# A plea for adequate data analysis methods

Norden E. Huang

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

MCMD Conference Taipei, Taiwan 2010

# A plea for adequate data analysis

- Whatever we want to do related to MCMD:
  - On forecast hazard occurrence
  - On increase Infrastructure resiliency
  - On quantify risk, and
  - On minimize hazard impact

- We need good data and, more importantly, adequate data analysis methods.

With the advance of IT and sensor technologies, there is an explosion of data.

We are now drowning in data

yet still thirsty for information.

# Scientific Activities

Collecting, analyzing, synthesizing, and theorizing are the core of scientific activities.

Data analysis is a key link in this continuous loop.

Data is the link to reality

# Data analysis for engineering

- Data for design specifications
- Data for health monitoring
- Data for quantify risk
- Data to minimizing hazard impacts

# Example: Health monitoring

- Most natural hazards are hard to predict; therefore, it is critical that we should increase the structures resiliency to withstand the impacts of the hazards. The easiest and most logic ways to maximize structural resiliency are these:
  - To design the structure intelligently, and
  - To keep the structure in healthy conditions and good repair.

# Land Slide damages

**Photograph by R.L. Schuster, U.S. Geological Survey, 1995.)**

# Earthquake Damage

**(www.cedim.de)**

# Hurricane Damage

**Southern Grenada after Hurricane Ivan (Photo WW-2004).**

# Data Analysis

Data analysis is too important to be left to the mathematicians.

Why?!

# Different Paradigms I

## Mathematics vs. Science/Engineering

- **<u>Mathematicians</u>**
  - **Absolute proofs**
  - **Logic consistency**
  - **Mathematical rigor**

- **<u>Scientists/Engineers</u>**
  - **Agreement with observations**
  - **Physical meaning**
  - **Working Approximations**

# Data Processing vs. Analysis

In pursue of mathematic rigor and certainty, however, we lost sight of physics and are forced to idealize, but also deviate from, the reality.

As a result, we are forced to live in a pseudo-real world, in which all processes are

Linear and Stationary

# On nonlinear oscillations

The key to structure health monitoring

# How to define nonlinearity?

Based on Linear Algebra: nonlinearity is defined based on input vs. output.

But in reality, such an approach is not practical. The alternative is to define nonlinearity based on data characteristics.

# Linear Systems

Linear systems satisfy the properties of superposition and scaling. Given two valid inputs

$$x_1(t) \; and \;\; x_2(t)$$

as well as their respective outputs

$$y_1(t) = H\{x_1(t)\} \;\; and$$
$$y_2(t) = H\{x_2(t)\}$$

then a linear system must satisfy

$$\alpha y_1(t) + \beta y_1(t) = H\{\alpha x_1(t) + \beta x_2(t)\}$$

for any scalar values **α** and **β**.

# Characterizing Nonlinear Processes from Data

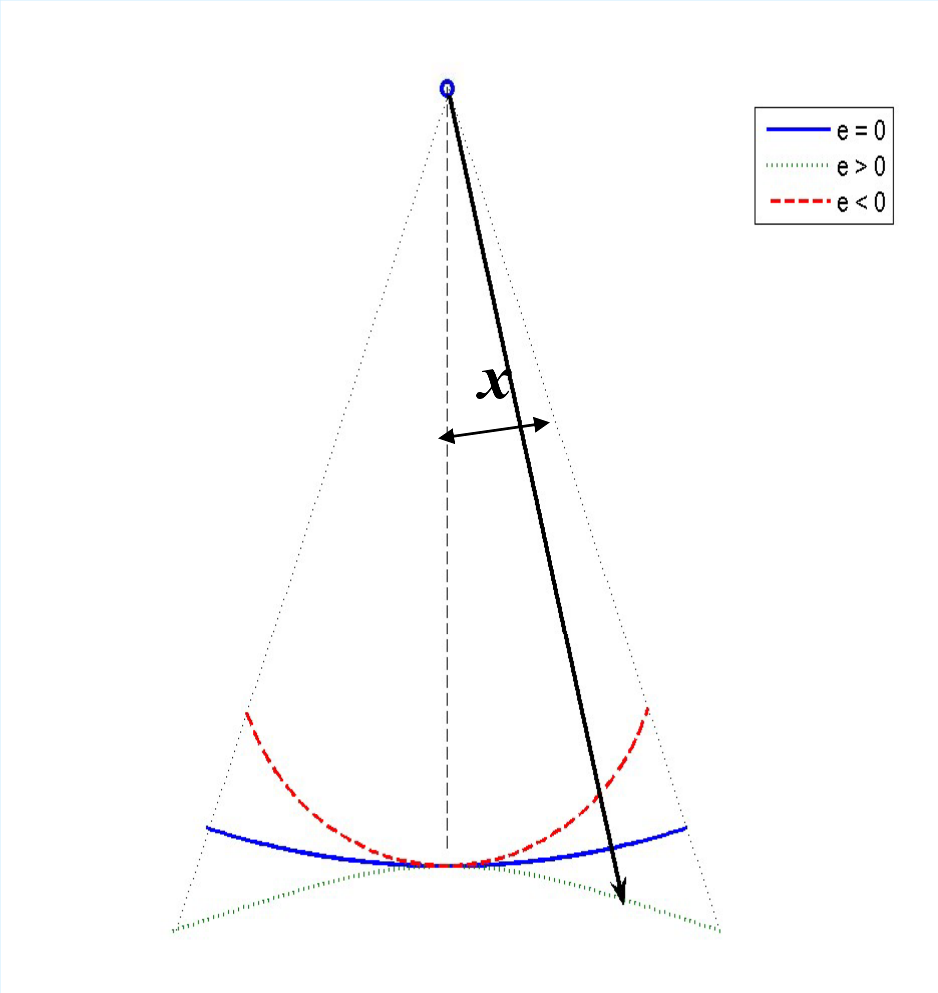
$$\frac{d^2 x}{dt^2} + x + \varepsilon x^3 = \gamma \cos \omega t$$

$$\Rightarrow \quad \frac{d^2 x}{dt^2} + x \left( 1 + \varepsilon x^2 \right) = \gamma \cos \omega t$$

⇒ *Spring with position dependent constant,*
*intra-wave frequency modulation;*
*therefore, we need instantaneous frequency.*

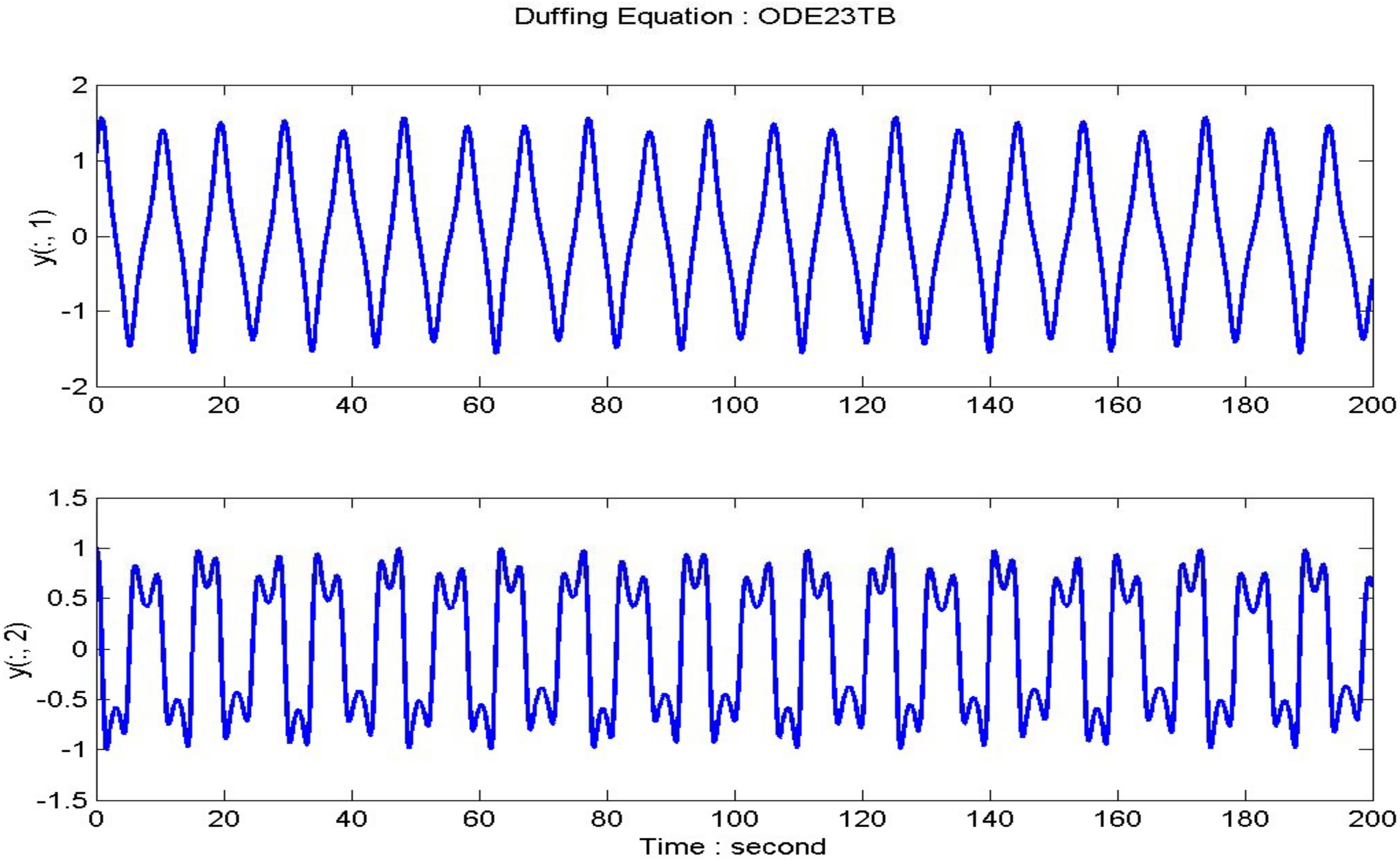
# Duffing Pendulum:

# Intra-Wave Frequency Modulation




$$\frac{d^2x}{dt^2} + x\,(1 + \varepsilon\,x^2) = \gamma\cos\omega t\ .$$

# Duffing Equation : Data



# Duffing Type Wave

## Perturbation Expansion

*For* $\varepsilon \ll 1$, *we can have*

$$x(t) = \cos\left(\omega t + \varepsilon \sin 2\omega t\right)$$

$$= \cos \omega t \cos\left(\varepsilon \sin 2\omega t\right) - \sin \omega t \sin\left(\varepsilon \sin 2\omega t\right)$$

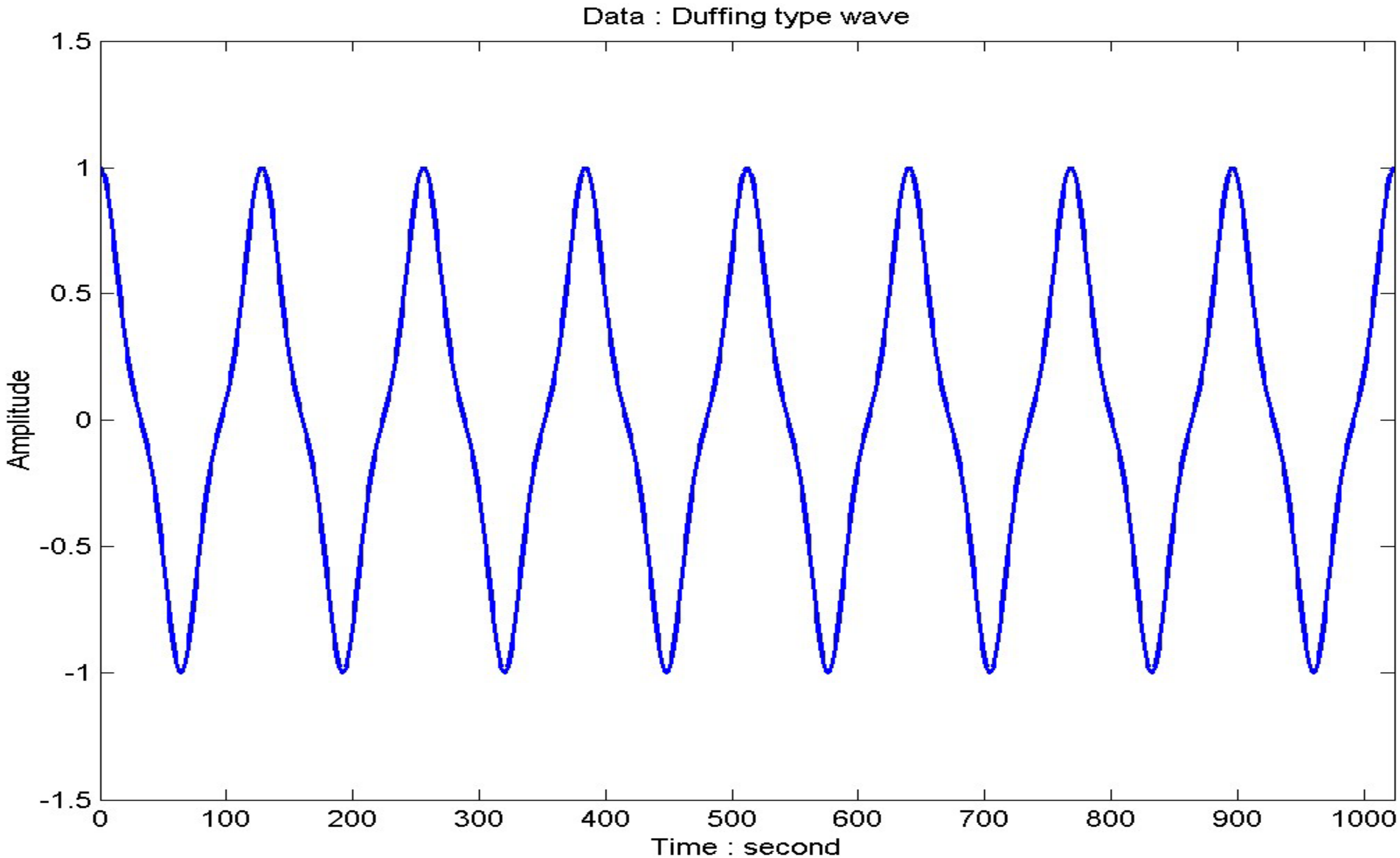
$$= \cos \omega t - \varepsilon \sin \omega t \sin 2\omega t + ....$$

$$= \left(1 - \frac{\varepsilon}{2}\right)\cos \omega t + \frac{\varepsilon}{2} \cos 3\omega t + ....$$

