Publication of Little Lion Scientific R&D, Islamabad PAKISTAN
Journal of Theoretical and Applied Information Technology
31st July 2011. Vol. 29 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

NON LINEAR AND NON-STATIONARY DATA ANALYSIS USING HILBERT-HUANG TRANSFORM

[1]Ch. HEMANTH*, [2]A. SIVASAI RAM , [3]N. RAMA KRISHNA, and [4]P. S. BRAHMANANDAM
[1]Dept. of ECE, K L University, Green fields, Vaddeswaram 522 502, India
[2]Dept. of ECE, K L University, Green fields, Vaddeswaram 522 502, India
[3]Dept. of ECE, K L University, Green fields, Vaddeswaram 522 502, India
[4]Dept. of ECE, K L University, Green fields, Vaddeswaram 522 502, India

ABSTRACT

The newly developed method known as Hilbert-Huang Transform (HHT) is ideal for nonlinear and non-stationary data analysis as it is totally adaptive in nature. This paper will discuss the fundamentals of HHT method which consists of the empirical mode decomposition and Hilbert spectral transform. As a part of analysis, the HHT method is applied on two different data sets in which one is the annual mean global surface temperature anomaly and the other is the noise component measured by the Equatorial Atmosphere Radar located in Indonesia. The analysis of two data sets indicates that the HHT method is able to extract each and every frequency component, which might not be possible with the Fourier spectral analysis. Specifically, this study indicates that the decomposed components in EMD of HHT, namely the intrinsic mode function components contain observable, physical information inherent to the original data. Finally, the study illustrates that the HHT-based Hilbert spectra are able to reveal the time-frequency distributions more precisely.

Key words: *Nonlinear and non-stationary signals, Fourier spectral analysis, Empirical Mode Decomposition (EMD), Hilbert Spectral Analysis (HAS), Instantaneous frequency*

1. INTRODUCTION

Data analysis is a crucial part in pure research and real time applications. Data analysis mainly contributes to design the parameters that are required to construct a model as well as in verifying the viability of the constructed model. Unfortunately most of the data from physical measurements or numerical modeling most likely will have some problems like: (*a*) the total data span is too short; (*b*) the data are non-stationary; and (*c*) the data represent nonlinear processes. Although each of the above problems can be real by itself, the first two are related, for a data section shorter than the longest time scale of a stationary process can appear to be non-stationary. Facing such data, we have limited options to use in the analysis.

2. CONVENTIONAL METHODS FOR DATA ANALYSIS

Historically, Fourier spectral analysis has provided a general method for examining the global energy-frequency distributions. As a result, the term `spectrum' has become almost synonymous with the Fourier transform of the data. Because of its prowess and simplicity, Fourier analysis has dominated the data analysis efforts since soon after its introduction, and has been applied to all kinds of data. Although the Fourier transform is valid under extremely general conditions (see, for example, Titchmarsh 1948), there are some crucial

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

restrictions of the Fourier spectral analysis: the system must be linear; and the data must be strictly periodic or stationary [1]; otherwise, the resulting spectrum will make little physical sense. Unfortunately most of the practical data are nonlinear and non-stationary. So we need better methods to analyze the so-called data.

3. HILBERT-HUANG TRANSFORM

The Hilbert-Huang Transform (HHT) developed by Huang et al. (1998, 1999); can represent non-stationary nonlinear data efficiently. The HHT, which consists of two steps— empirical mode decomposition (EMD) and Hilbert spectral analysis (HSA), is summarized by Huang et al. (1998, 1999).

3.1 Empirical Mode Decomposition

The EMD step builds on the assumption that any data set consists of different, simple, intrinsic modes of oscillation that need not be sinusoidal, with the non sinusoidal character of each mode of oscillation derived from the data. At any given time, the recorded data may have many different coexisting modes of oscillation, which may or may not relate to different seismological phases. Each of these oscillatory modes, called an intrinsic mode function (IMF), is defined by the following conditions:

1. In the whole data set, the number of extrema and the number of zero crossings must either equal or differ at most by one; 2. At any point, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero.

