The empirical mode decomposition (EMD) method in MT data processing

J. Chen and M. Jegen-Kulcsar

SFB 574 IFM-GEOMAR Wischhofstr.1-3, 24148, Kiel, Germany
jchen@ifm-geomar.de, mjegen@ifm-geomar.de

Abstract

The natural magnetotelluric (MT) geophysical prospecting method utilizes the spectra of associated time-varying horizontal electric and magnetic fields at the Earth's surface to determine a frequency-dependent impedance tensor. Most current methods of analysis determine the spectra based on Fourier transform and therefore must assume either that the signals under analysis are stationary over the record length or that any distortion in the spectral estimations due to non-stationarity will occur in an equivalent manner in the spectra of both the electric and magnetic fields and thus have no effect on the impedance estimates. A new method for dealing with non-stationarity of the MT time series based upon empirical mode decomposition (EMD) method and Hilbert spectrum is proposed in this paper. In this paper, we use the EMD method, Hilbert transform and Marginal Hilbert spectrum to determine the impedance tensor and compare the results with the traditional data processing method.

1 Introduction

The classic Fourier transform is based on the decomposition of the signal according to fixed basis functions and a fixed frequency set determined by the sampling frequency and the data or window length. It assumes that the signal is periodic or stationary. The Fourier spectrum defines uniform harmonic components globally, therefore, it needs many additional harmonic components to simulate non-stationary data that are non-uniform globally. As a result, it spreads the energy over a wide frequency range. Constrained by the energy conservation, these spurious harmonics and the wide frequency spectrum cannot faithfully represent the true energy density in the frequency space. The geomagnetic time series, however, are characterized by a change of frequency content with time and non-stationarity. This change of frequency content may not imaged by the Fourier analysis as the Fourier spectra averages over time. A way out is Hilbert spectrum, which does not assume a set of fixed frequencies and allows the imaging of frequency content as a function of time. However calculation of Hilbert spectra is unstable applied on geomagnetic time series directly. But with the development of the EMD method, a new way of decomposing data exists, which lead to stable calculation of the Hilbert spectra.

Hilbert-Huang transform (HHT), introduced by Huang on the basis of the classic Hilbert transform (Huang et al., 1998) [1] [2] [3], is a new non-stationary signal processing technique. The key part of the method is the *Empirical Mode Decomposition* (EMD), with which any complicated data set can be decomposed into a finite and often small number of *intrinsic mode functions* (IMFs) that admit a well-behaved *Hilbert Transform* to obtain the physical meaningful *instantaneous frequencies* (IF). The final presentation of the results is an energy-frequency-time distribution, designated as the *Hilbert spectrum*. The EMD is an adaptive decomposition of the data based on local characteristic time scales of a signal, it is applicable to non-stationary processes, and therefore, it is highly efficient. With the Hilbert transform, the frequency content in each IMF is not fixed but determined by the signal itself, and may change with time. Furthermore, since the method eliminates the need for spurious harmonics to represent non-stationary signals, the corresponding Hilbert spectrum will not lead to energy diffusion and leakage.

In this paper, the Empirical Mode Decomposition (EMD) method will be introduced in section 2. Based on a simple example, we show how to apply the EMD to a time series to obtain the Intrinsic Mode Function (IMF) and we illustrate the concepts of the Hilbert transform, instantaneous amplitude, instantaneous frequency, Hilbert spectrum and marginal spectrum. In section 3, we apply the EMD method to the real

MT data set and calculate the impedance tensor based on the Hilbert spectra. The results are compared to the impedance calculated by the classical Fourier approach. Conclusion and outlook is given in section 4.

2 EMD method

The empirical mode decomposition (EMD) is based on the direct extraction of the energy associated with various intrinsic time scales to generate a collection of intrinsic mode functions (IMF). Each IMF allows a well-behaved Hilbert transform, from which the instantaneous frequency can be calculated. Thus, we can localize any event on the time as well as the frequency axis. The decomposition can be viewed as an expansion of the data in terms of the IMFs. Each IMF can be, just like the underlying time series, non-stationary. Most important of all, the IMFs are adaptive. The requirements for locality and adaptivity are very crucial for non-stationary data, since we need the instantaneous frequency and energy rather than the global frequency and energy defined by the Fourier spectral analysis.

