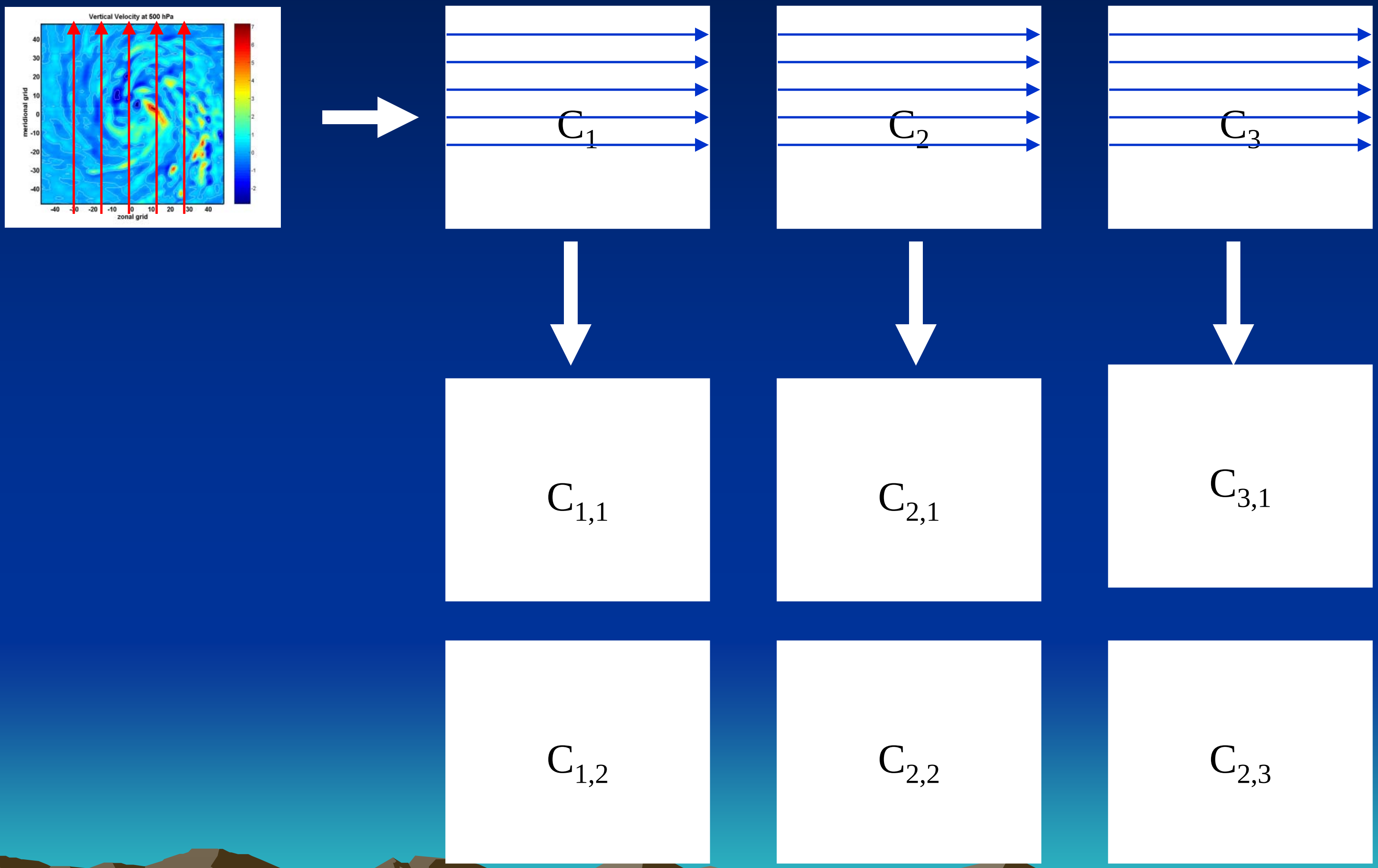


C_1

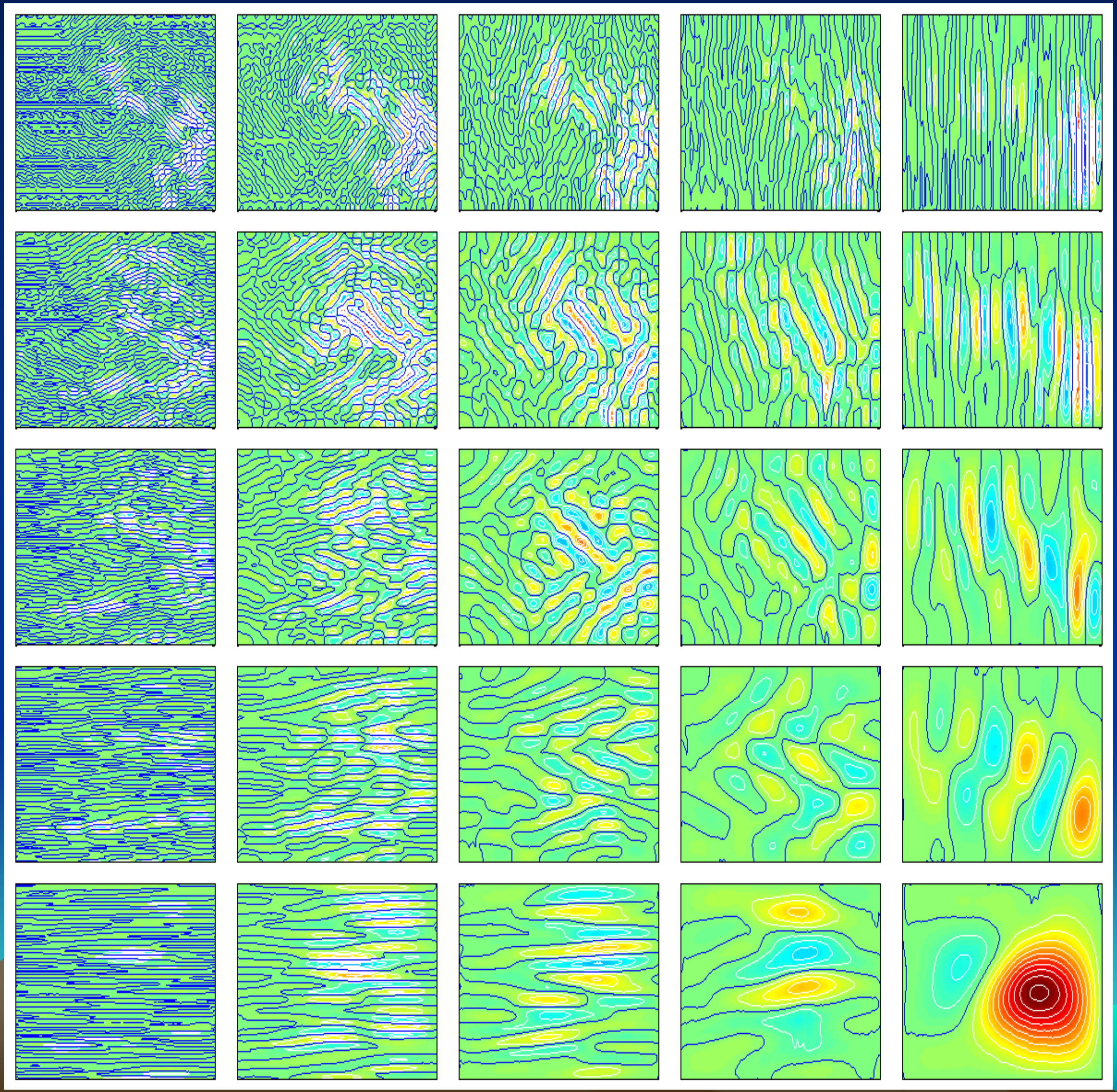
C_2

C_3