*This is very similar to the solution of Duffing equation .*

# Duffing Type Wave

Data: $x = cos(wt+0.3\ sin2wt)$



# 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*

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

# Instantaneous Frequency

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

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

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

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

$$So\ that\ \ both\ \ v\ and\ \ \omega\ can\ appear\ \ in\ differential\ \ equations.$$

The combination of Hilbert Spectral Analysis and Empirical Mode Decomposition is designated as

HHT

(HHT vs. FFT)

# Comparison between FFT and HHT

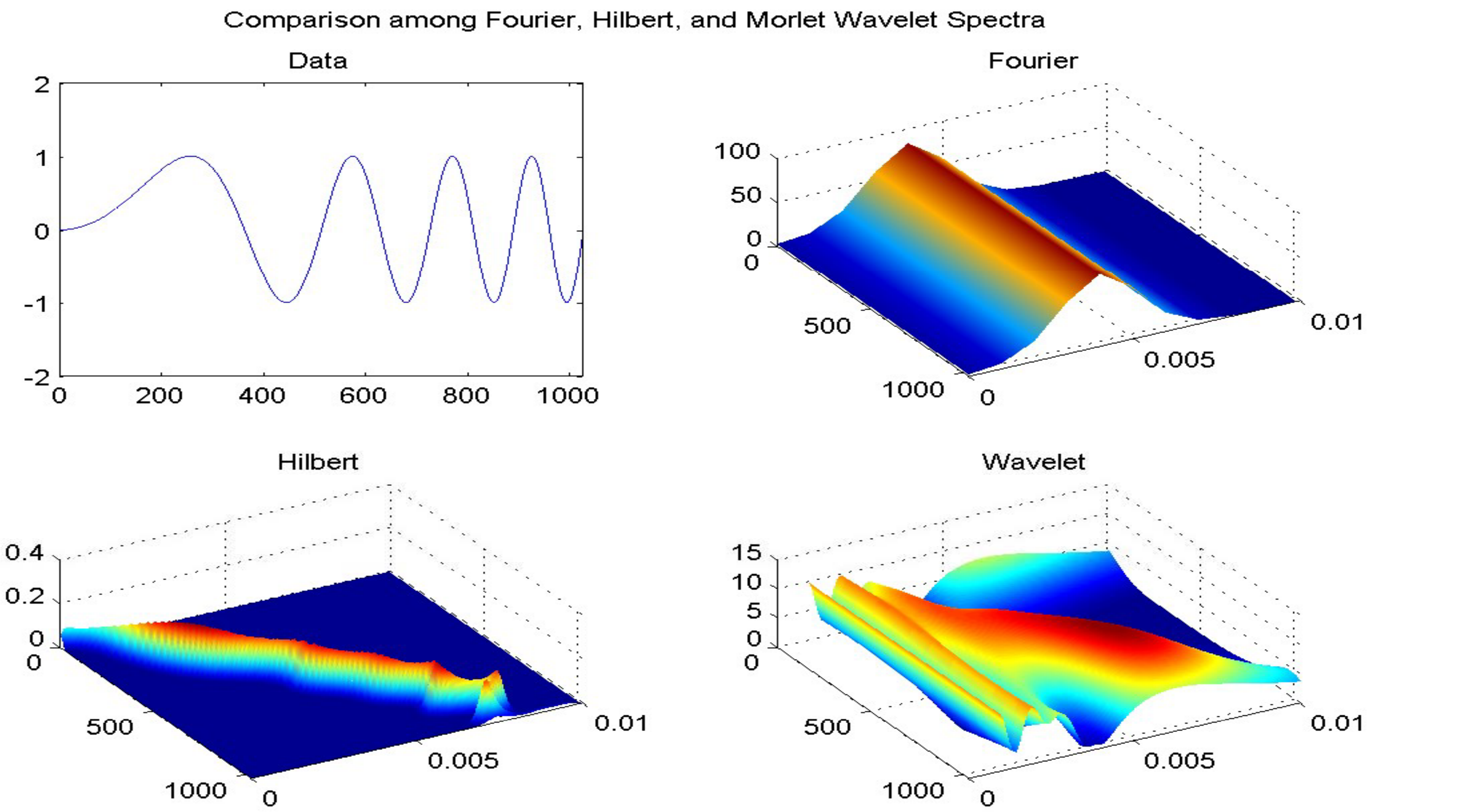
*1. FFT :*

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

*2. HHT :*

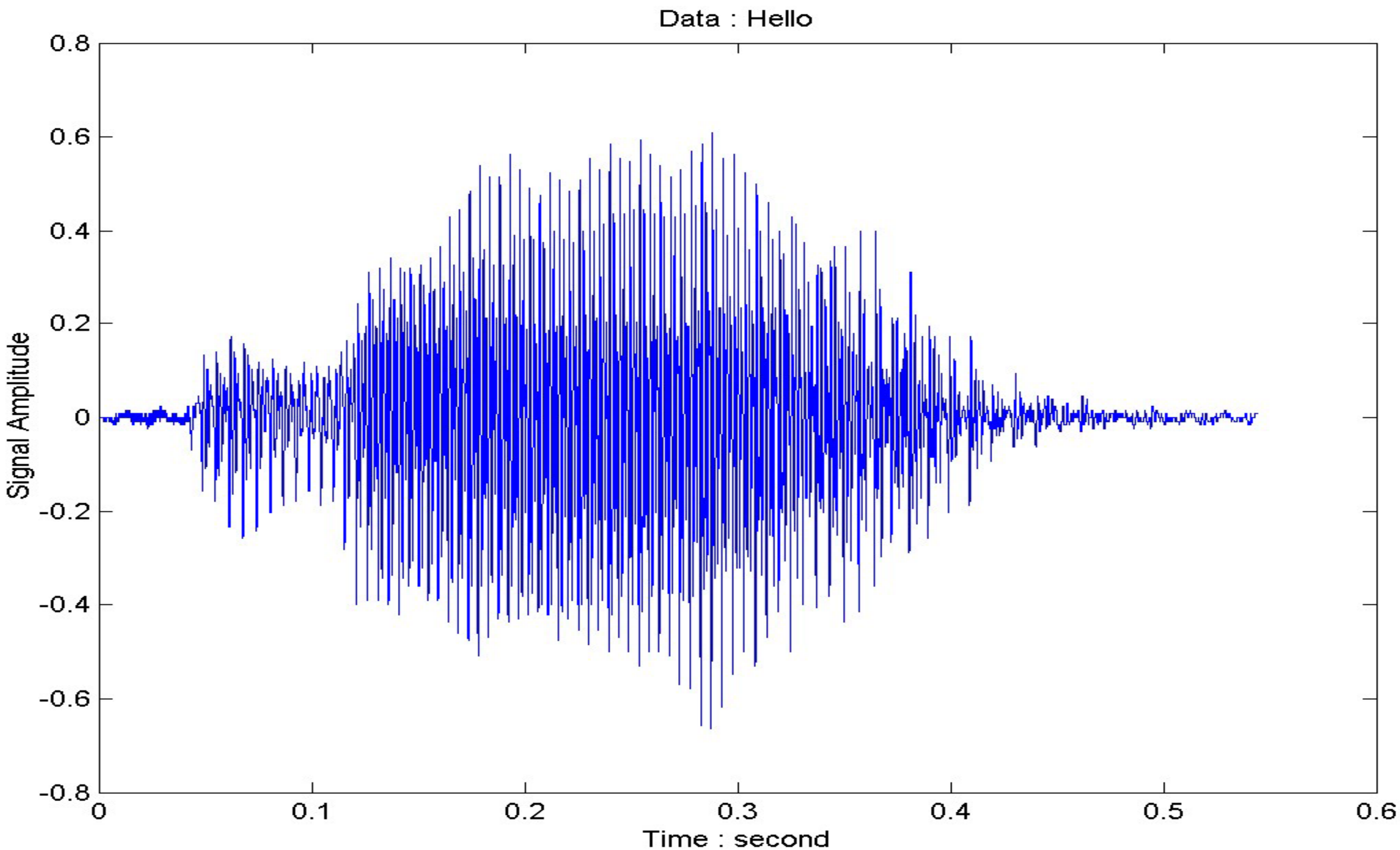
$$x(t) = \Re \sum_{j} a_j(t)\, e^{i\int_t \omega_j(\tau)\, d\tau} .$$

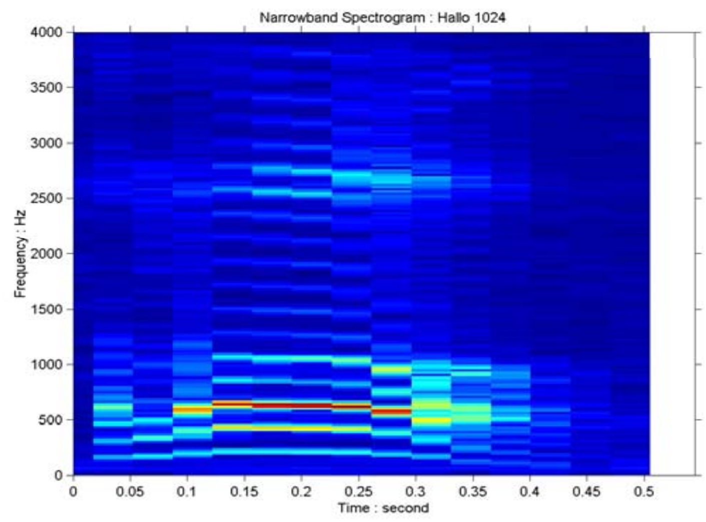
# Comparisons:
# Fourier, Hilbert & Wavelet