An IMF represents a simple oscillatory mode similar to a component in the Fourier-based simple harmonic function. One can decompose any waveform as follows. First, identify all the local extrema. Connect all local maxima by a cubic spline to produce the upper envelope, and repeat the procedure for the local minima to produce the lower envelope. The upper and lower envelopes should encompass all the data between them. The mean of these two envelopes is designated as m_1, and the difference between the data X and m_1 is the first component h_1; i.e.

$$X(t) - m_1 = h_1 \quad (1)$$

Ideally, h_1 should be an IMF, since the construction of h_1 described above should have made it satisfy all the conditions set in the definition of an IMF. Yet, in practice, all the conditions of an IMF cannot be achieved until the previous process, called the *sifting process*, is repeated. In the subsequent sifting process, h_1 is treated as the data, then

$$h_1 - m_{11} = h_{11} \quad (2)$$

where m_{11} is the mean of the upper and lower envelopes of h_1. After repeated shifting, up to k times which is usually less than 10, h_{1k} *given* by

$$h_{1(k-1)} - m_{1k} = h_{1k} \quad (3)$$

is designated as the first IMF component c_1 from the data, or

$$c_1 = h_{1k}; \quad (4)$$

Typically, $c1$ will contain the finest-scale or the shortest-period component of the signal. One can remove $c1$ from the rest of the data to obtain the residue

$$X(t) - c_1 = r_1 \quad (5)$$

This sifting process has two effects: (*a*) to eliminate riding waves; and (*b*) to smooth uneven amplitudes. The residue r_1, which contains longer-period components, is treated as the new data and subjected to the same sifting process as described above. This procedure can be repeated to obtain all the subsequent r_j functions as follows:

$$r_1 - c_2 = r_2 \ldots rn-1 - c_n = r_n \quad : \quad (6)$$

The sifting process can be terminated by either of the following predetermined criteria: (1) either the component c_n or the residue r_n becomes so small that it is less than a predetermined value of consequence; and (2) the residue r_n becomes a monotonic function, from which no more IMF can be extracted. If the data have a trend, the final residue will be that trend. The original data are thus

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

the sum of the IMF components plus the final residue

$$X(t) = \sum_{j=1}^{n} c_j + r_n \qquad (7)$$

Thus, the data are decomposed into *n* IMF components and a residue r_n that can be either the mean trend or a constant.

3.2 Hilbert Spectral Analysis

For given data, *X (t)*, the Hilbert transform, *Y (t)*, is defined as

$$\mathrm{Y}(t) = \frac{1}{\pi} P \int \frac{C(t')}{t-t'} dt \qquad (8)$$

where *P* denotes the Cauchy principal value. With this definition, *C* (*t*) and *Y* (*t*) can be combined to form analytical signal *Z* (*t*), given by

$$Z(t) = X(t) + iY(t) = a(t)\, e^{i\Theta}(t); \qquad (9)$$

where time-dependent amplitude *a* (*t*) and phase *Θ* (*t*) are found as

$$a(t) = [X_2(t) + Y_2(t)]^{1/2}; \qquad (10)$$

$$\theta(t) = \arctan(\mathrm{Y}(t)/\mathrm{X}(t)) \qquad (11)$$

From the polar coordinate expression, the instantaneous frequency can be defined

$$\omega = \frac{d\theta}{dt} \qquad (12)$$

Applying the Hilbert transform to the *n* IMF components, the data *X* (*t*) can be written as

$$\mathrm{X}(t) = \mathrm{R} \sum_{j=1}^{n} a_j(t) e^{i\Theta\int \omega_j(t)dt} \qquad (13)$$

where R is the real part of the value to be calculated and a_j – the analytic signal associated with the *j*th IMF. The residue *rn* is not included because of its monotonic property (Huang et al. 1998). The above Equation (13) is written in terms of amplitude and instantaneous frequency associated with each component as functions of time, which differ from the time-independent amplitude and phase in the Fourier series representation of

$$\mathrm{X}(t) = \mathrm{R} \sum_{j=1}^{n} A_j e^{i\Omega_j t} \qquad (14)$$

where A_j=Fourier transform of *X (t)*, a function of frequency Ω_j. A comparison of the two representations in Equations (13) and (14) suggests that the Hilbert transform of the IMF can be considered as a generalized Fourier expansion. The time-dependent amplitude and instantaneous frequency in above equation might not only improve the flexibility of the expansion, but also enable the expansion to accommodate non-stationary data. The frequency-time distribution of the amplitude is designated as the Hilbert amplitude spectrum, *H* (ω, *t*), or simply Hilbert spectrum, defined as

$$H(\omega, t) = \sum_{j=1}^{n} \tilde{H}_j(\omega, t) \equiv \sum_{j=1}^{n} a_j(t) \qquad (15)$$

where ≡ denotes ''by definition'' and $\tilde{H}_j$=*j*th component of the total Hilbert spectrum *H*. The square of *H* reveals the evolutionary energy distribution or energy density. The marginal spectrum, *h* (ω), defined as

$$h(\omega) = \sum_{j=1}^{n} \tilde{h}_J(\omega) \equiv \sum_{j=1}^{n} \int_0^T a_{jd}(t)dt \qquad (16)$$

provides a measure of total amplitude contribution from each frequency value, in which *T* denotes the time duration of data. It should be noted that the Hilbert transform described in Equations. (8)– (12) is not new. However, the incorporation of the Hilbert transform into the IMF components and thus the HHT representation of data in Eq. (13) are entirely novel. Huang et al. (1998) show that the instantaneous frequency has physical meaning only through its definition on each IMF component; by contrast, the instantaneous frequency defined through the Hilbert transform of the original data might be less directly related to frequency content because of the violation of the mono component condition on the Hilbert transform.