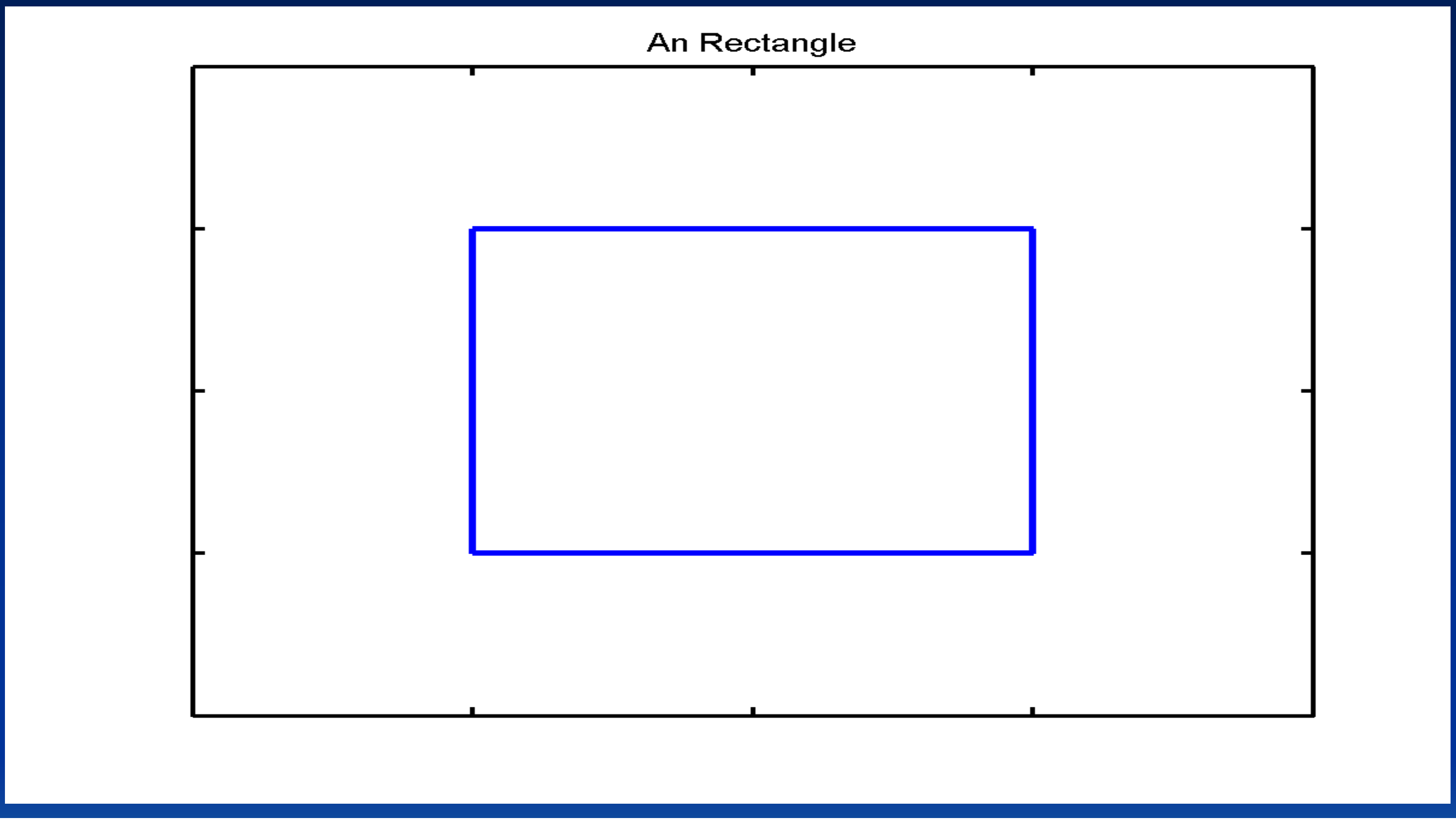
SCHEMATIC OF DECOMPOSITION



DECOMPOSITION EXAMPLE