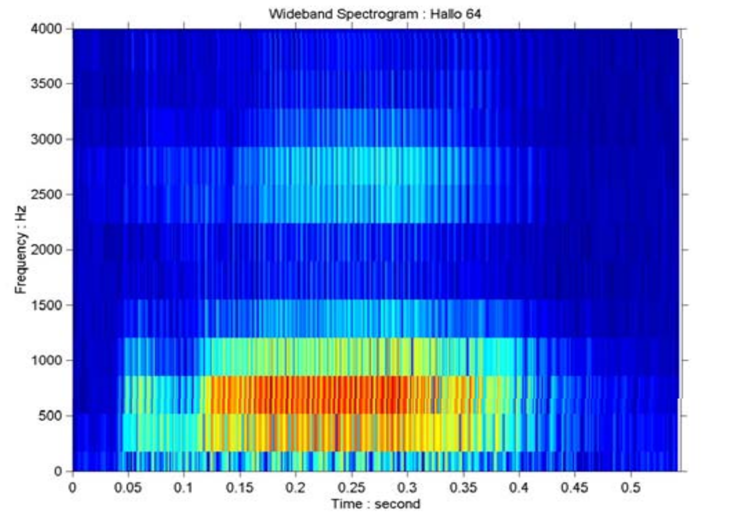
# Speech Analysis

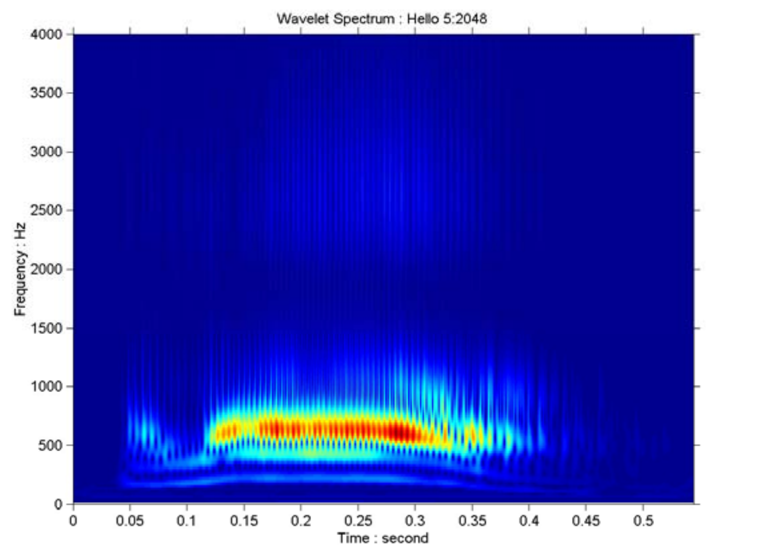
**Hello : Data**

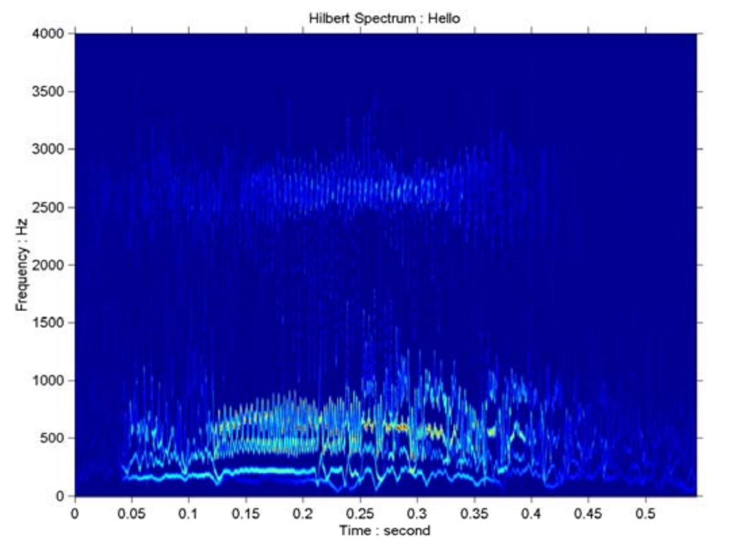





Four comparsions D



# Hilbert's View on Nonlinear Data

# Duffing Type Wave

## Perturbation Expansion

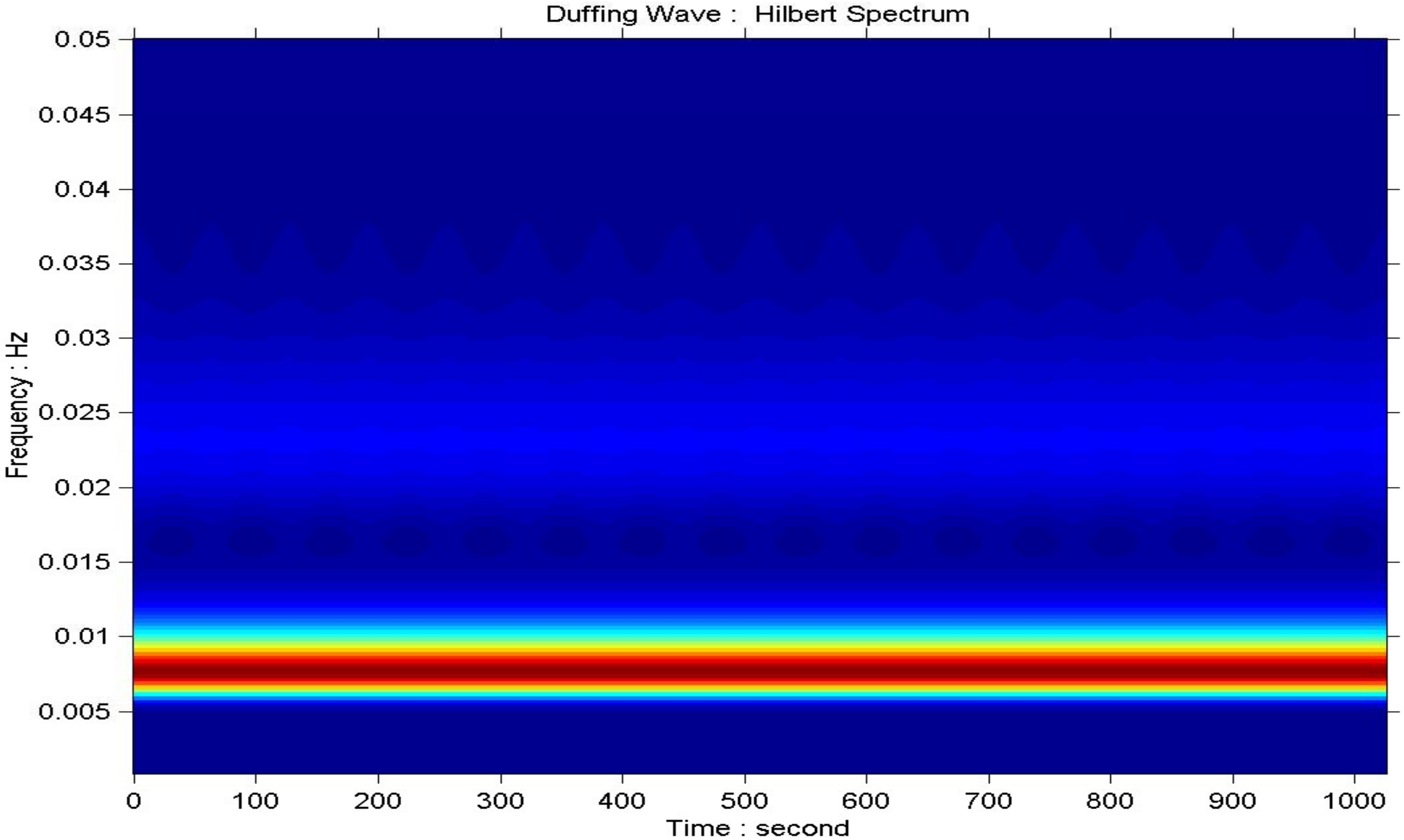
*For* $\varepsilon \ll 1$, *we can have*

$$
\begin{aligned}
x(t) &= \cos\left(\omega t + \varepsilon \sin 2\omega t\right) \\
&= \cos\omega t \cos\left(\varepsilon \sin 2\omega t\right) - \sin\omega t \sin\left(\varepsilon \sin 2\omega t\right) \\
&= \cos\omega t - \varepsilon \sin\omega t \sin 2\omega t + \ldots. \\
&= \left(1 - \frac{\varepsilon}{2}\right)\cos\omega t + \frac{\varepsilon}{2}\cos 3\omega t + \ldots.
\end{aligned}
$$

*This is very similar to the solution of Duffing equation.*

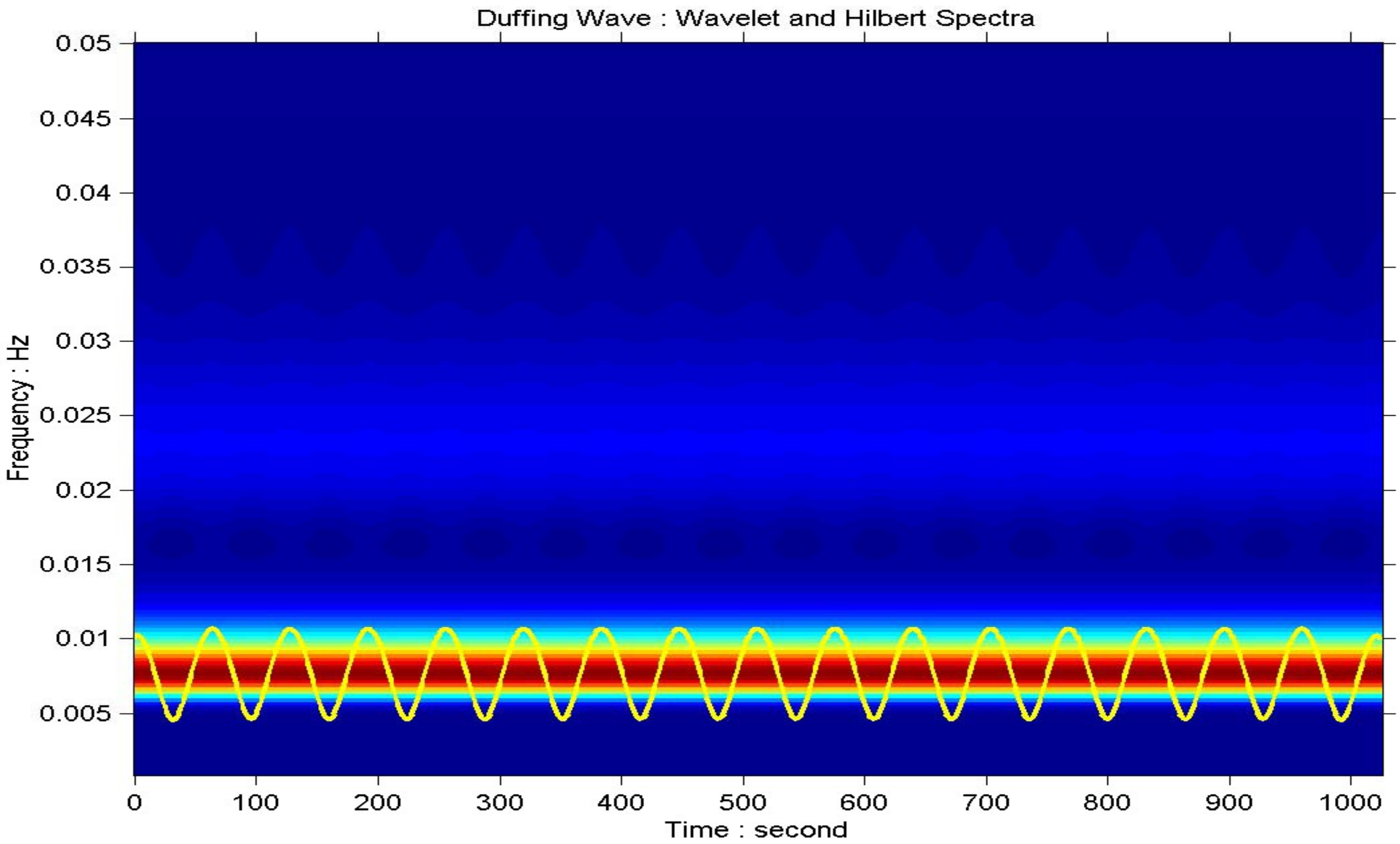
# Duffing Type Wave

## Wavelet Spectrum



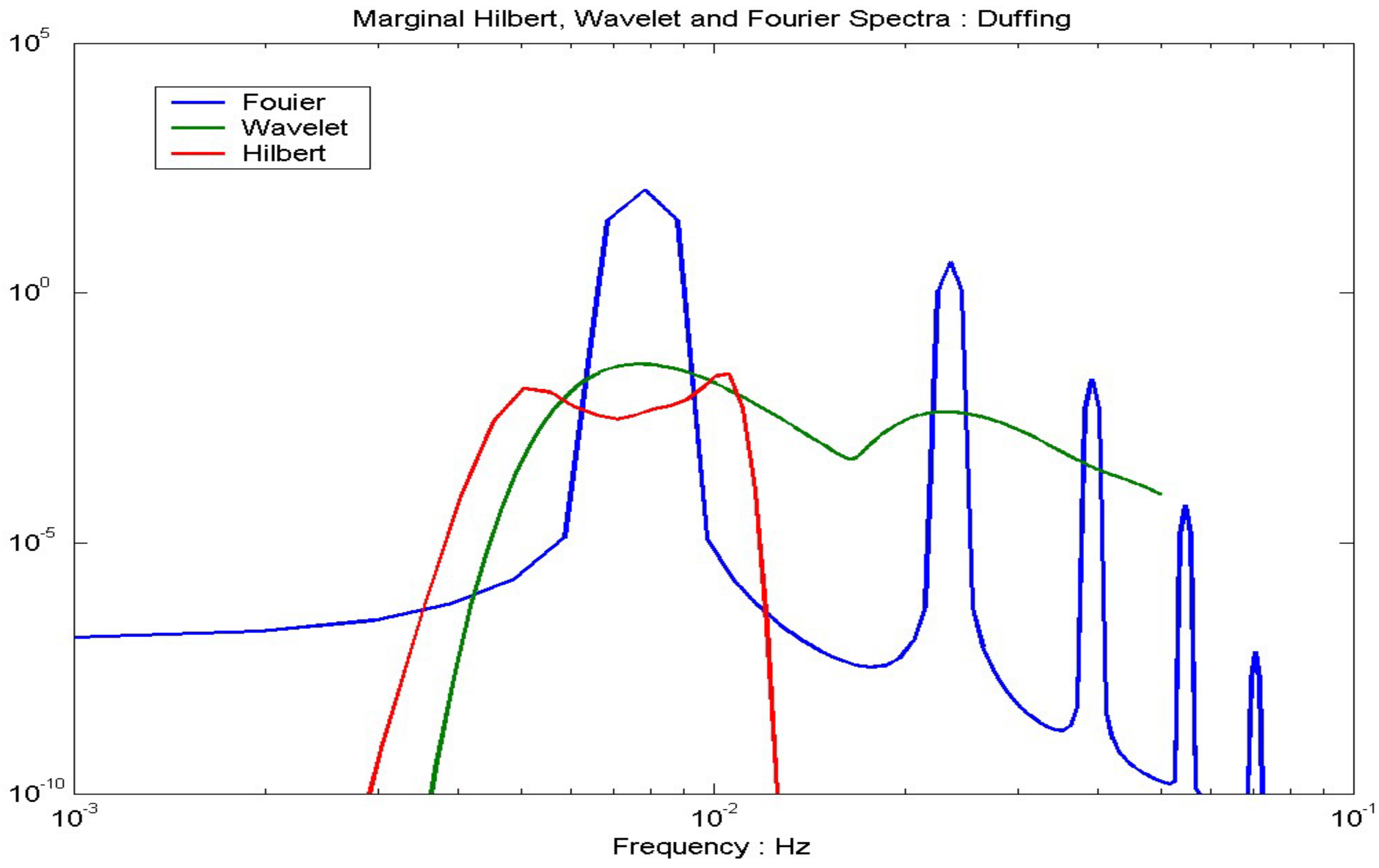
# Duffing Type Wave

## Hilbert Spectrum



# Duffing Type Wave: Marginal Spectrum

Different harmonics from different Method: not physical



# Degree of Nonlinearity

Quantify, rather than qualify, nonlinearity

# Degree of nonlinearity

$$Consider \quad x(t) = \cos(\omega t + \varepsilon \sin \eta \omega t)$$

$$\Rightarrow \quad IF = \frac{d\theta}{dt} = \omega\left(1 + \eta\varepsilon \cos \eta\omega t\right)$$

$$DN\ (Degree\ of\ Nolinearity) = \left\langle \left( \frac{IF - IF_z}{IF_z} \right)^2 \right\rangle^{1/2} = \frac{\eta\varepsilon}{\sqrt{2}}$$

# Degree of Nonlinearity

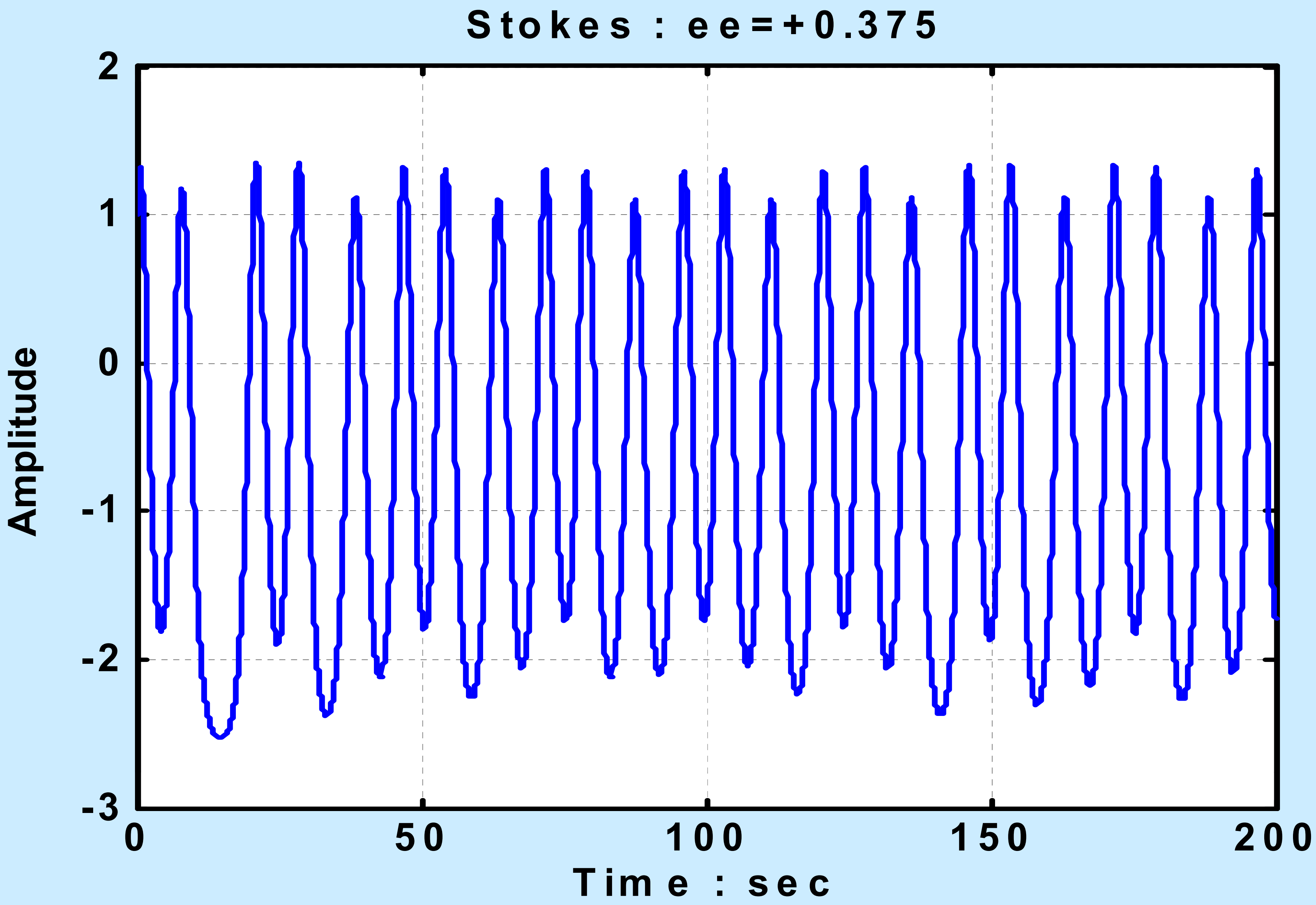
- We can determine DN precisely with Hilbert Spectral Analysis.
- We can also determine ***ε*** *and* ***η*** separately.
- ***η*** can be determined from the frequency modulation.
- ***ε*** can be determined from DN with ***η*** given.