4. APPLICATION OF HHT ANALYSIS

Publication of Little Lion Scientific R&D, Islamabad PAKISTAN
Journal of Theoretical and Applied Information Technology
31st July 2011. Vol. 29 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

In this section, we used annual mean global surface temperature anomaly data from one of the internet based data to illustrate the HHT analysis potential in nonlinear and non-stationary data analyses. It is also used a database on the noise component that present in the signal at various time intervals which is taken from the Equatorial Atmosphere Observatory website. The HHT calculation is carried out in MATLAB.

4.1 HHT on annual mean global surface temperature anomaly

Fig 1 represents the annual mean global temperature anomaly belongs to years from 1850 to 2010. It is clear that the data is nonlinear and non-stationary. In order to show the superiority of HHT over traditional data analysis method, it was calculated the Fourier spectrum for this data. Fig 2 represents the Fourier spectrum of the above data calculated by using FFT algorithm. Fourier spectrum explored all the frequency components present in the data corresponding to their respective amplitude

FIGURE 1 location

FIGURE 2 location

There are two major problems with this spectral analysis as pointed out in below

1) Fourier spectral analysis is priory based analysis i.e. it explains data in terms of a superposition of trigonometric functions. If the actual data is very far from this assumption, then its leads to loss of important spectral components.

2) Fourier spectral analysis doesn't give any information regarding frequency components at instant time which is most important for many real time analyses.

So, Fourier spectral analysis is not completely viable to analyze this type of data.

HHT is applied to the same data. Fig 3 represents the Empirical Mode Decomposition (EMD) of the given data. The entire data can be split into 4 IMFs and a residue also known as trend whose behavior is explained in the above section.

FIGURE 3 location

FIGURE 4 location

We can observe all IMFs are non-overlapping. Splitting the data into finite number of data sets makes the spectral analyses easier. The Hilbert spectrum in HSA shows a clear picture of temporal-frequency energy distribution. Fig 4 represents the Hilbert spectrum of 3rd IMF. From this figure, it is possible to know the instantaneous frequency of that particular IMF. Since the HHT speaks only about data without any pre assumptions, no frequency component gets missed. So it's entirely reliable and adaptive in nature.

4.2 Application of HHT on Equatorial Atmospheric Radar data

We now demonstrate the efficiency of HHT by using the data measured by the Equatorial

Publication of Little Lion Scientific R&D, Islamabad PAKISTAN
Journal of Theoretical and Applied Information Technology
31st July 2011. Vol. 29 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Atmospheric Radar located in Indonesia. Fig 5 represents the noise level at different intervals of time. It is obvious to note that the above data is nonlinear and non-stationary in nature. We have also checked the efficiency of the Fourier spectral analysis in this case as we did for the global temperature anomaly data. As mentioned above, same drawbacks can also be seen here. In addition, Fourier spectrum does not provide information that is specific to the localized time scale (Figure 6). But HHT overcomes those problems.

FIGURE 5 location

FIGURE 6 location

Clearly, the data are associated with quite complicated local extrema, but no zero crossings are present. The mean can be treated as a zero reference, although defining it is hard, for the whole process it is transient. This example illustrates the advantage of adopting the successive extrema for defining the time scale; it also illustrates the difficulties of dealing with non-stationary data and even a meaningful. Comparing this with the traditional Fourier expansion, one can immediately see the efficiency of the EMD: the expansion of a turbulence data set with only six terms. From the result, one can see a general separation of the data into locally non-overlapping time scale components. Figure 8 represents the corresponding instantaneous frequencies of respective IMFs.

FIGURE 7 location

FIGURE 8 location

. All the IMFs are non-overlapping as mentioned earlier. Further, as we don't assume the data in terms of predefined functions, most of the data can be reconstructed. Thus, it is possible to reconstruct the entire data from these IMFs including residue with an accuracy of 99.99%. It is observed that the waveform gets smoothened as IMF number increases. It is also obvious that the IMF components have provided a variable amplitude and frequency representation. However, in some components the signals are intermittent then the neighboring components might contain oscillations of the same scale. But signals of the same time scale would never occur at the same locations in two different IMF components. This allows us to remove the unwanted data according to its corresponding instantaneous frequency that is larger than a presumed threshold value. If we observe first IMF it constitutes higher frequency components but the amplitude is small when compared with the remaining IMFs. The next IMFs have a bit bigger amplitude with larger periods. It is quite reasonable to conclude that the HHT is a better method in analyzing nonlinear and non-stationary data.