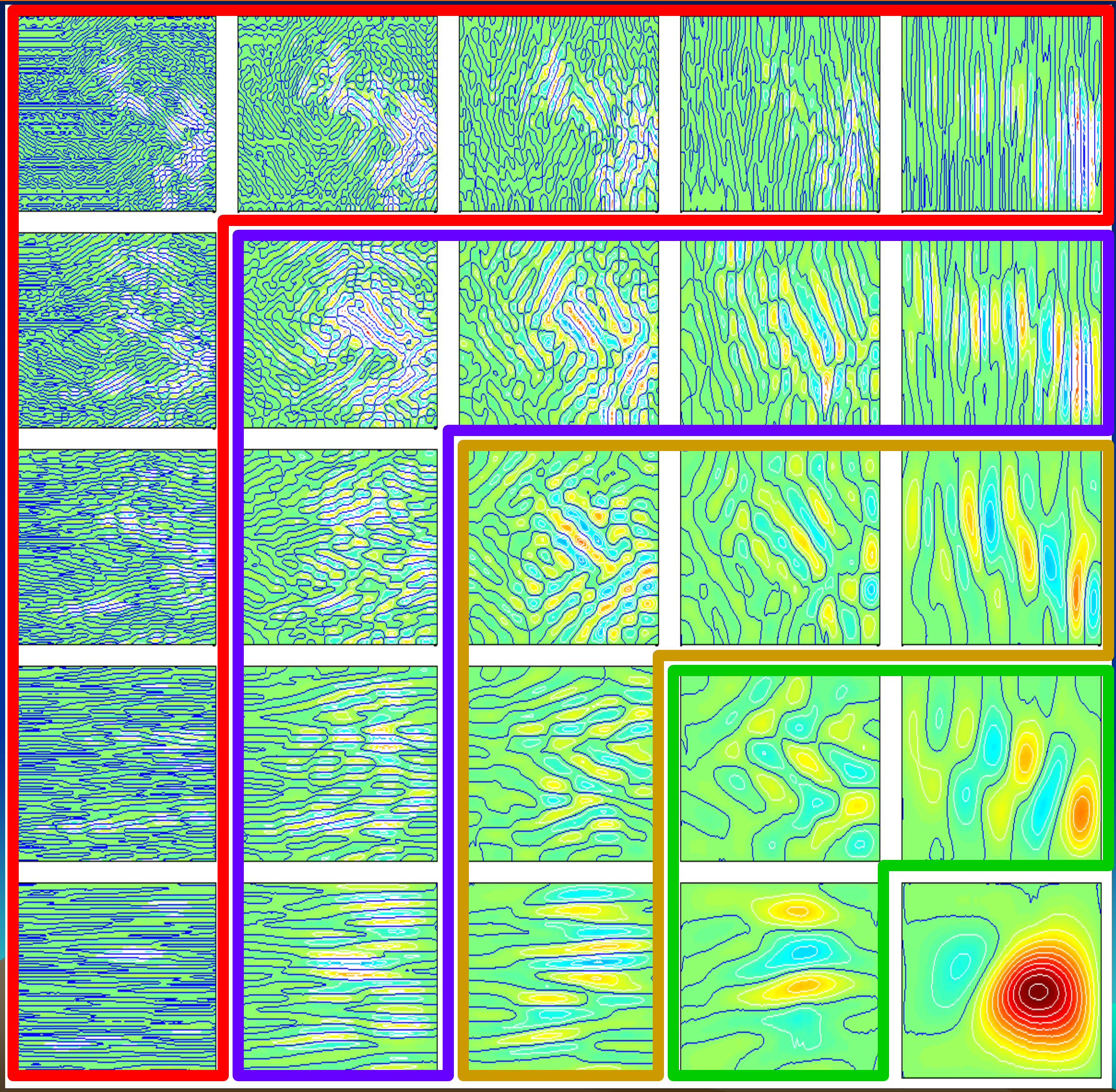


MINIMUM SCALE STRATEGY

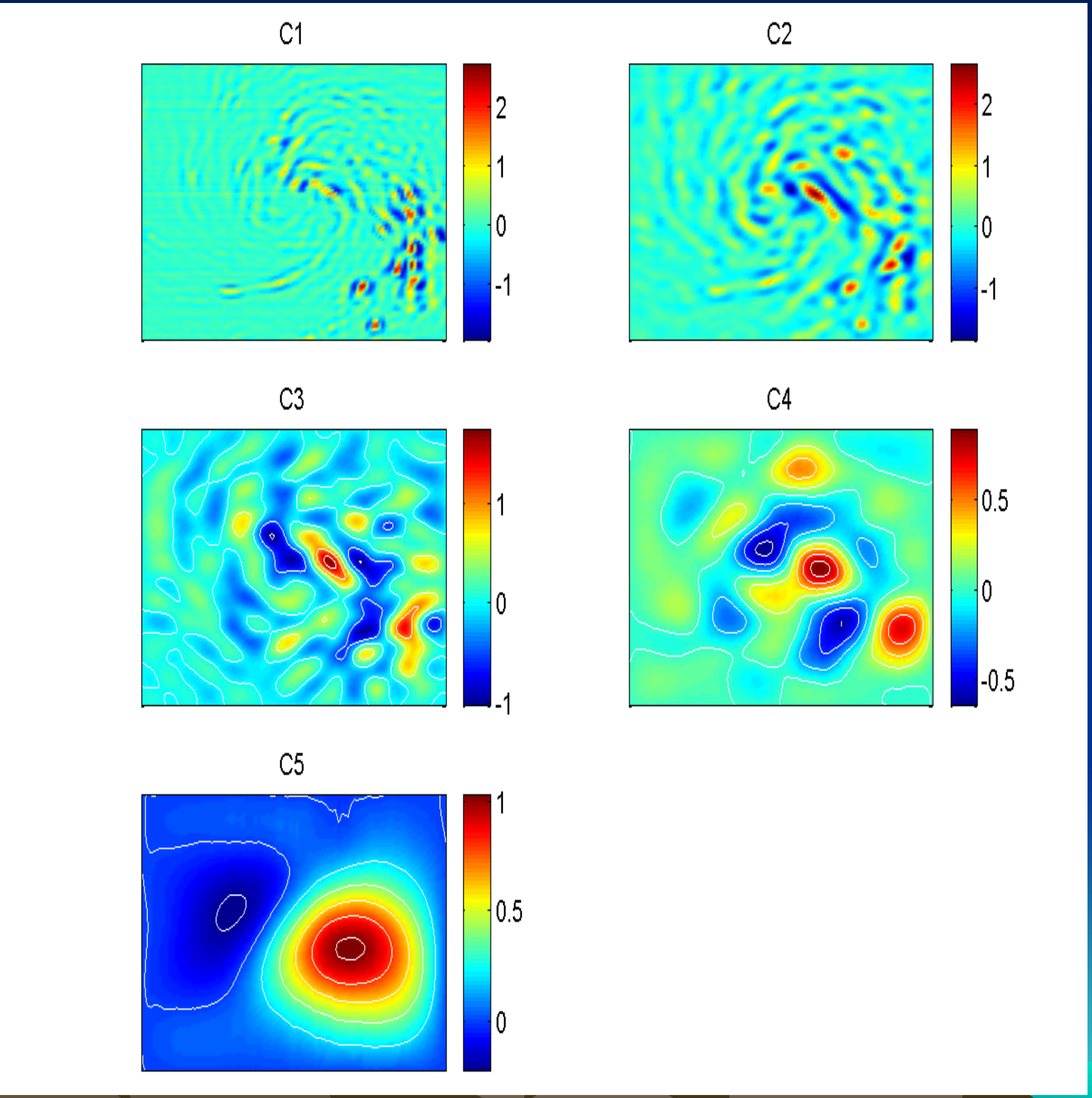


Among all the components resulted from applying of EEMD in two orthogonal directions, the components that have approximately the same minimal scales are combined to one component