# Stokes Waves

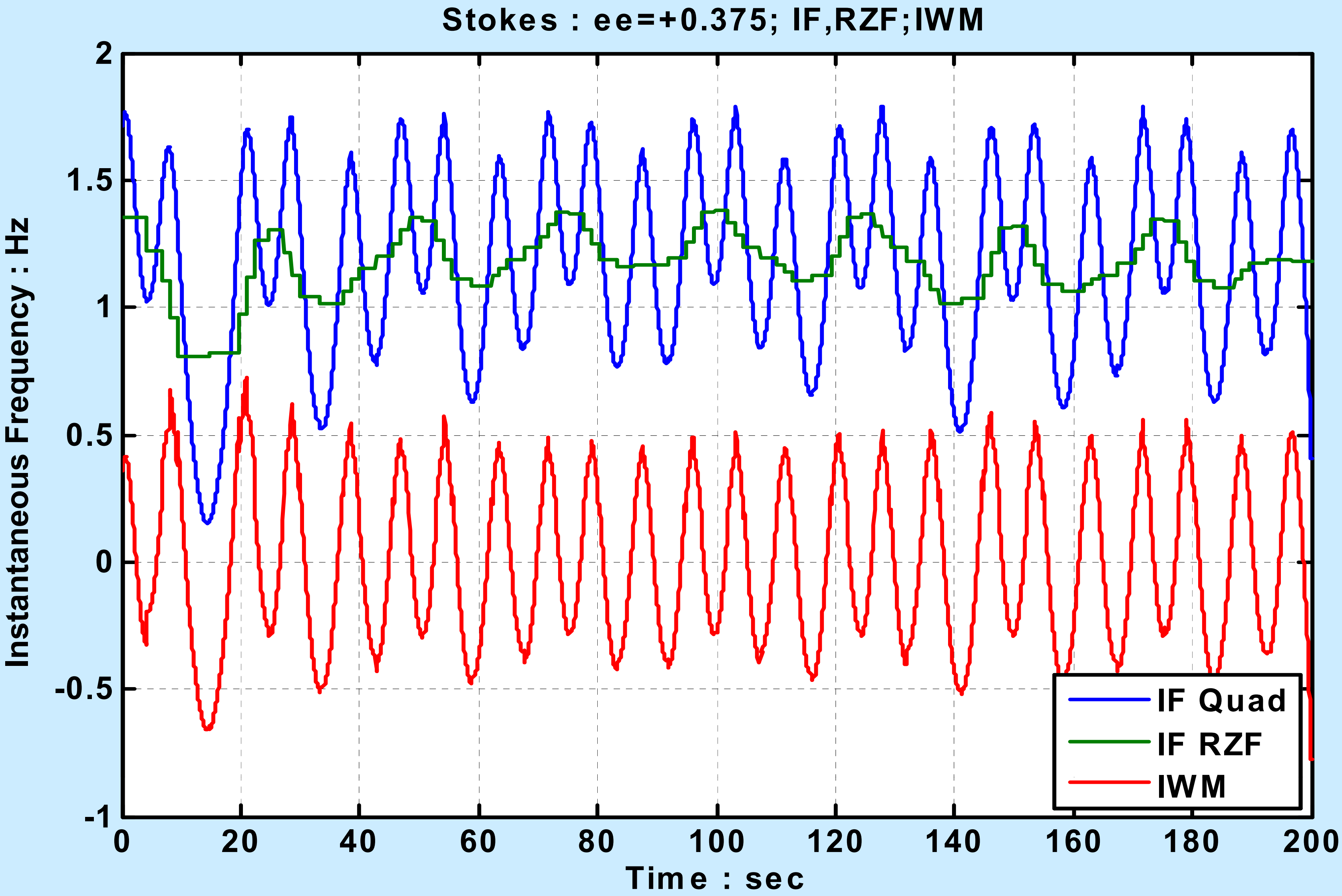
$$\frac{d^2x}{dt^2} + x + \varepsilon x^2 = \gamma \cos \omega t$$