5. CONCLUSIONS

HHT offers a potentially viable method for nonlinear and non-stationary data analysis, especially for time-frequency representations when compared with the traditional data analysis methods. Due to its adaptive natured analysis, the results are more accurate than any other traditional analysis methods. The IMF components contain the observable, physical information inherent to the original data and HSA provides us with the time-frequency relations which are very crucial for any data analysis. From the discussed examples, it proves the fact that this method is the best alternative for analyzing nonlinear and non-stationary signals. HHT has wide applications in geophysics, image analysis, oceanography and radar data analysis etc.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

ACKNOWLEDGEMENTS

The writers gratefully acknowledge support, in part, from the management of Koneru Lakshmaiah University, Vaddeswaram, India. We also would like to thank Research Institute for Sustainable Humanosphere of Kyoto University for providing us the EAR data.

REFERENCES

Huang, N. E., Tung, C. C. & Long, S. R. 1990*a* Wave spectra. *The Sea* 9, 197-237.

Huang, N. E., Tung, C. C. & Long, S. R. 1990*b* the probability structure of the ocean surface. *The Sea* 9, 335-366.

N. E. Huang, Z. Shen, S. R Long, et al, "The Empirical Mode Decomposition and Hilbert Spectrum for Non-linear and Non-stationary Time Series Analysis", *Proc.* R. *Soc. London A, 454,* 1998

N. E. Huang, Z. Shen, *S.* R Long, "A new view of nonlinear water waves: The Hilbert spectrum", *Annu. Rev. Fluid Mech. 31,* 1999, pp.417-457.

Priestley, M. B. (1988). *Nonlinear and non-stationary time series analysis*, 1st Ed., Academic, San Diego.

Shekel, J. 1953 `Instantaneous' frequency. *Proc. IRE* 41, 548.

Titchmarsh, E. C. 1948 *Introduction to the theory of Fourier integrals*, Oxford University Press.

Wu, Z., and N. E. Huang, 2005: Statistical significant test of intrinsic mode functions. In *Hilbert-Huang Transform : Introduction and Applications*, pp 125-148, Ed. N. E. Huang and S. S. P. Shen, World Scientific, Singapore, 360pp.

***Corresponding author**: Ch. Hemanth (hemanthchimata@gmail.com)

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

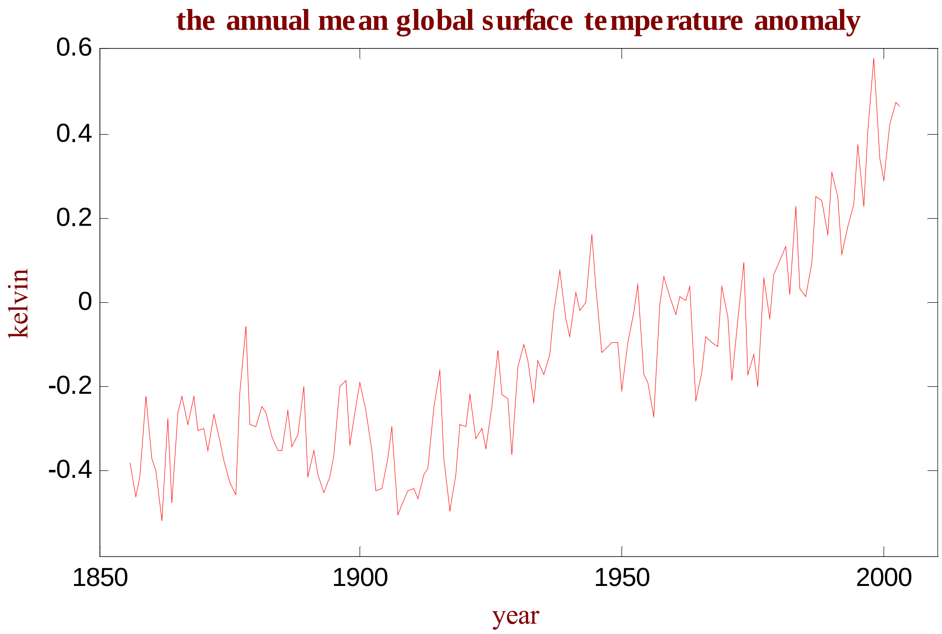

Figure 1: Annual mean global temperature anomaly data (courtesy from www.rcada.ncu.edu.tw).