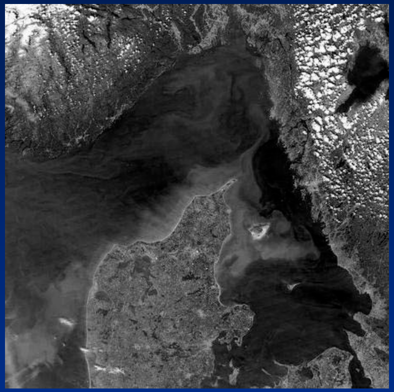
MINIMUM SCALE STRATEGY

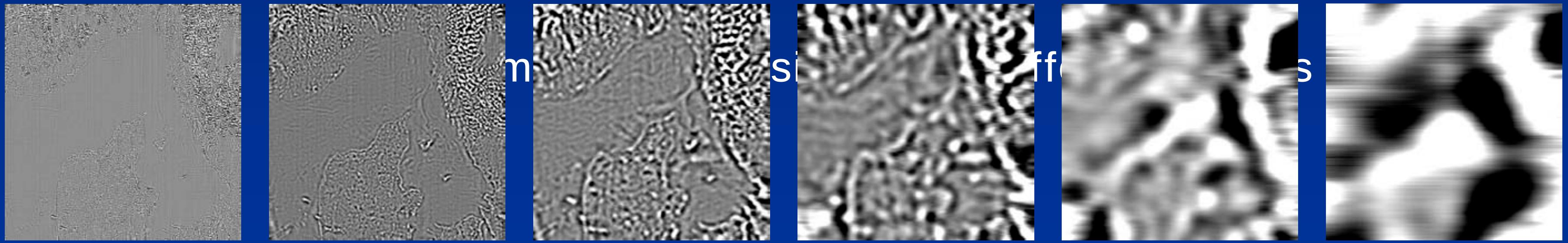


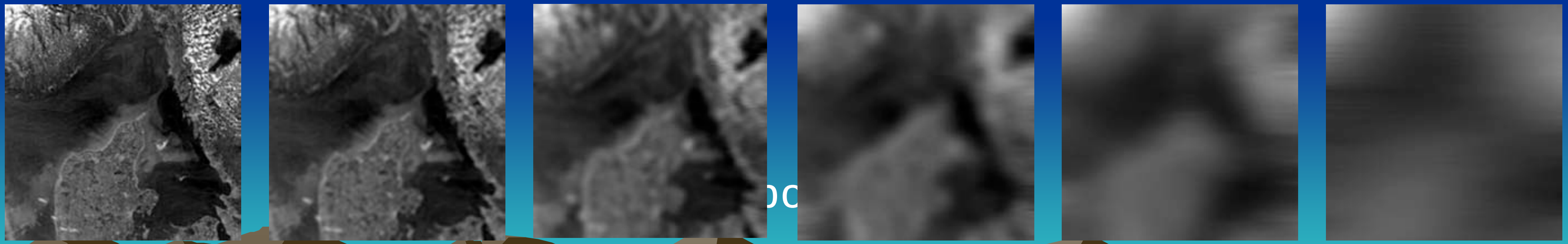
THE FINAL COMPONENTS



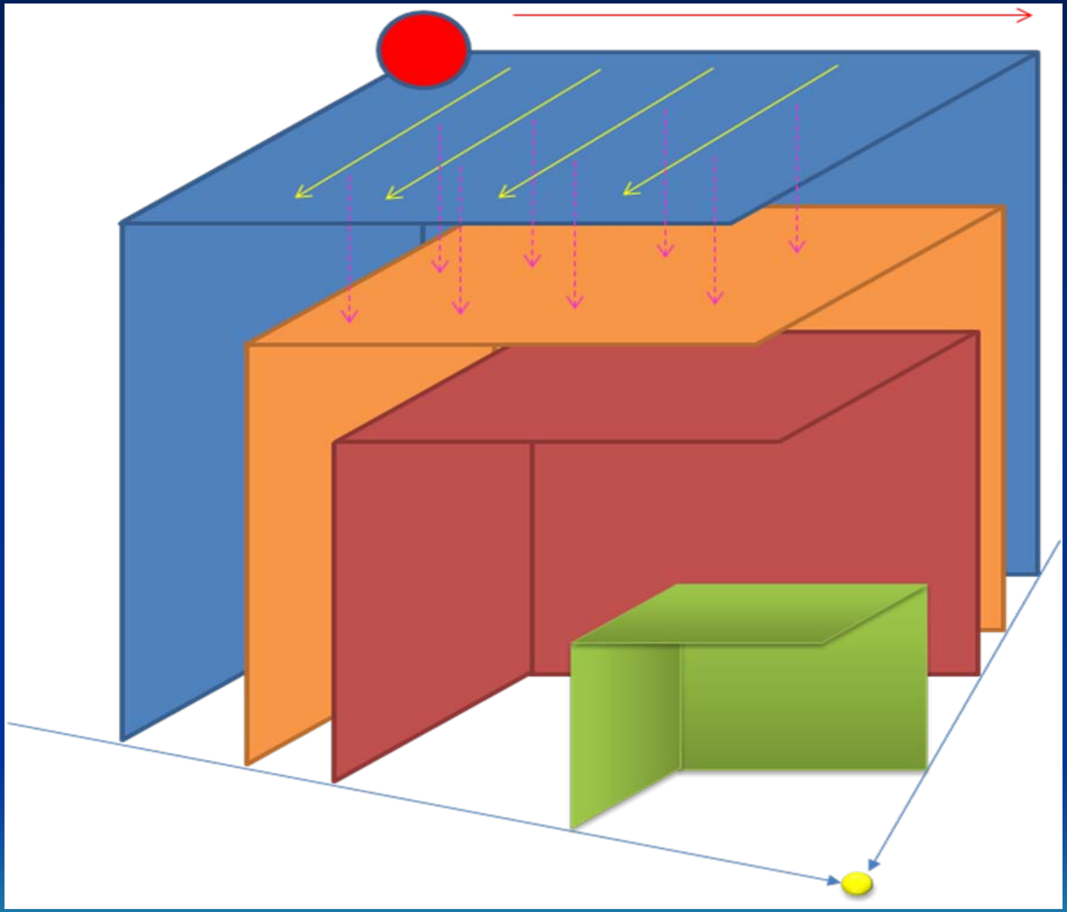
OCEAN COLOR PICTURE







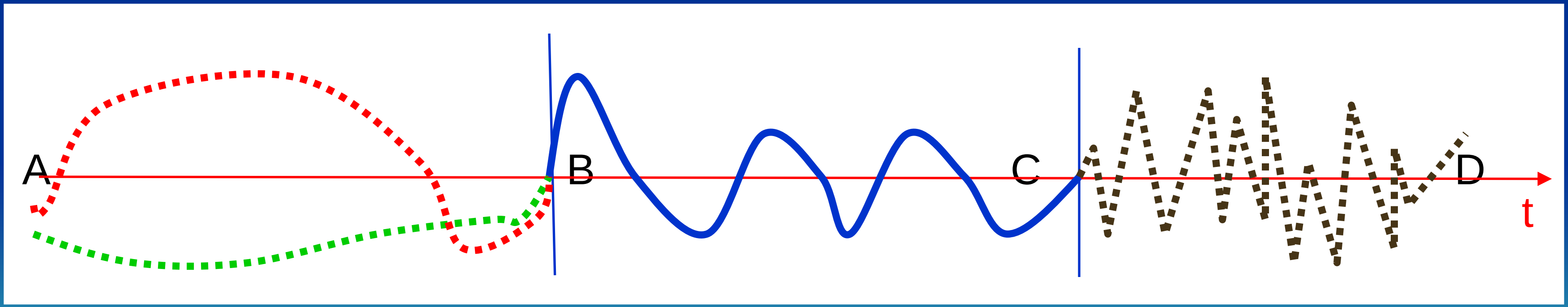
EXTENSION TO MULTI-DIMENSIONAL EEMD



COMPUTATIONAL SPEED: O(NlogN), N the total data points