$$\frac{d^2x}{dt^2} + x\left(1 + \varepsilon x\right) = \gamma \cos \omega t$$

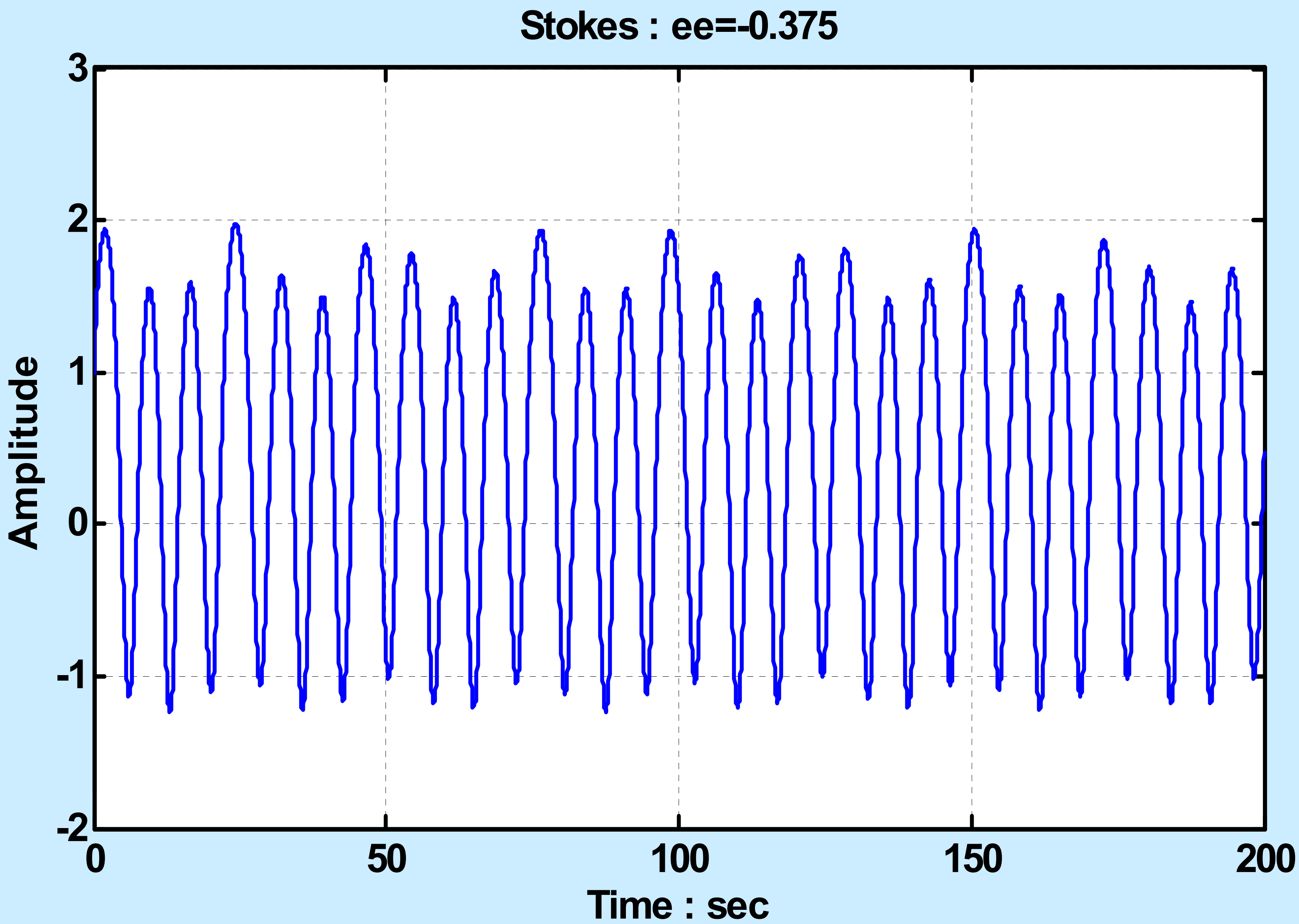
# Stokes ee= +0.375; DN=0.0407



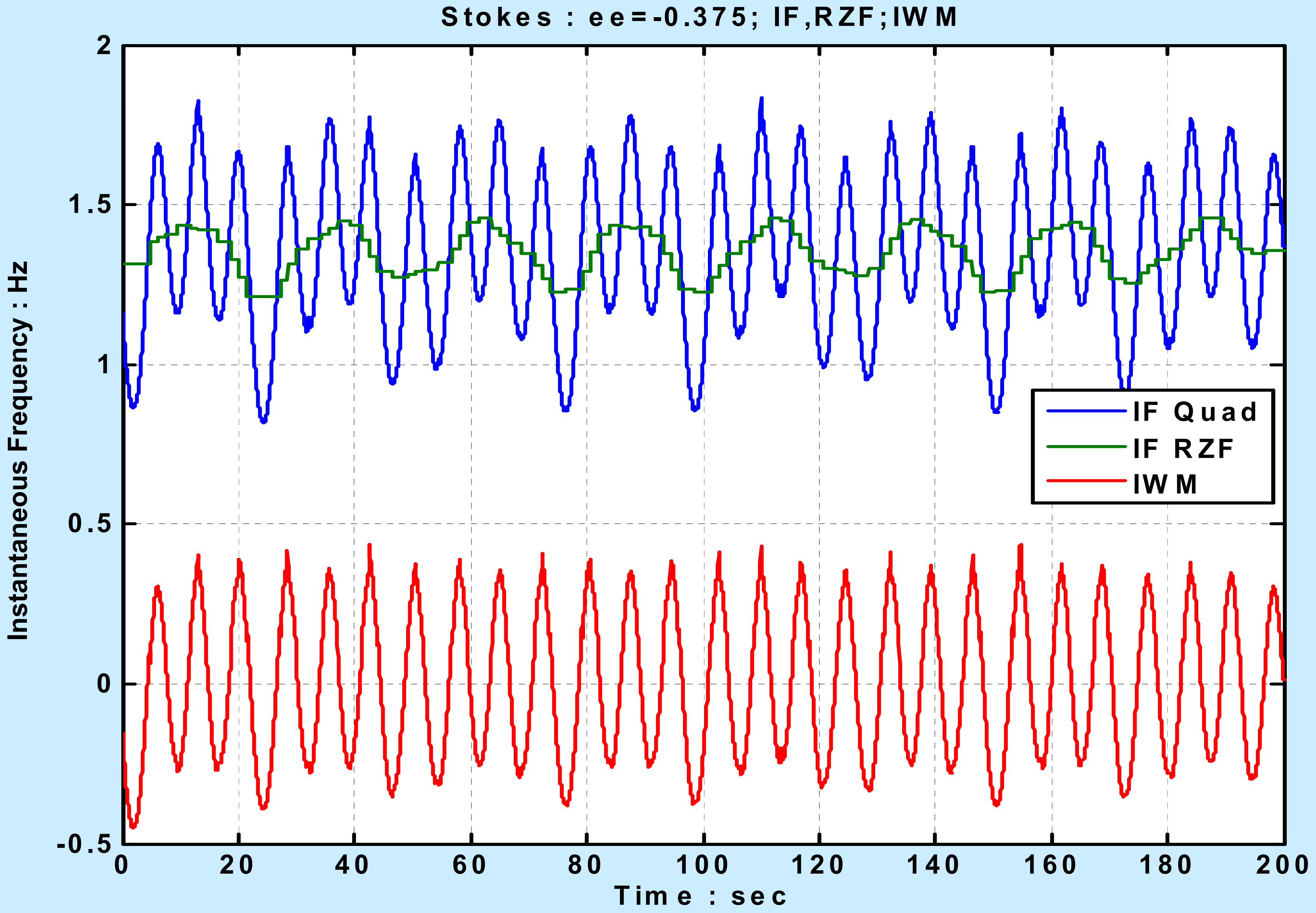
# Stokes ee= +0.375; DN=0.0893



# Stokes ee=-0.375; DN=0.0407



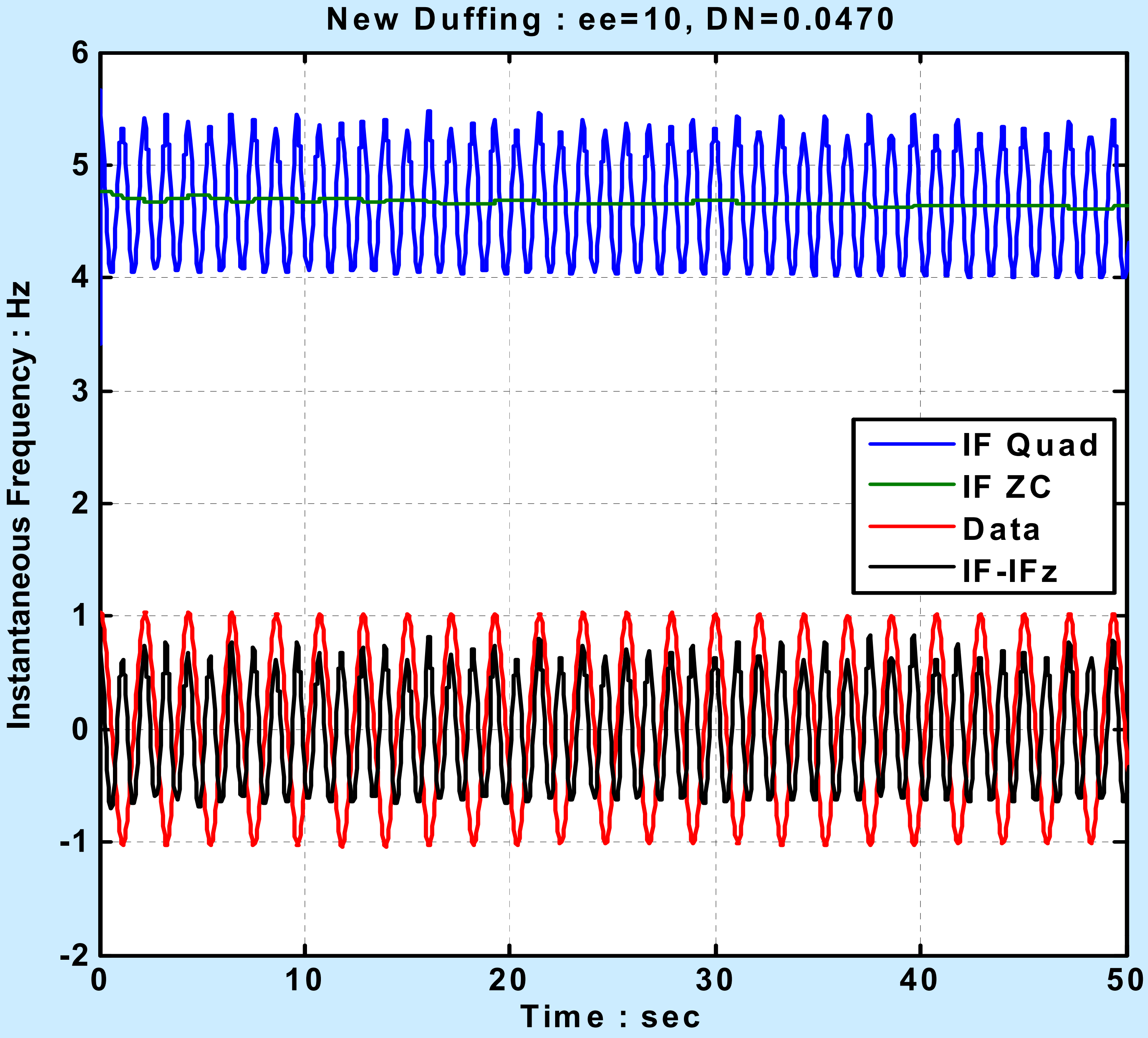
# Stokes ee=-0.375; DN=0.0407



# Duffing Waves

$$\frac{d^2x}{dt^2} + x + \varepsilon x^3 = \gamma \cos \omega t$$

$$\frac{d^2x}{dt^2} + x\left(1 + \varepsilon x^2\right) = \gamma \cos \omega t$$

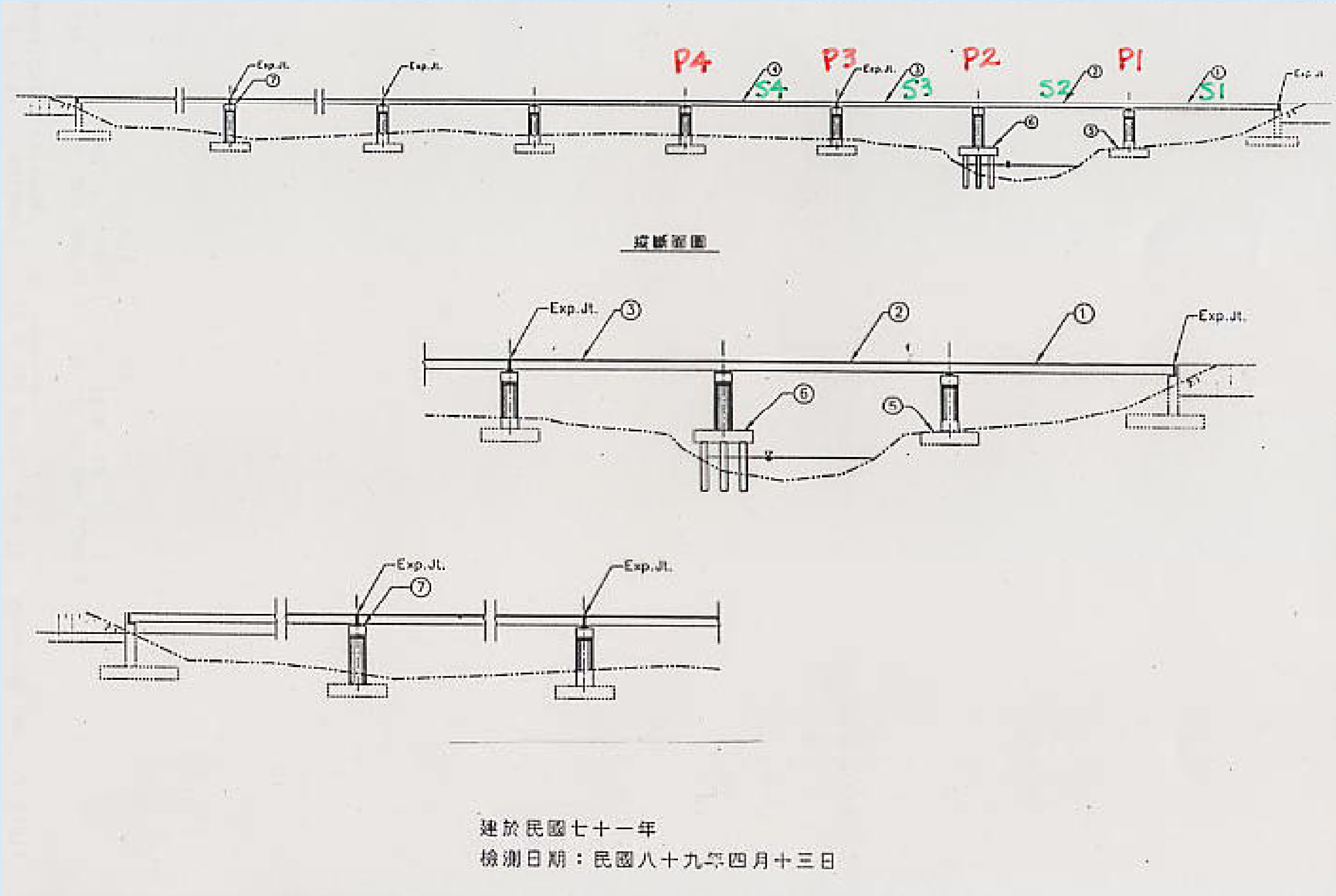


New Duffing : ee=10, DN=0.0470
Instantaneous Frequency : Hz
Time : sec
IF Quad
IF ZC
Data
IF-IFz
6
5
4
3
2
1
0
-1
-2
0
10
20
30
40
50