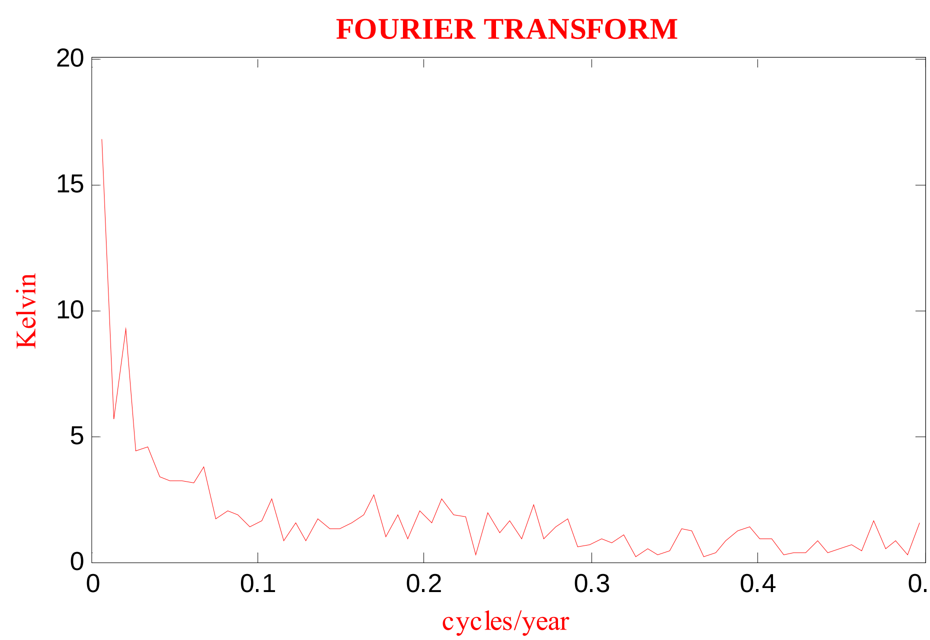

Figure 2: Fourier Transform of temperature data

Publication of Little Lion Scientific R&D, Islamabad PAKISTAN
Journal of Theoretical and Applied Information Technology
31st July 2011. Vol. 29 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