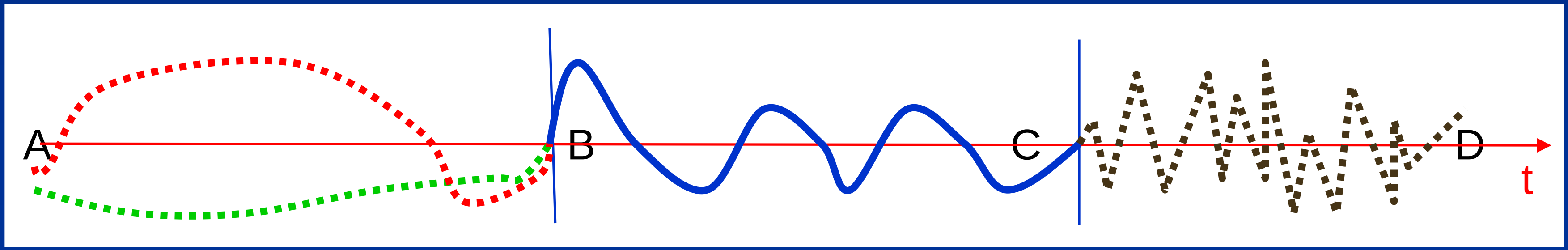
PHYSICAL CONSTRAINTS

1. Later evolution can not change the past
2. What matter to a dynamic system's future evolution are its initial condition boundary condition, and external forcing



TEMPORAL LOCALITY

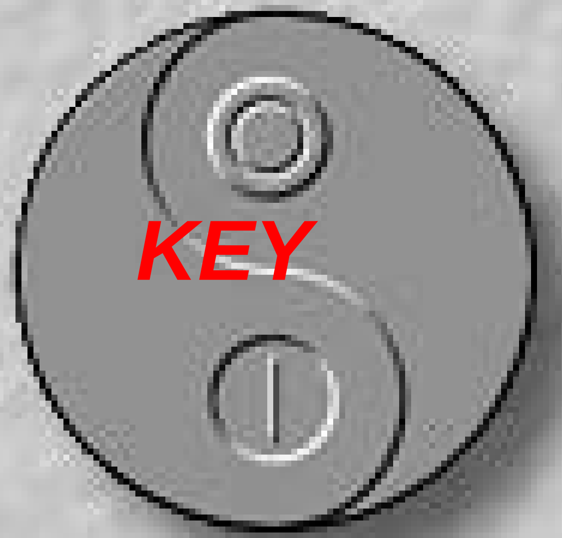
Suppose that the data ***BC*** contains physically meaningful oscillation (signal) and an analysis method extracts that oscillation. If the data is extended to ***AD*** and the same method is applied to ***AD***, the physically meaningful oscillation within BC should not be changed.



When a scientific data analysis method is designed, "**temporal locality**" should be checked.

CONCLUSION

Noise is a



KEY

to unlock the signals in data