In order to define a meaningful instantaneous frequency, the intrinsic mode functions (IMF) have to satisfy two conditions [1] [2] [3]:

(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; and
(2) at any point, the mean value of the envelope defined by the local maxima and the envelope by the local minima is zero.

These requirements to intrinsic mode function are adopted because they represent the oscillation mode imbedded in the data. Each IMF is capable of containing a modulated frequency and amplitude and therefore might be of non-stationary character.

An IMF represents a simple oscillatory mode as opposed to a simple harmonic function. Based on the above definitions, any signal $x(t)$ can be decomposed as follows [1] [2] [3]:

(1) Identify all the local extrema, and then connect all the local maxima by a cubic spline line as the upper envelope.
(2) Repeat the procedure for the local minima to produce the lower envelope. The upper and lower envelopes should cover all the data between them.
(3) The mean of upper and low envelope value is designated as m_1; and the difference between the signal $x(t)$ and m_1 is the first component, h_1; i.e.

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

Ideally, if h_1 is an IMF, then h_1 is the first component of $x(t)$.
(4) If h_1 is not an IMF, h_1 is treated as the original signal and repeat (1), (2), (3); then

$$h_1 - m_{11} = h_{11}$$

After repeated sifting, i.e. up to k times, h_{1k} becomes an IMF, that is

$$h_{1(k-1)} - m_{1k} = h_{1k}$$

Then it is designated as

$$c_1 = h_{1k}$$

The first IMF component is obtained from the original data. c_1 should contain the finest scale or the shortest period component of the signal.
(5) Separate c_1 from $x(t)$ by

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

where r_1 is treated as the original data and repeat the above processes until the second IMF component c_2 of $x(t)$ has been derived. The above process is repeated n times until n-IMFs of the signal $x(t)$ have been determined.

$$r_1 - c - 2 = r_2$$
$$\vdots$$
$$r_{n-1} - c_n = r_n$$

The decomposition process can be stopped when r_n becomes a monotonic function from which no more IMFs can be extracted. We finally obtain

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

Thus, one can achieve a decomposition of the signal into n-empirical modes and a residue r_n, which is the mean trend of $x(t)$. The IMFs $c_1, c_2, \cdots, c_n$ include different frequency bands, where highest frequencies are usually found in the first IMF and lower frequencies in subsequent IMFs. The frequency content in each IMF changes with time and the frequency band found in one IMF might overlap with the frequency band in another IMF. However, at each point in time, no two IMFs contain the same frequency.

For example, figure 1 shows a time series that is the sum of two sine waves with an modulated amplitude wave given by:

$$x(t) = 2sin(2\pi \times 15t) + sin(2\pi \times 5t)sin(2\pi \times 0.1t) + 4sin(2\pi \times t)$$

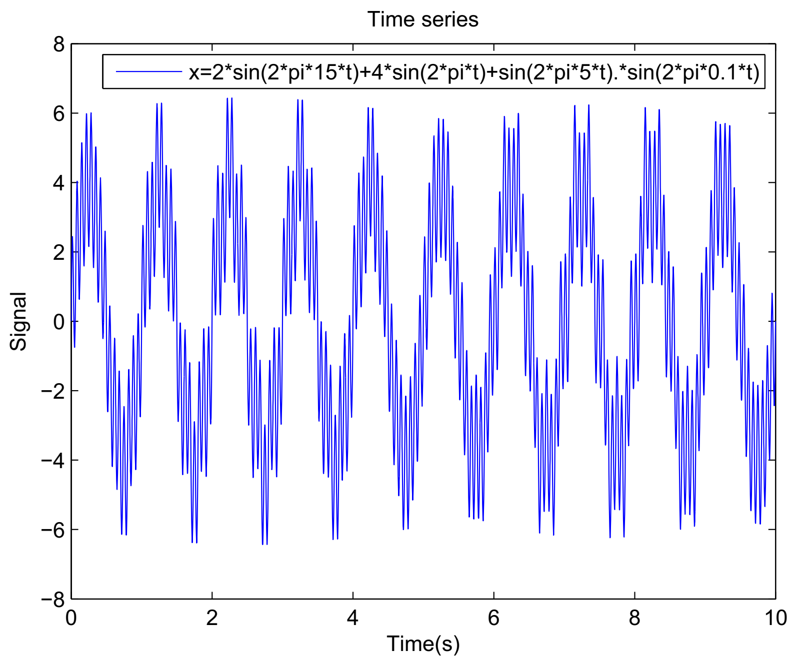

Figure 1: The time series example.