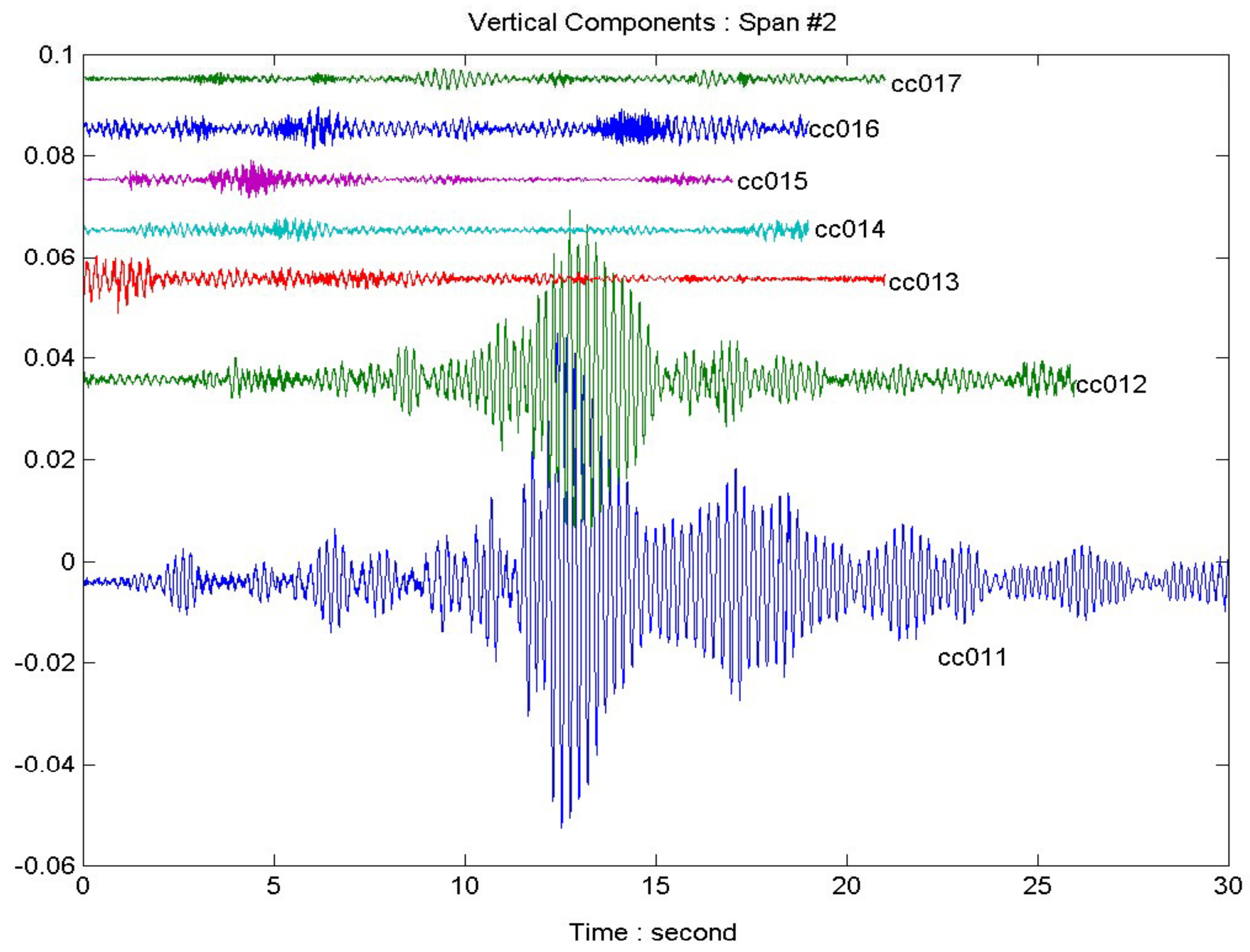
# Case Study

A Bridge in Central Taiwan

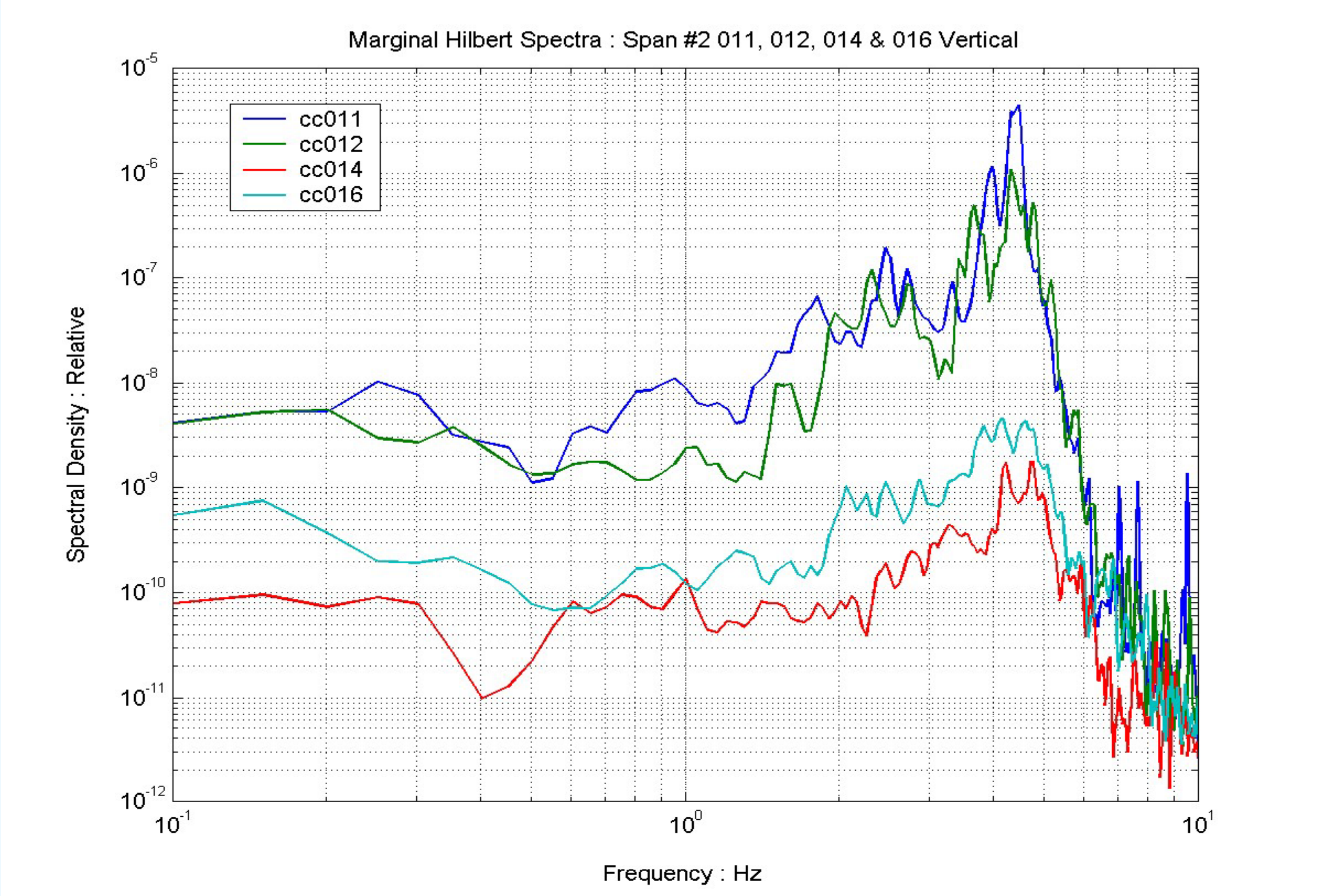
# A Bridge in Southern Taiwan



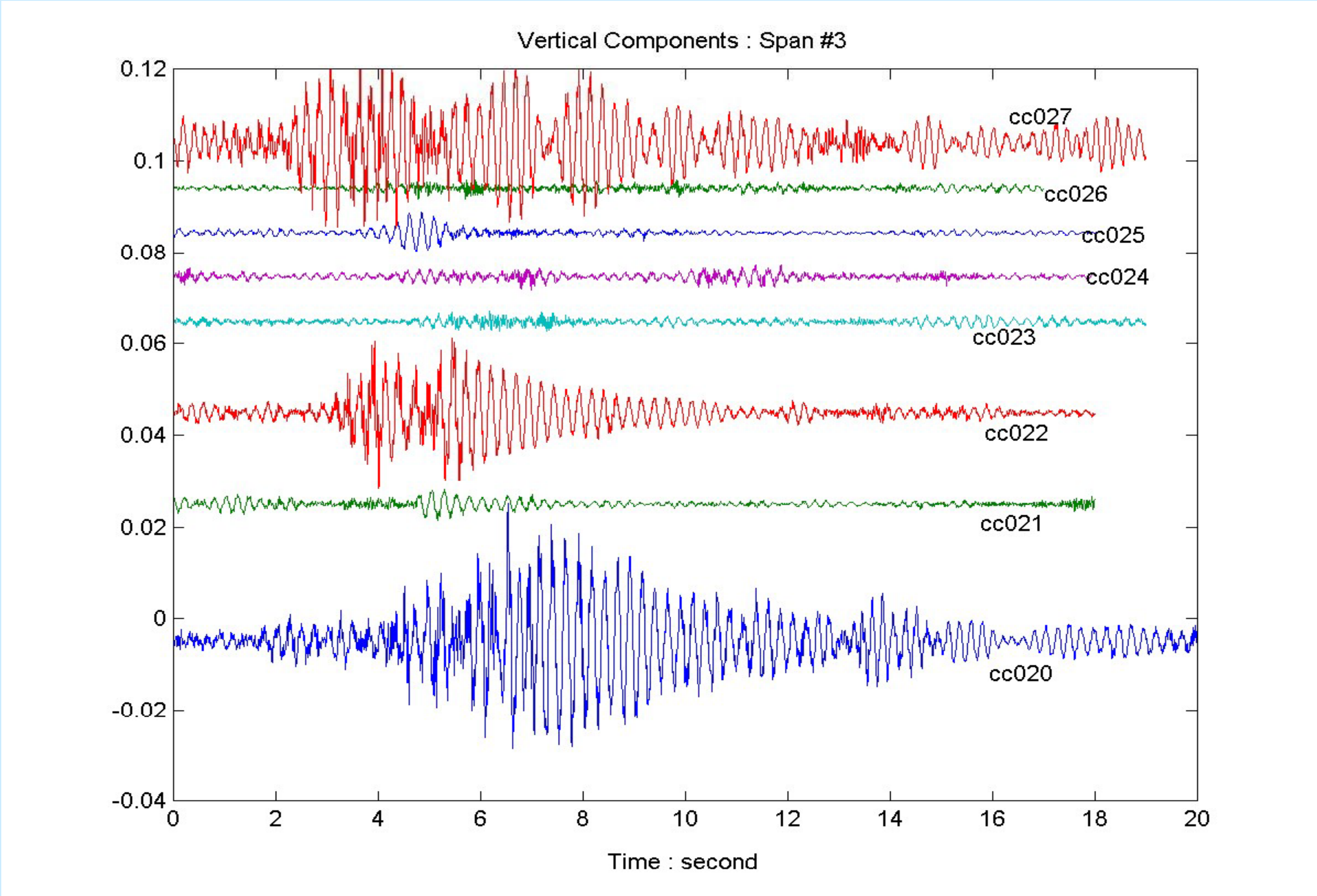
# Data : Hsin Nan Bridge Span 2 Vertical



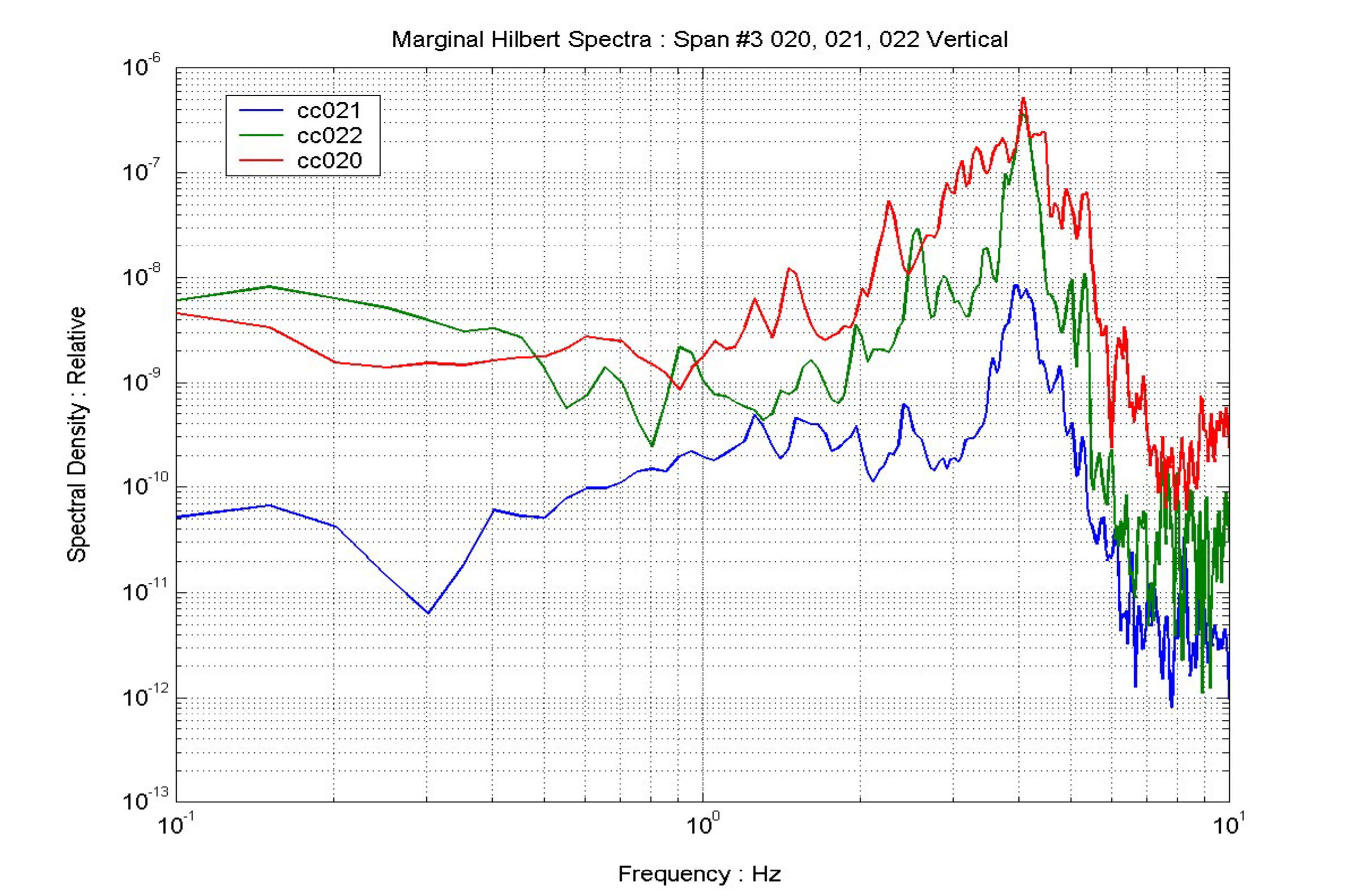
# Marginal Hilbert Spectra : Hsin Nan Bridge Span 2 Vertical



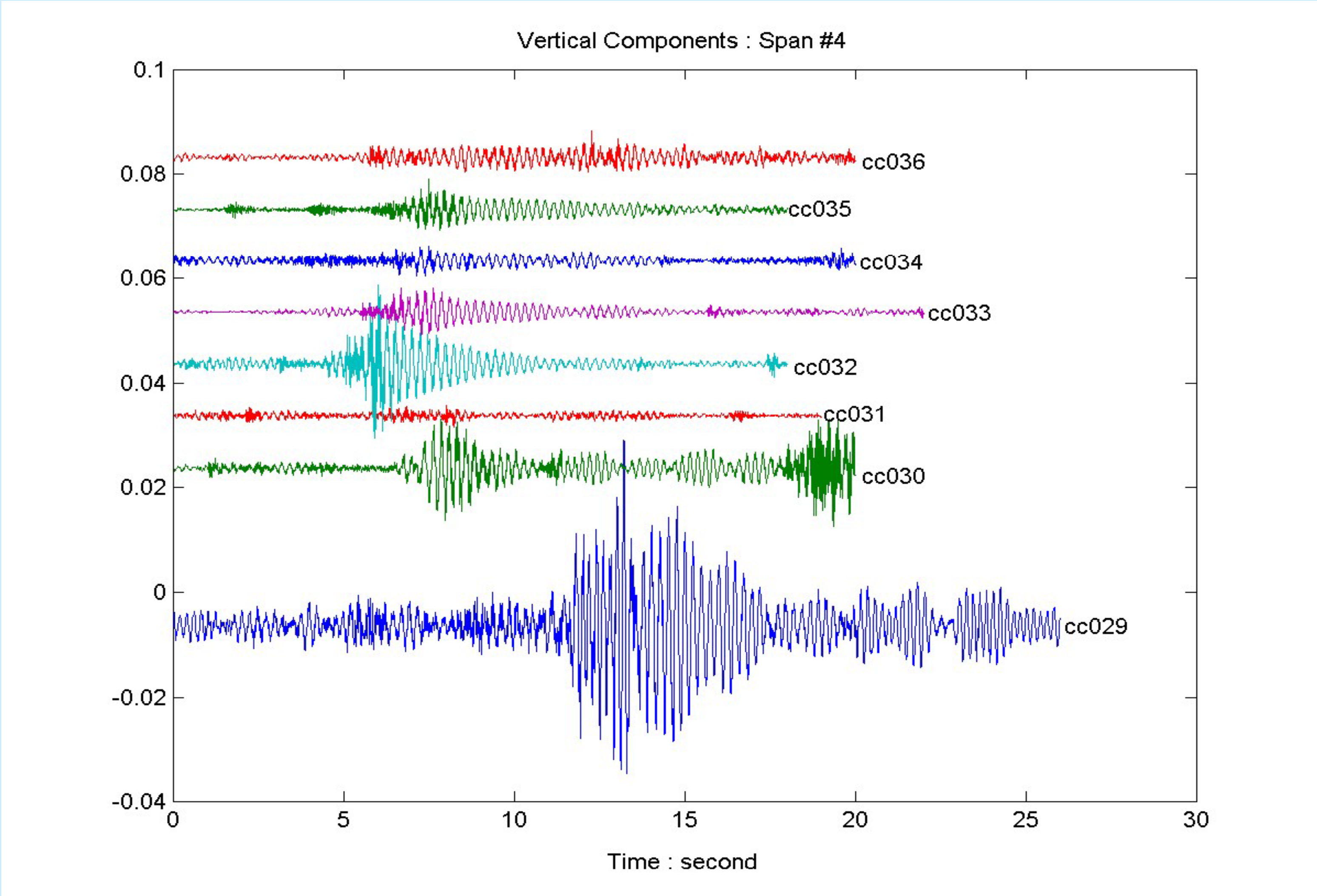
# Data : Hsin Nan Bridge Span 3 Vertical



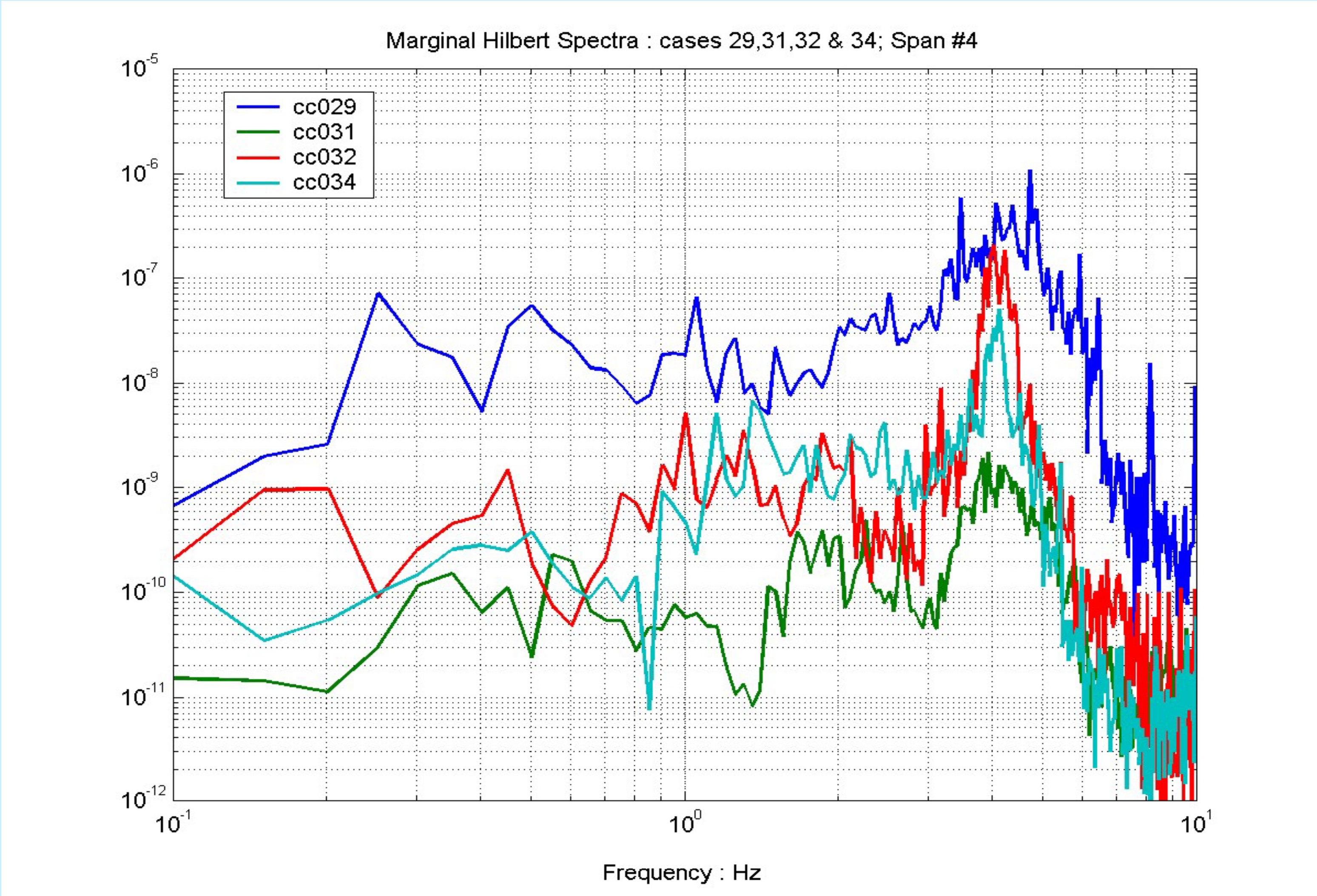
# Marginal Hilbert Spectra : Hsin Nan Bridge Span 3 Vertical