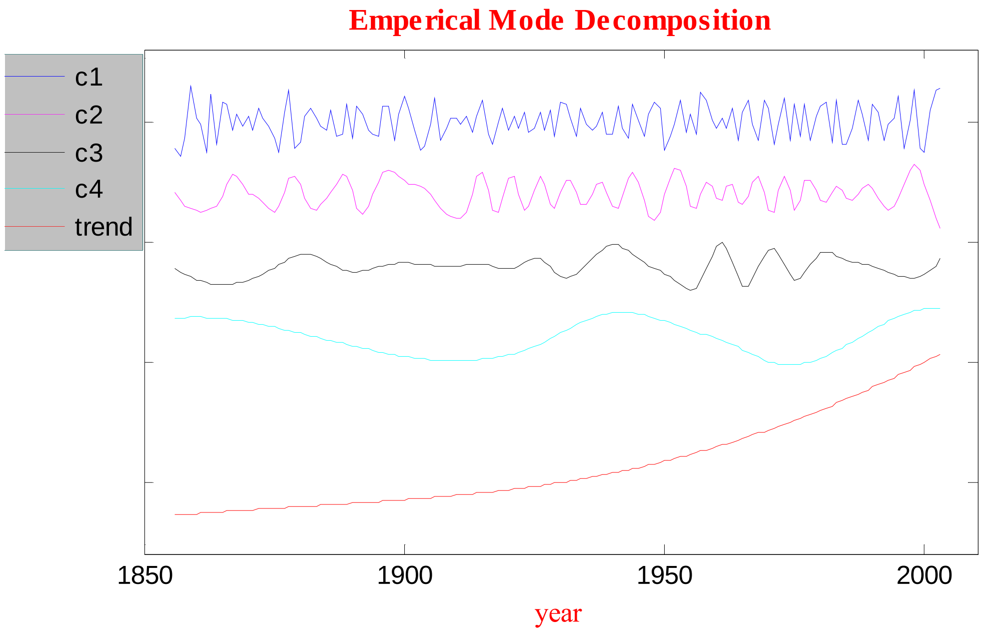

Figure 3: Empirical Mode Decomposition of temperature data

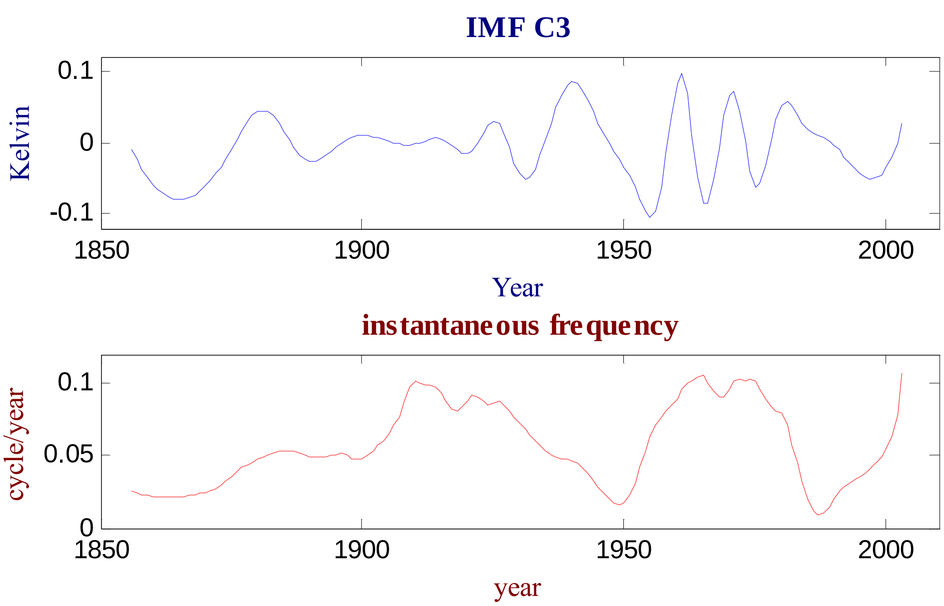

Figure 4: The IMF 3 (upper panel) and its instantaneous frequency (lower panel)