Figure 2 shows the IMFs of the time series. It is clear that the first IMF shows the 15Hz component of the time series, the second IMF the 5Hz modulated signal and the third IMF the 1Hz signal, respectively. The last IMF shows the trend of the signal.

In a second step, the IMFs are submitted to the Hilbert transformation process, which is defined as:

$$Y(t) = \frac{1}{\pi} P \int_{\infty}^{\infty} \frac{X(t')}{t - t'} dt'$$

where P indicates the Cauchy principle value. With this definition, $X(t)$ and $Y(t)$ form the complex conjugate pair, which can be composed to an analytic signal $Z(t)$, as

$$Z(t) = X(t) + iY(t) = A(t)e^{i\theta(t)}$$

where

$$A(t) = \sqrt{X^2(t) + Y^2(t)}$$

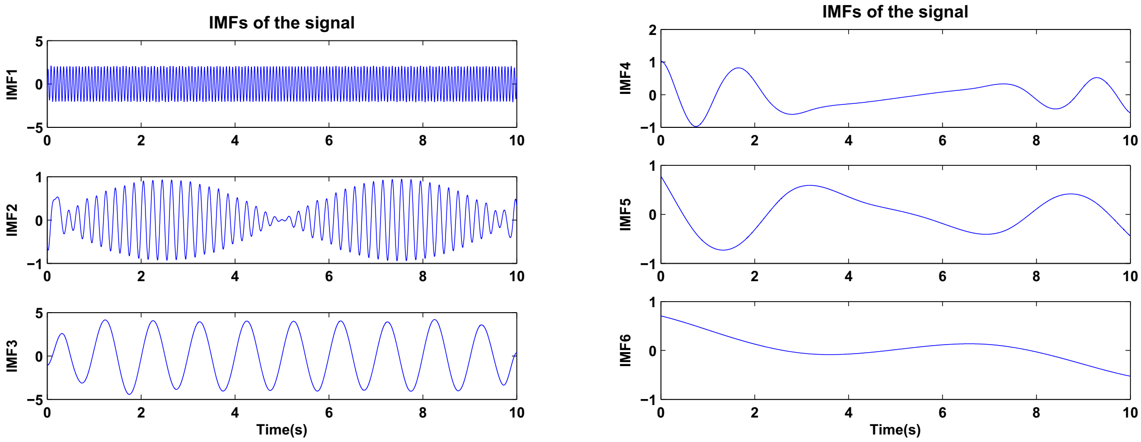

Figure 2: The 6 IMF components of time series.

and

$$\theta(t) = \text{atan}(\frac{Y(t)}{X(t)})$$

We call $A(t)$ *instantaneous amplitude* and $\theta(t)$ *instantaneous phase*.

If we express the phase with a Taylor series then

$$\theta(t) = \theta(t_0) + (t - t_0)\theta'(t_0) + R$$

where R is small when t is close to t_0. The analytic signal becomes

$$Z(t) = X(t) + iY(t) = A(t)e^{i\theta(t)} = A(t)e^{i(\theta(t_0) - t_0\theta'(t_0))}e^{it\theta'(t_0)}e^{iR}$$

and we see that $\theta'(t_0)$ has the role of frequency if R is neglected. This makes it natural to introduce the notion of *instantaneous (angular) frequency*, that is

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

After performing the Hilbert transform to each IMF component, the original signal can be expressed as the real part (RP) of the analytic signal in the following form:

$$x(t) = \text{RP}\sum_{j=1}^{n} A_j(t)e^{i\theta_j(t)} = \text{RP}\sum_{j=1}^{n} A_j(t)e^{i\int \omega_j(t)dt}$$

Here we left out the residue r_n on purpose, for it is either a monotonic function or a constant.

The above equation enables us to represent the amplitude and the instantaneous frequency as functions of time in a three-dimensional plot, in which the amplitude can be contoured on the frequency-time plane. This frequency-time distribution of the amplitude is designated as the *Hilbert spectrum* $H(\omega, t)$.