# Data : Hsin Nan Bridge Span 4 Vertical



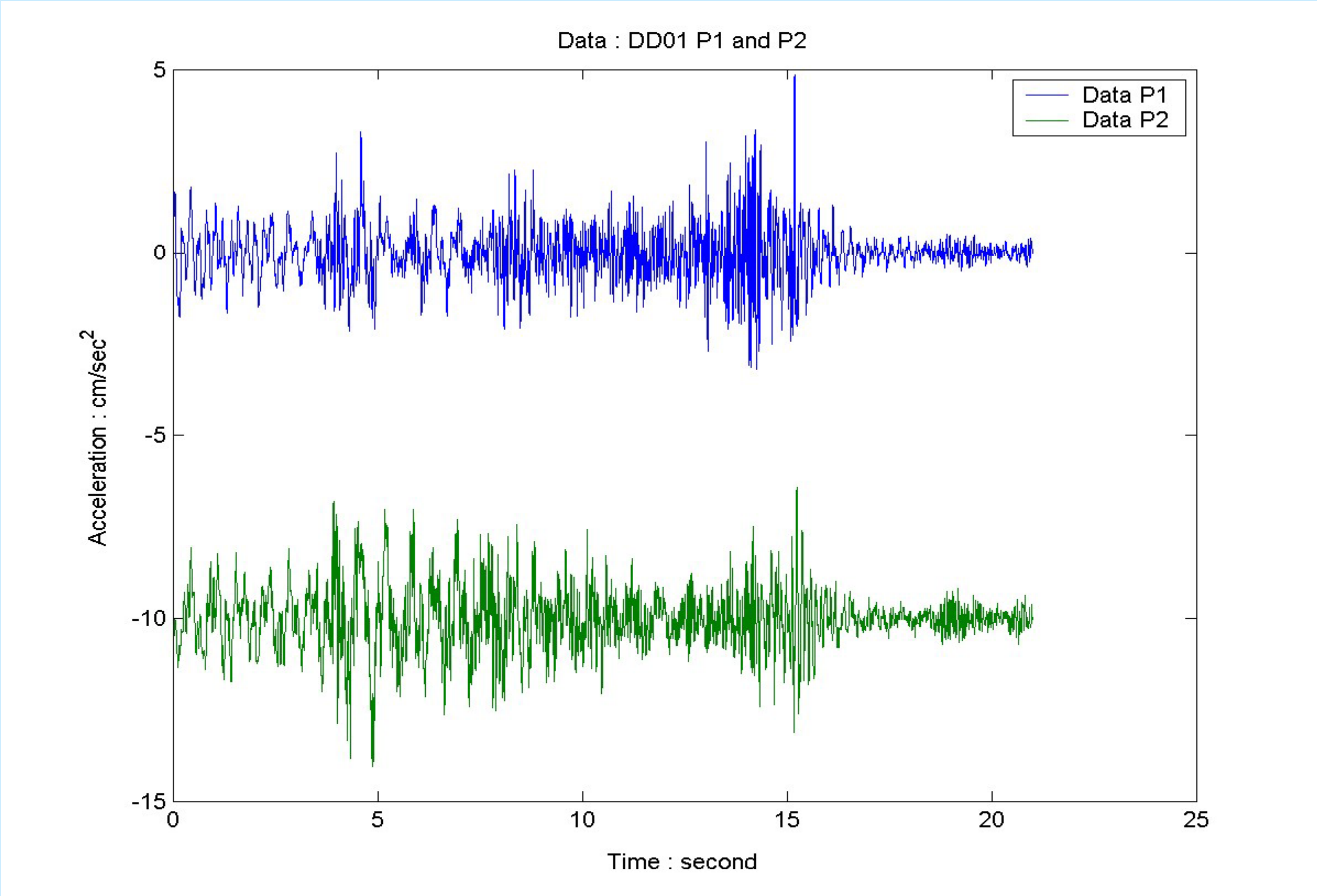
# Marginal Hilbert Spectra : Hsin Nan Bridge Span 4 Vertical



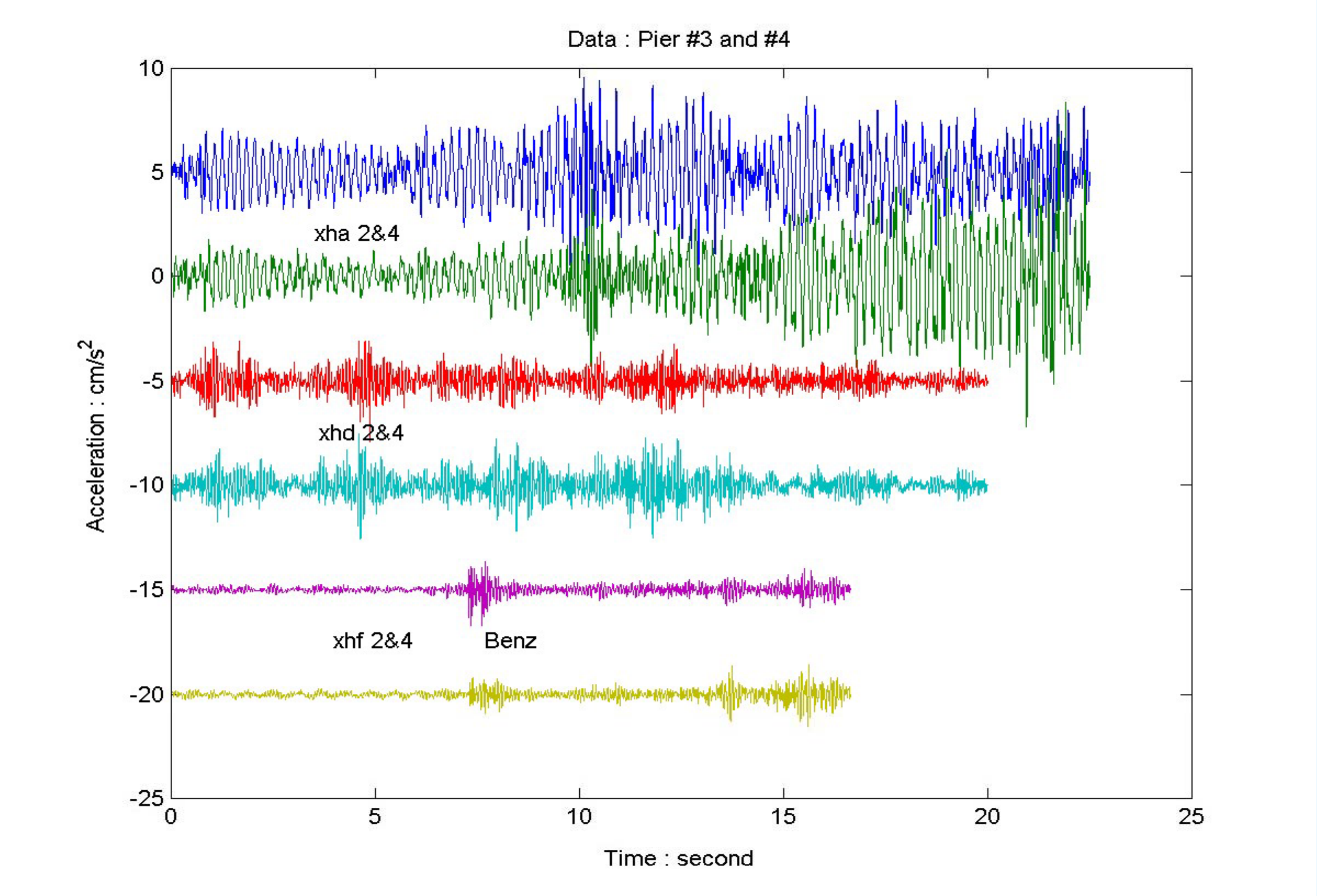
# The foundations: the Piers

Check the piles

# Data : Bridge Piers 1 & 2

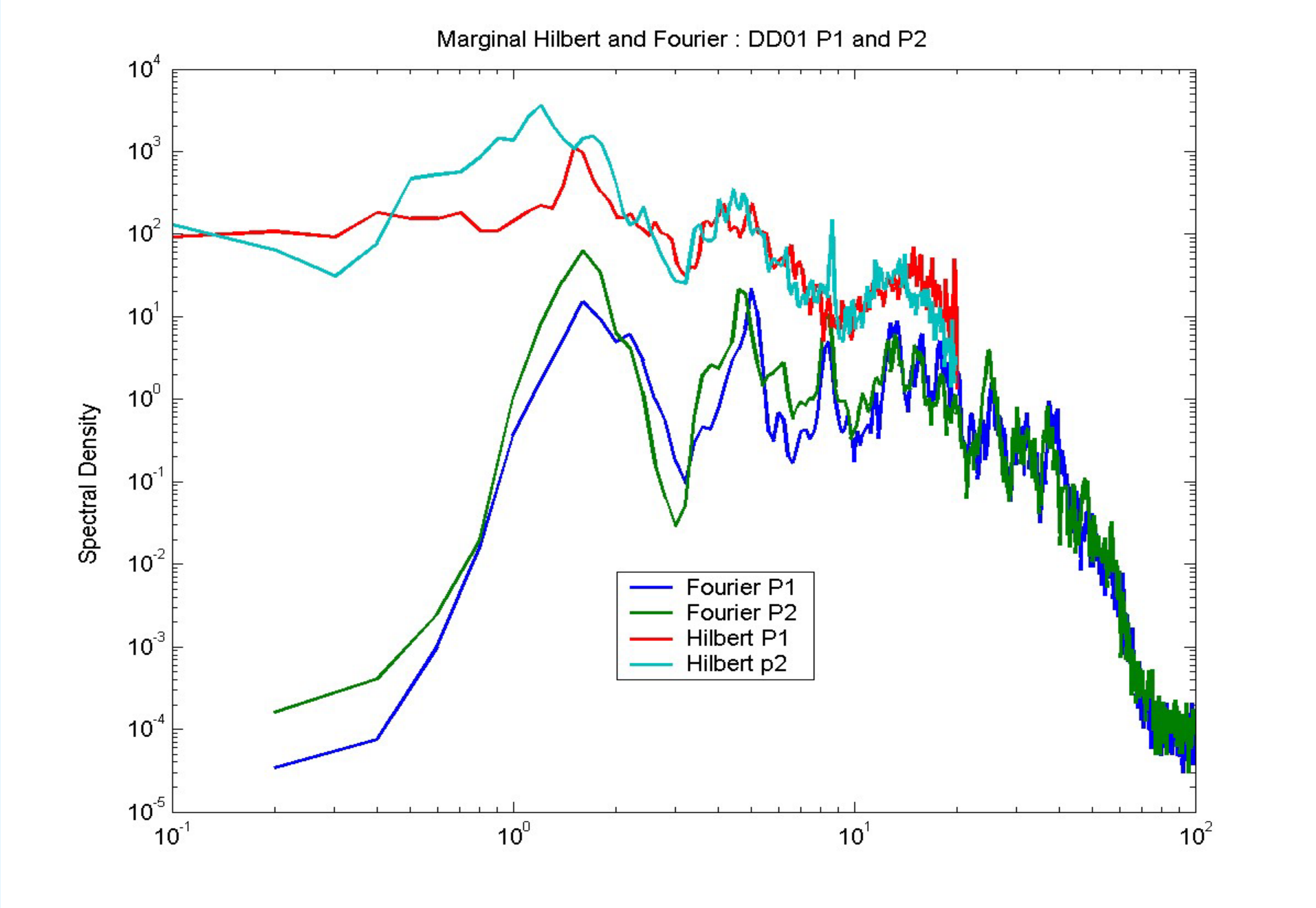


# Data : Bridge Piers 3 & 4



# Marginal Hilbert & Fourier Spectra :

## Bridge Piers 1 & 2 Vertical



# What This Means

- Instantaneous Frequency offers a total different view for nonlinear data: instantaneous frequency with no need for harmonics and unlimited by uncertainty.

- Adaptive basis is indispensable for nonstationary and nonlinear data analysis

- HHT establishes a new paradigm of data analysis with which we can quantify nonlinearity

# Comparisons

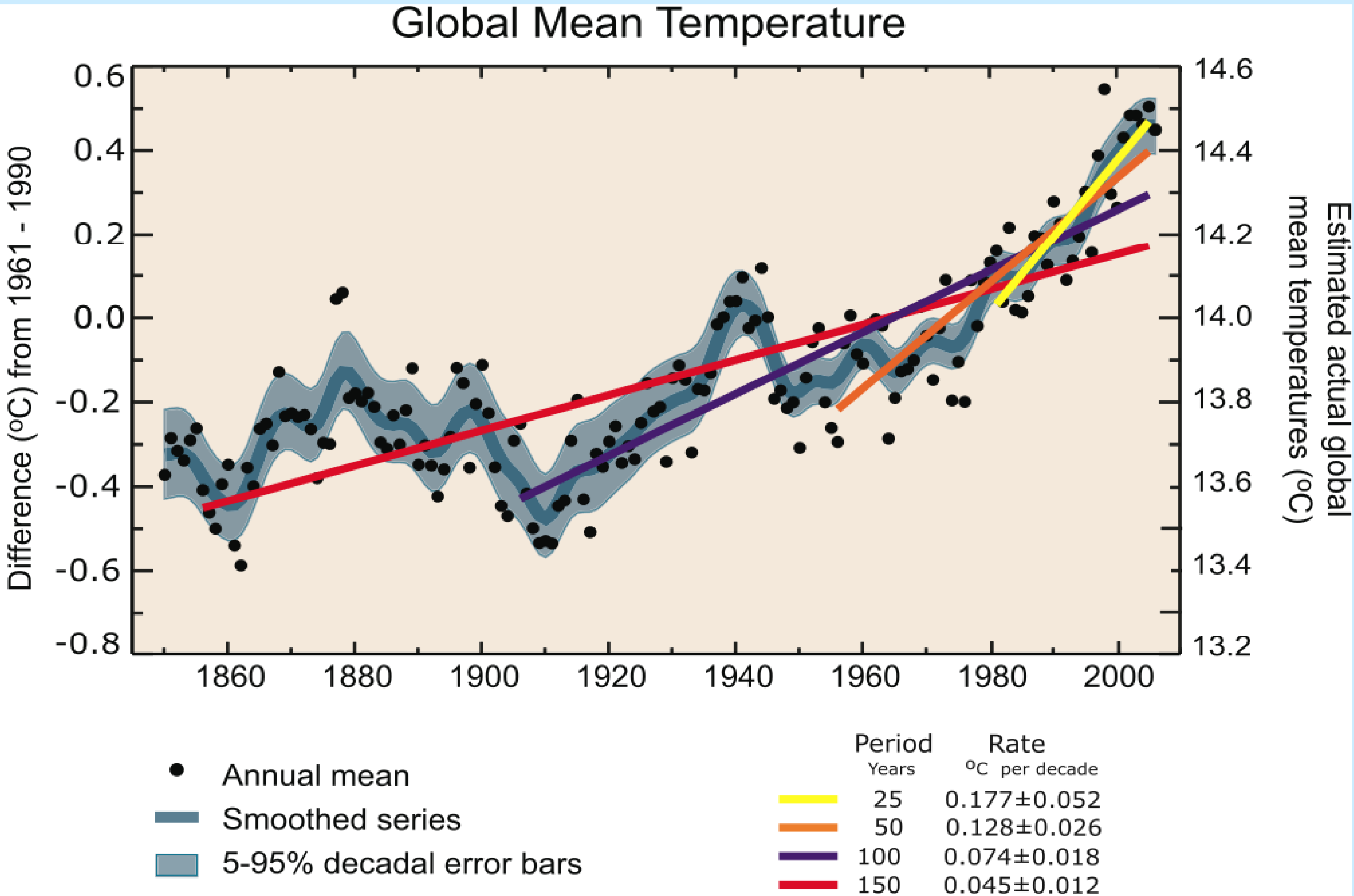
| | Fourier | Wavelet | Hilbert |
|---|---|---|---|
| **Basis** | a priori | a priori | Adaptive |
| **Frequency** | Integral transform: Global | Integral transform: Regional | Differentiation: Local |
| **Presentation** | Energy-frequency | Energy-time-frequency | Energy-time-frequency |
| **Nonlinear**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

# On Hazard Occurrence

Global Warming

HHT: Trend and prediction

# IPCC Global Mean Temperature Trend



*“Note that for shorter recent periods, the slope is greater, indicating accelerated warming.”*

IPCC 4th Assessment Report 2007

# The State-of-the arts: Trend

“One economist’s trend is another economist’s cycle”

Engle, R. F. and Granger, C. W. J. 1991 *Long-run Economic Relationships*. Cambridge University Press.