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

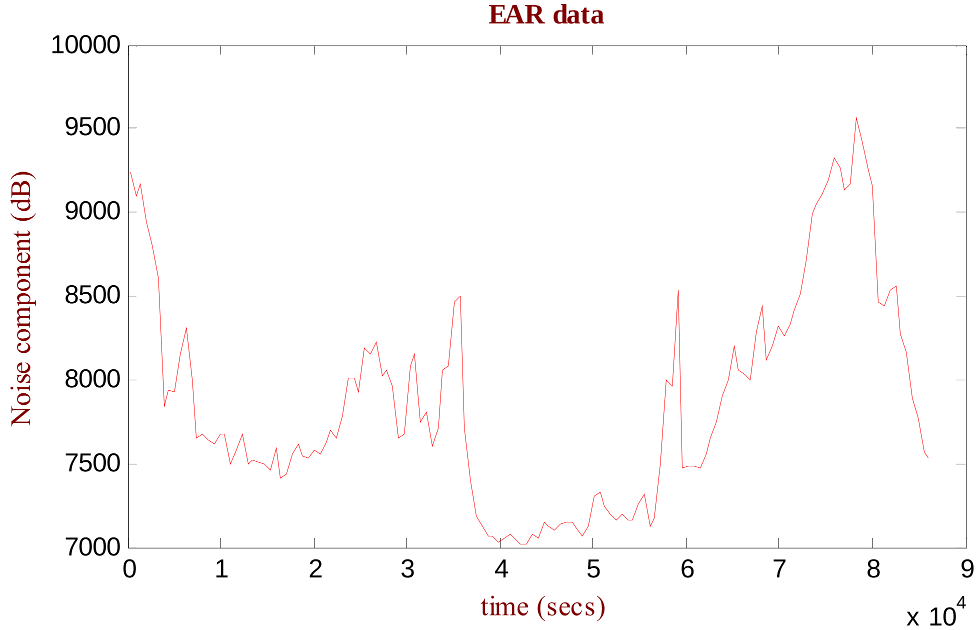

Figure 5: Noise level of signal measured by the Equatorial Atmospheric Radar

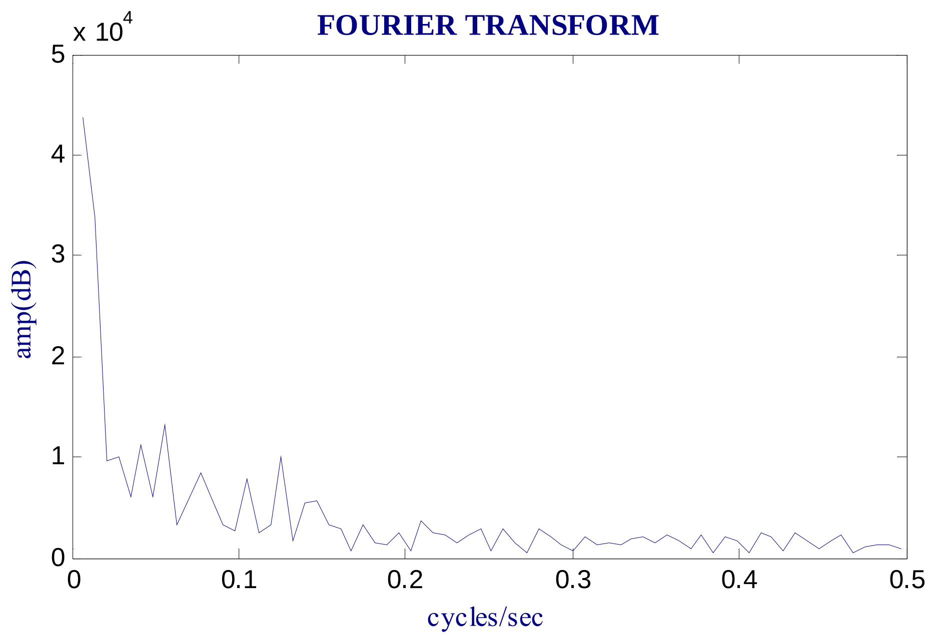

Figure 6: Fourier transform of EAR data

Publication of Little Lion Scientific R&D, Islamabad PAKISTAN
Journal of Theoretical and Applied Information Technology
31st July 2011. Vol. 29 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

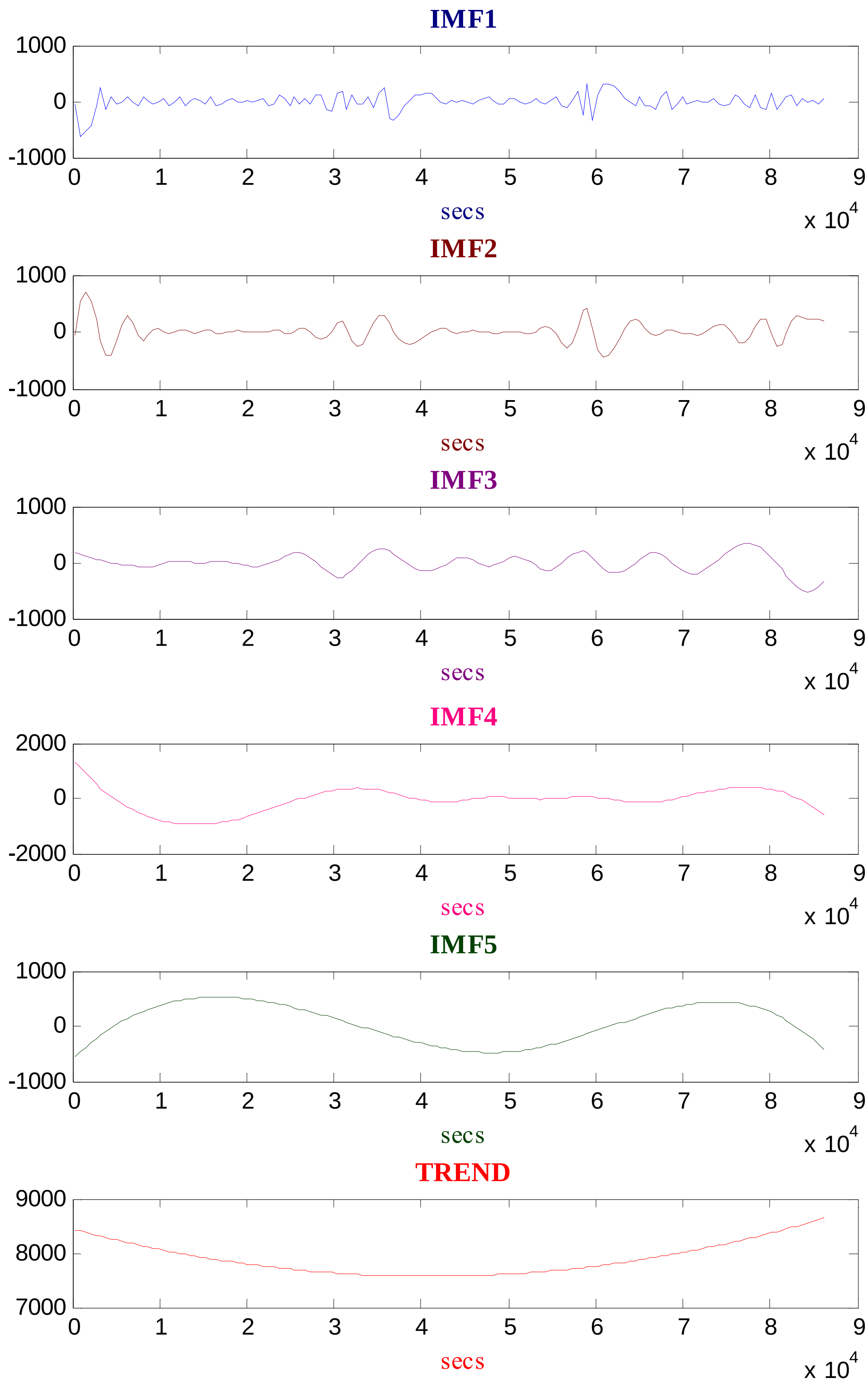

Figure 7: Intrinsic Mode Functions of the EAR data (except the last one which is called trend)