With the Hilbert spectrum defined, we can also define the marginal spectrum, $h(\omega)$ as

$$h(\omega) = \int_0^T H(\omega, t)dt$$

where T is the total data length. The Hilbert spectrum offers a measure of amplitude contribution from each frequency and time, while the marginal spectrum offers a measure of the total amplitude contribution from each frequency.

Instantaneous Frequency An ambiguity inherent in the instantaneous phase renders equation above impractical for calculating the instantaneous frequency: only the principal values of the phase are computed, which causes 2π phase discontinuities. Instantaneous frequency is instead calculated by another equation, directly derived from equation above:

$$\omega(t) = \frac{X(t)Y'(t) - X'(t)Y(t)}{X^2(t) + Y^2(t)}$$

where the primes denote differentiation with respect to time [4].

Above equation requires two differentiations to calculate the instantaneous frequency. To avoid these differentiations, three formulas that approximate instantaneous frequency and are faster to compute are used[4].

$$\omega_a(t) = \frac{1}{T}\text{atan}\frac{X(t)Y(t+T) - X(t+T)Y(t)}{X(t)X(t+T) + Y(t)Y(t+T)}$$

$$\omega_b(t) = \frac{1}{2T}\text{atan}\frac{X(t-T)Y(t+T) - X(t+T)Y(t-T)}{X(t-T)X(t+T) + Y(t-T)Y(t+T)}$$

$$\omega_c(t) = \frac{4}{T}\text{atan}\frac{X(t)Y(t+T) - X(t+T)Y(t)}{(X(t) + X(t+T))^2 + (Y(t) + Y(t+T))^2}$$

where T is sample period.

Now we can calculate the instantaneous frequencies and instantaneous amplitudes of each IMFs of the above example. Figure 3 shows the instantaneous frequencies (left) and amplitude (right). Frequencies components can be clearly seen in the left figure. It is obvious that the time series just contains discrete frequencies as opposed to the continues frequency band in Fourier analysis. The small oscillations in the frequency band and the inaccurate frequencies at the boundary are due to the numerical calculation.

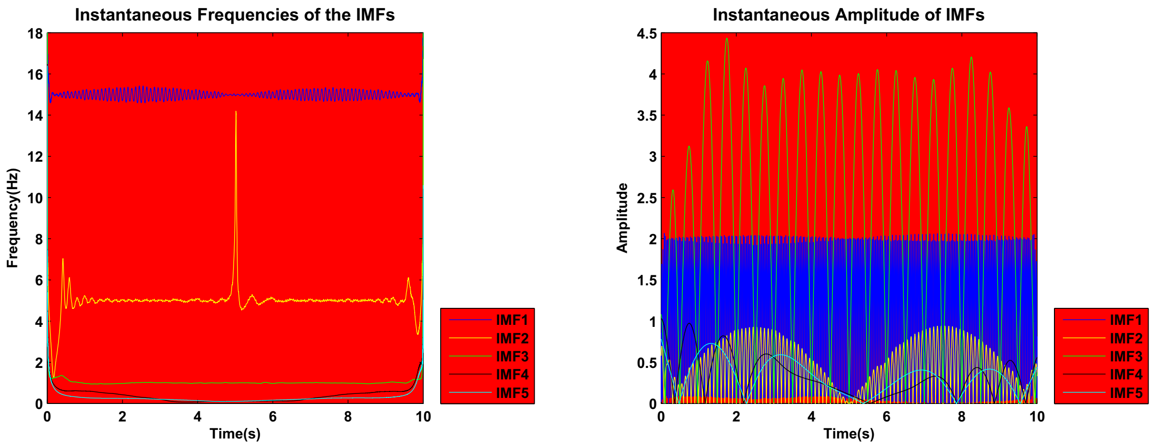

Figure 3: The instantaneous frequency and amplitude of IMF components of the time series.

We can calculate the Hilbert marginal spectrum of the time series. The comparison with the Fourier spectrum is shown in Figure 4. One can see that the peaks of the three components are clearly separated and the resolution ratio is higher in the Hilbert marginal spectrum since there is severe energy leakage in Fourier spectrum. This gives us an idea to apply this efficient method to estimate the apparent resistivity from measured electro- and magneto- fields time series.