Regression method is arbitrary and ad hoc.

# Philosophical Problem Anticipated

名不正則言不順

言不順則事不成

——孔夫子

# On Definition

Without a proper definition,
logic discourse would be impossible.
Without logic discourse,
nothing can be accomplished.

**Confucius**

# Definition of the Trend

Huang et al, *Proc. Roy. Soc. Lond.,* 1998
Wu et al. *PNAS* 2007

**Within the given data span, the trend is an intrinsically fitted monotonic function, or a function in which there can be at most one extremum.**

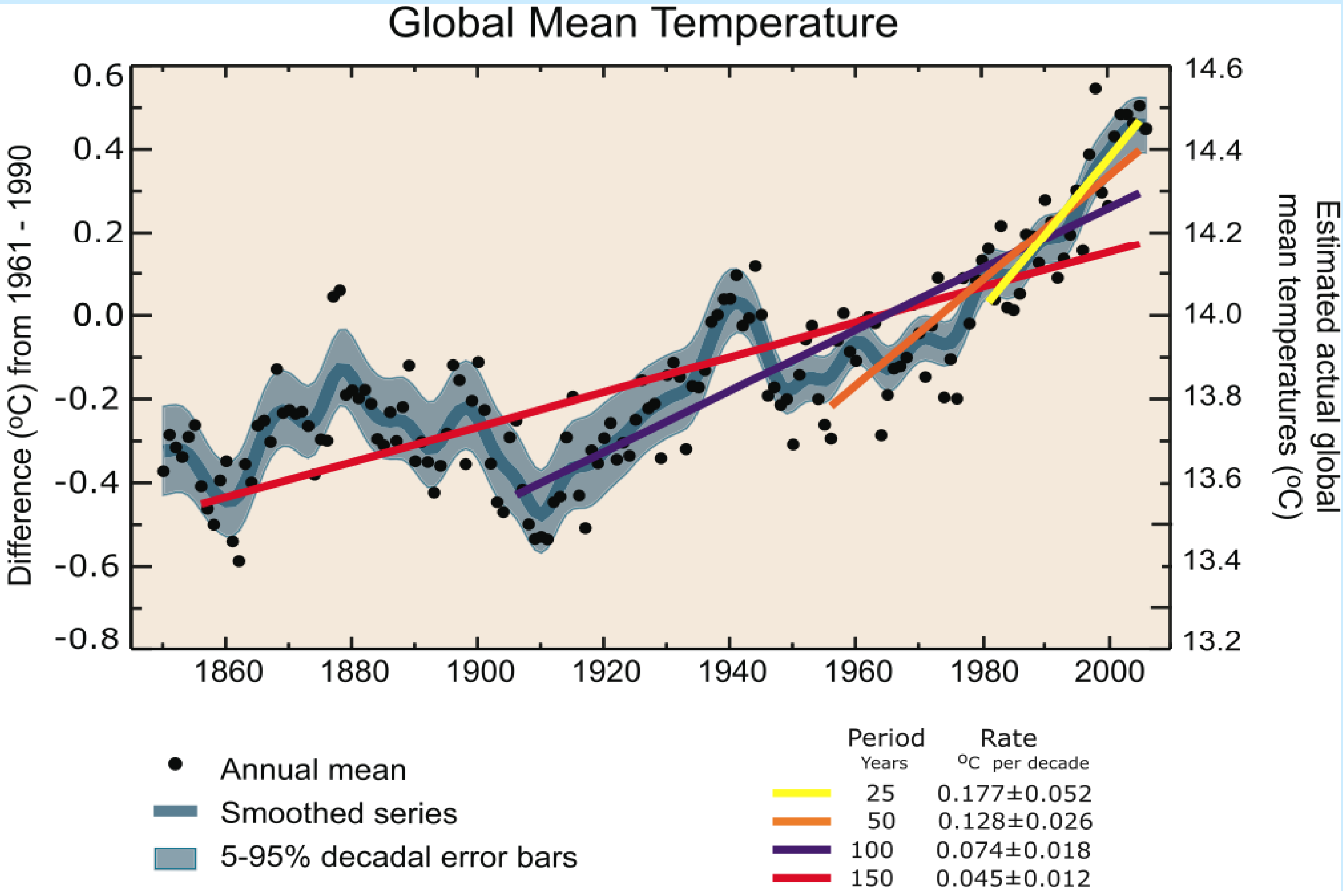
The trend should be an intrinsic and local property of the data; it is determined by the same mechanisms that generate the data.

Being local, it has to associate with a local length scale, and be valid only within that length span, and be part of a full wave length.

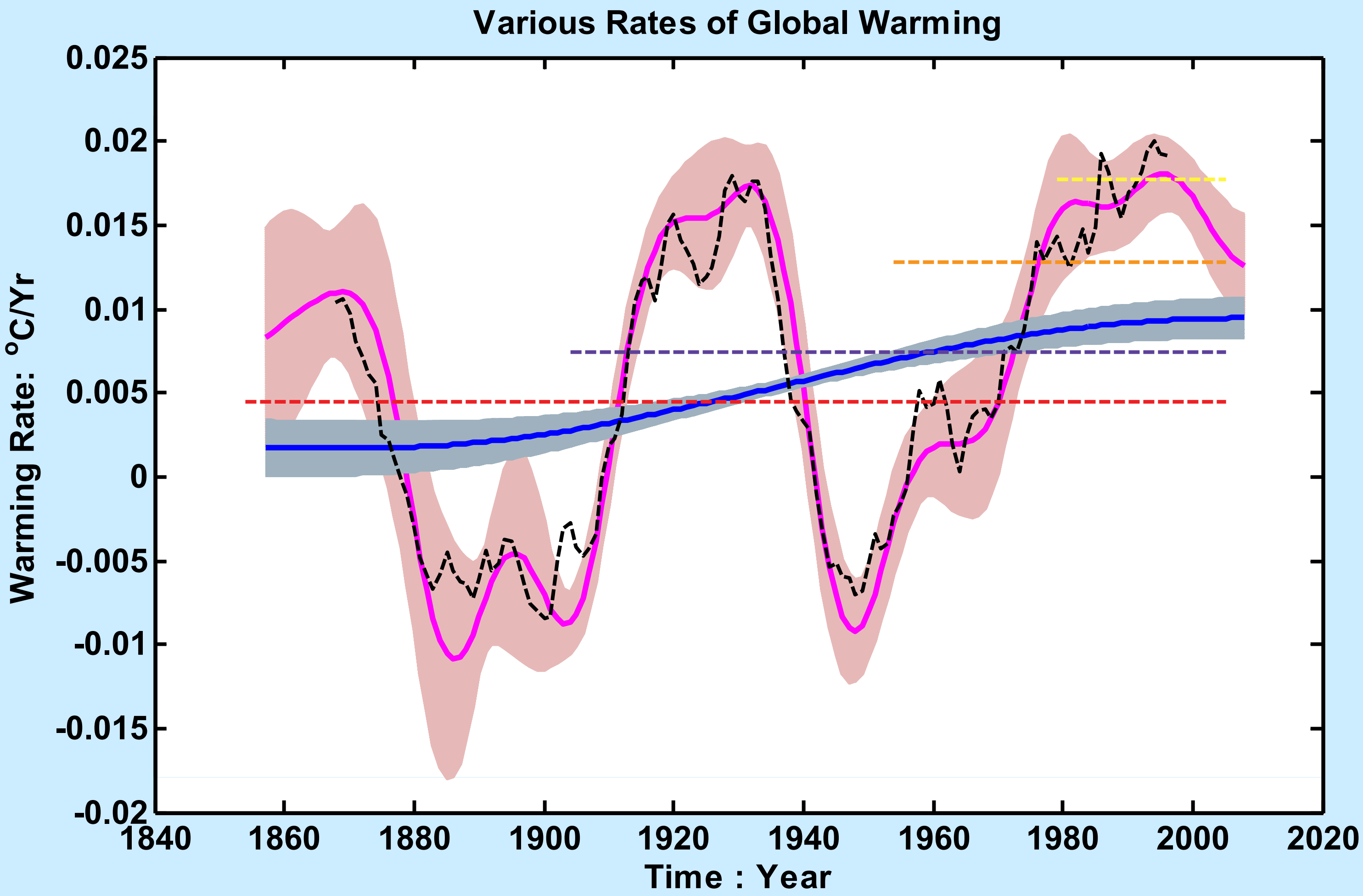
The method determining the trend should be intrinsic. Being intrinsic, the method for defining the trend has to be adaptive.

All traditional trend determination methods are extrinsic.

# IPCC Global Mean Temperature Trend

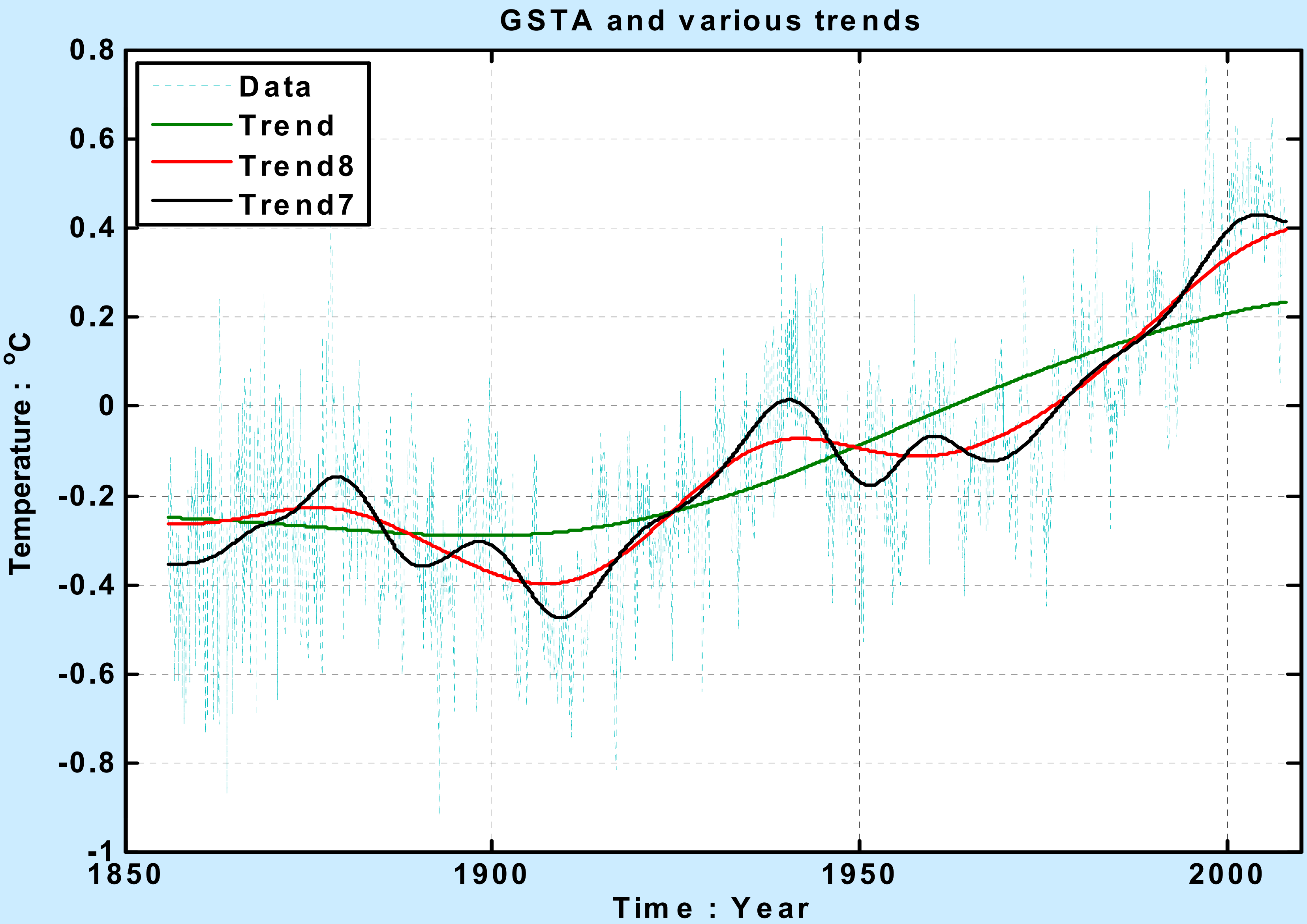


# Comparison between non-linear rate with multi-rate of IPCC




Blue shadow and blue line are the warming rate of non-linear trend.
Magenta shadow and magenta line are the rate of combination of non-linear trend and AMO-like components.
Dashed lines are IPCC rates.

# GSAT Data and Various Trends



# Annual Temperature Ranking : 2008

| GISS | NCDC | CRU | Rank |
|---|---|---|---|
| 2005 | 2005 | 1998 | 1 |
| 1998 | 1998 | 2005 | 2 |
| 2002 | 2002 | 2003 | 3 |
| 2007 | 2003 | 2002 | 4 |
| 2003 | 2006 | 2004 | 5 |
| 2006 | 2007 | 2006 | 6 |
| 2001 | 2004 | 2001 | 7 |
| 2004 | 2001 | 2007 | 8 |
| 2008 | 2008 | 1997 | 9 |
| 1997 | 1997 | 2008 | 10 |

# 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; it is ideal for engineering applications.