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

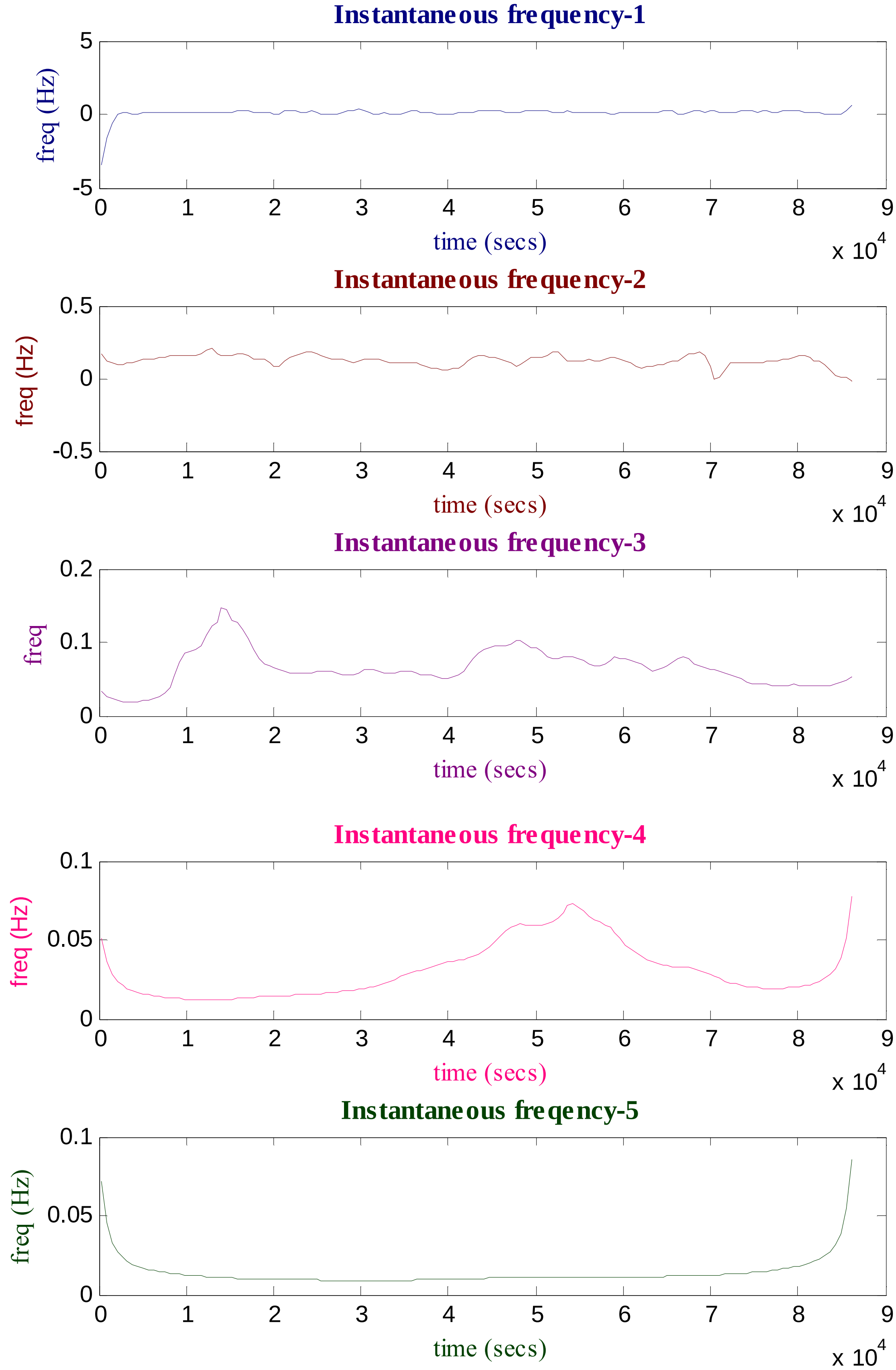

Figure 8: The corresponding instantaneous frequencies of the IMFs