3 EMD and HT to the MT data processing

Now, we apply the EMD and HT to the magnetotelluric "FLARE10" raw data measured near Faroe island. The MT stations and the E- and B-field time series of MT station 11 are shown in Figure 5.

We applied EMD method to the E- and B-field time series to obtain the IMFs of both time series, which for the E-field is shown in Figure 6. Through a Hilbert transform of each IMF, one can obtain the

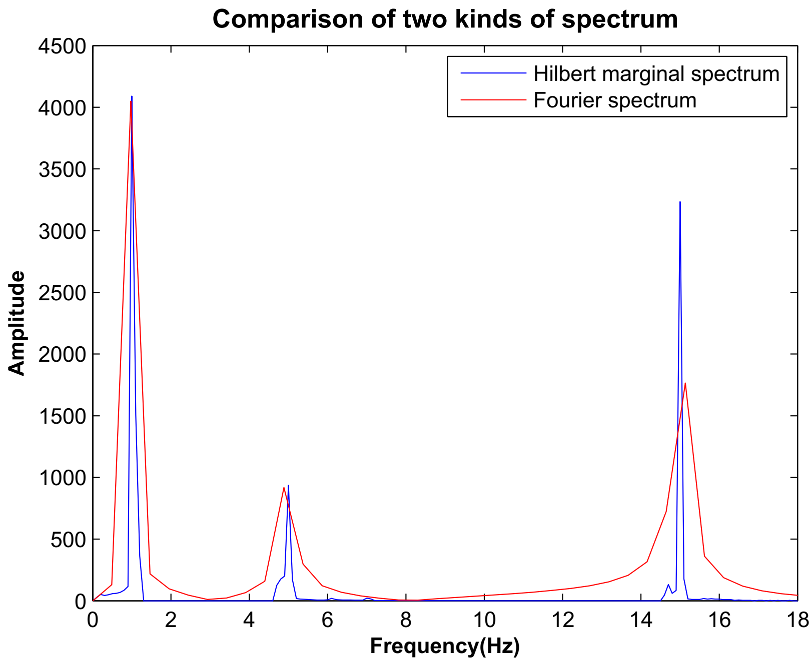

Figure 4: The comparison of Hilbert marginal spectrum and Fourier spectrum of the time series.

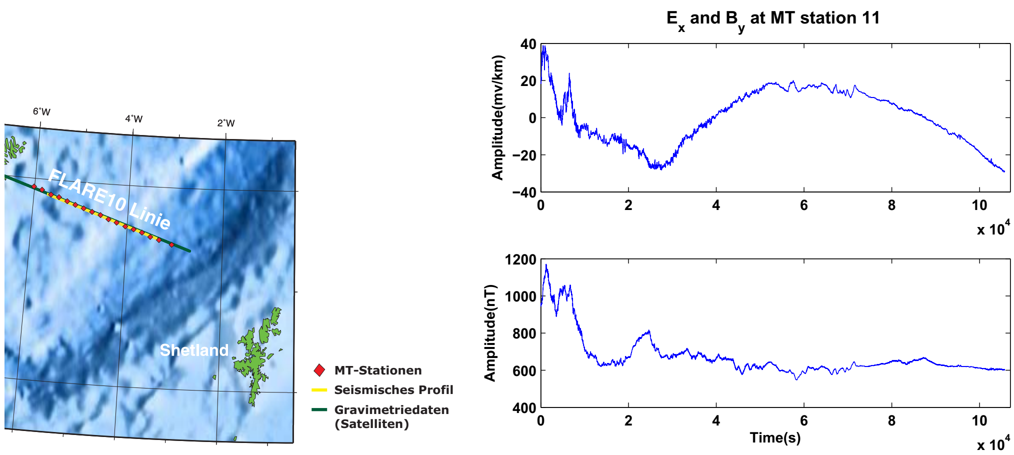

Figure 5: MT stations near Faroe island. The time series of E- and B-field at MT station 11.

instantaneous frequencies and amplitudes of the signal. The instantaneous frequencies for all IMFs are are shown in Figure 7. Through the Hilbert spectra it becomes obvious that the frequencies contained in the signal are discrete a fact that is not visible in the Fourier analysis.

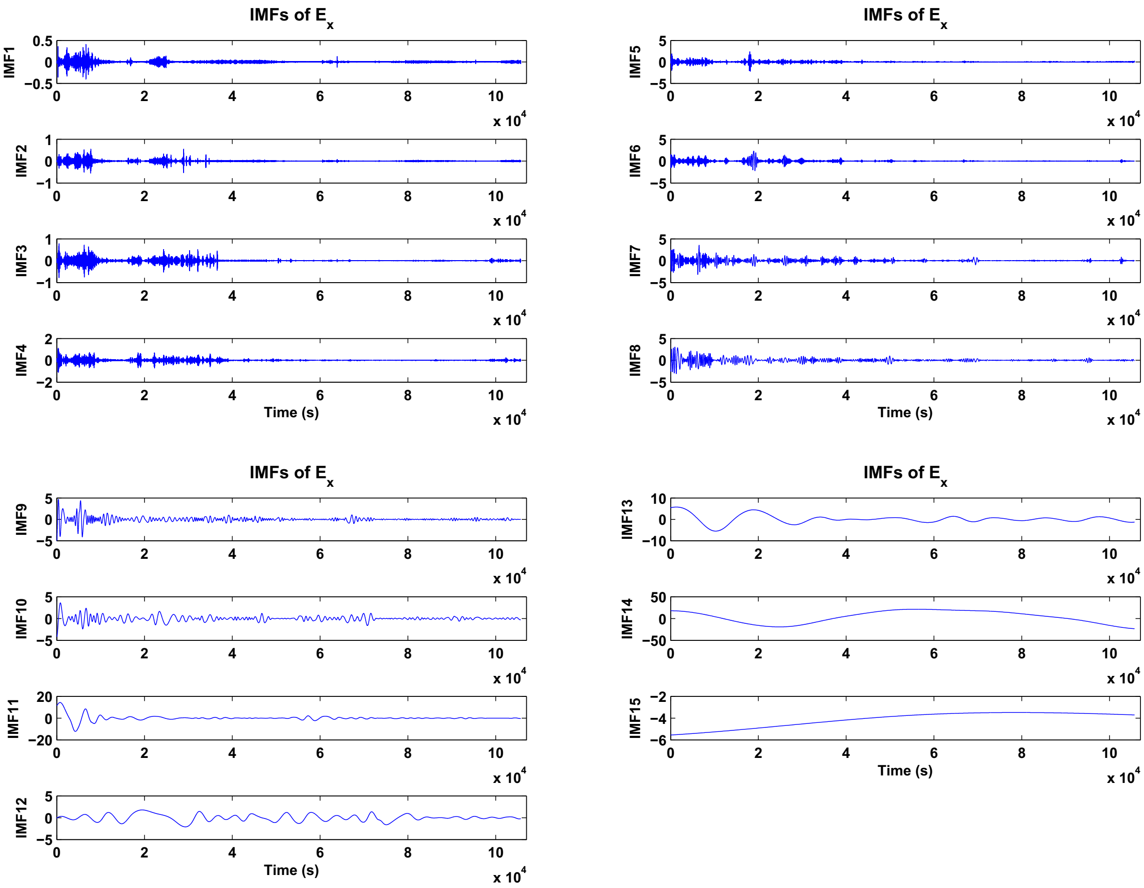

Figure 6: The 15 IMF components of $e(t)$.

Now, we can choose some frequency bands to calculate the Hilbert marginal spectrum of each frequency band to calculate the impedance given by[5] [8] [9] [7] [10] [6]

$$\begin{pmatrix} E_x(\omega) \\ E_y(\omega) \end{pmatrix} = \frac{1}{\mu_0} \begin{pmatrix} Z_{xx}(\omega) & Z_{xy}(\omega) \\ Z_{yx}(\omega) & Z_{yy}(\omega) \end{pmatrix} \begin{pmatrix} B_x(\omega) \\ B_y(\omega) \end{pmatrix}$$

to calculate the transfer function.

For simplification, we just consider a 2D case in which case the above formula reduces to:

$$\begin{cases} Z_{xy} = \frac{\mu_0 E_x(\omega)}{B_y(\omega)} \\ Z_{yx} = \frac{\mu_0 E_y(\omega)}{B_x(\omega)} \end{cases}$$

In order to obtain the impedances, we introduce three strategies:

a). Chose a time window, compute the Hilbert marginal spectrum pair of E- and B-field w.r.t. certain frequency band.

b). Shift the window with some percentage overlapping and compute another spectrum pair, and so on.

c). For all Hilbert marginal spectrum pairs, make a robust least square fit to obtain the ratio $E(\omega)/B(\omega)$ and the error.

One of the robust fit for frequency 0.0045Hz is shown in Figure 8.

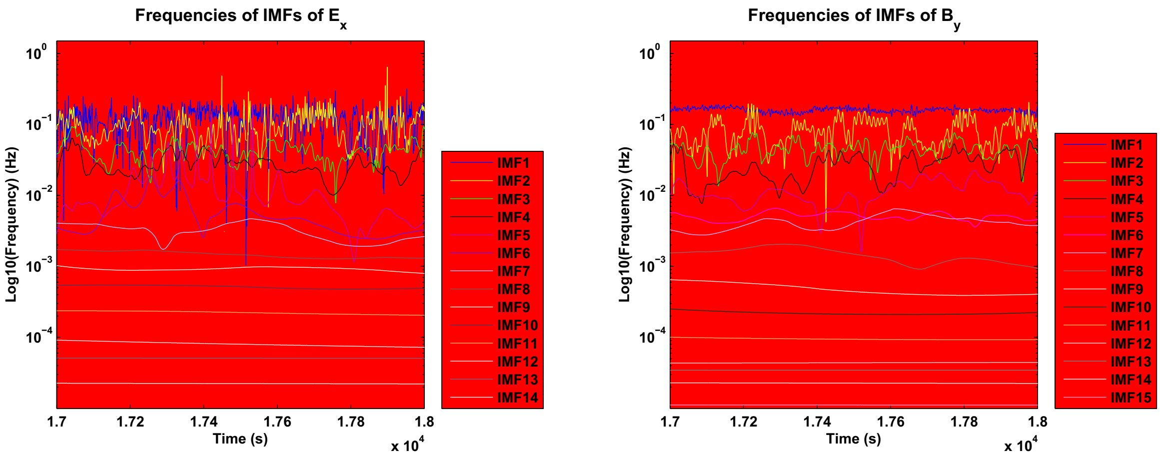

Figure 7: The instantaneous frequencies of 15 IMF components of $e(t)$ and $b(t)$. section: 1000s

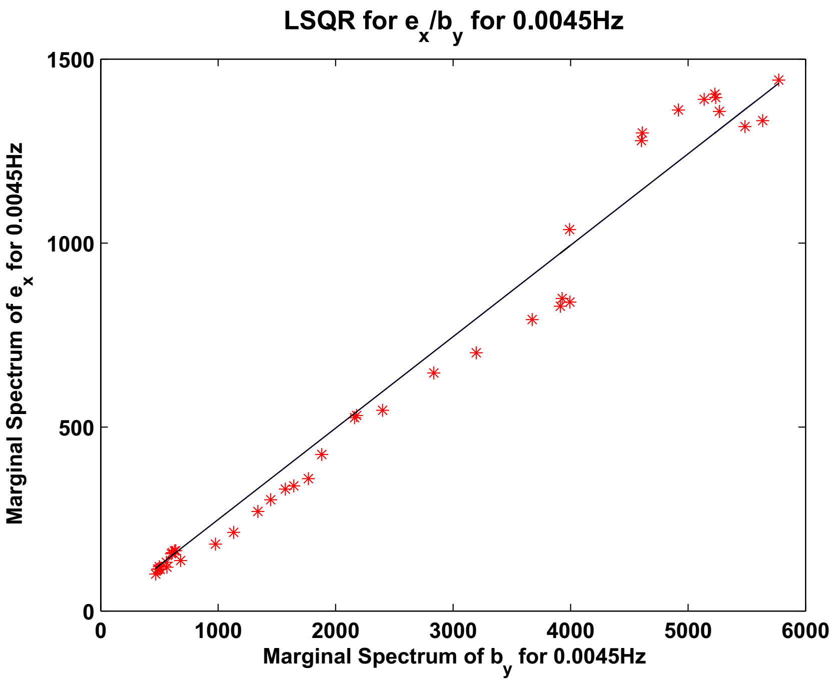

Figure 8: The robust least square fit of the Hilbert marginal spectrum pairs.

After estimation of the ratios $E(\omega)/B(\omega)$ for each chosen frequency band, we can calculate the impedance ρ_{xy} and ρ_{yx} and obtain the apparent resistivities. One estimation of the apparent resistivity of TM mode is shown in Figure 9.

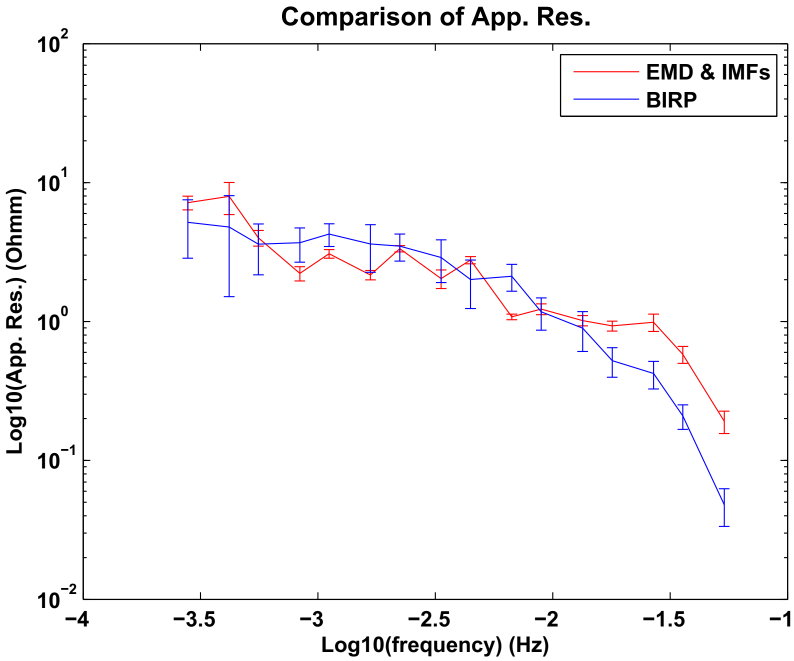

Figure 9: The robust least square fit of the Hilbert marginal spectrum pairs.

4 Conclusion and outlook

In this paper, a new method to deal with the non-stationary MT time series is introduced. The method is easily handled, delivers satisfactory results and may be applied to raw data. Since the EMD method is not tied to specific basis functions but on the data itself, it is adaptive and highly efficient. EMD and HT provide means to analysis the change in frequency content of geomagnetic time series at a high resolution. We are investigating ways of using EMD and HT for transfer function calculations. We believe that it is an interesting new approach to MT data processing which is worth developing forward.

References

[1] Norden E. Huang et al., "*The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis.*", Proc. R. Soc. Lond. A(1998), 454, 903-995.

[2] Gabriel Rilling, Patrick Flandrin and Paulo Gonçalvès, "*On empirical mode decomposition and its algorithms.*".

[3] Patrick Flandrin, "*Empirical mode decompositions as data-driven wavelet-like expansions.*", International Journal of Wavelets, Multiresolution and Information Processing, Vol.2, No.4 (2004) 1-20.

[4] Arture E. Barnes, "*Short note: The calculation of instantaneous frequency and instantaneous bandwidth .*", Geophysics, Vol.57, No.12 (Nov.1992) 1520-1524.

[5] Gary D. Egbert, "*Robust multiple-station magnetotelluric data processing.*", Geophys. J. Int.,(1997) 130 475-496.

[6] D. Sutarno, K. Vozoff, "*Phase-smoothed robust M-estimation of magnetotelluric impedance functions.*", Geophysics, Vol.56, No.12 (1991) 1999-2007.

[7] Philip E. Wannamaker, John R. Booker, Alan G. Jones, et al. "*Resistivity Cross Section Through the Juan de Fuca Subduction System and Its Tectonic Implications.*", Journal of Geophysical Research, Vol.94, No.B10, 14127-14144, Oct. 10, (1989).

[8] I.J. Chant, L.M. Hastie, "*Time-frequency analysis of magnetotelluric data.*", Geophys. J. Int., (1992)111, 399-413.

[9] G. Lamarque, "*Improvement of MT data processing using stationary and coherence tests.*", Geophysical Prospecting, (1999), 47,819-840.

[10] Umberto Spagnolini, "*Time-domain estimation of MT impedence tensor.*", Geophysics, Vol.59, No.5 (May 1994) 712-721.