Advanced spectral analysis and cross correlation based on the empirical mode decomposition: application to the environmental time series

Ben Ismail Dhouha Kbaier 1, *, Lazure Pascal 2, Puillat Ingrid 2

1 IFREMER, REM RDT LSCM, F-29280 Plouzane, France.
2 IFREMER, ODE DYNECO PHYSED, F-29280 Plouzane, France.

* Corresponding author: Dhouha Kbaier Ben Ismail, email address: dhouha.Kbaier@ifremer.fr

Abstract:

Abstract—In marine sciences, time series are often nonlinear and nonstationary. Adequate and specific methods are needed to analyze such series. In this paper, an application of the Empirical Mode Decomposition method (EMD) associated to the Hilbert Spectral Analysis (HSA) is presented. Furthermore, EMD based Time Dependent Intrinsic Correlation (TDIC) analysis is applied to consider the correlation between two nonstationary time series. Four temperature time series obtained from automatic measurements in nearshore waters of the Réunion island are considered, recorded every ten minutes from July 2011 to January 2012. The application of the EMD on these series and the estimation of their power spectra using the HSA are illustrated. The authors identify low-frequency tidal waves and display the pattern of correlations at different scales and different locations. By TDIC analysis, it was concluded that the high frequency modes have small correlation, whereas the trends are perfectly correlated.

Keywords: Augmented Dickey-Fuller tests, cross correlation, empirical mode decomposition (EMD), Hilbert spectral analysis (HSA), Hilbert-Huang transform (HHT), stationarity, time-dependent intrinsic correlation (TDIC), time series, wavelets

I. INTRODUCTION

Generally in Earth sciences and especially in the marine environment, the recorded time series are often nonlinear, nonstationary and interact with each other.

Surface tides are the heartbeat of the ocean and are deterministic since they are controlled by the relative movement of earth, moon and sun. In addition, internal tides are created in a stratified ocean by
the interaction of the surface tide currents with the bathymetry. They are ubiquitous in the ocean and can lead to strong vertical oscillations of isotherms. They can propagate over hundreds of km and since their propagation condition is related to the stratification of the ocean, they are distorted by the circulation which modifies the 3D density field. As a result, in the coastal area, they often appear as highly non linear oscillations with wide spectral content that are usually not predictable, even on short time scale [1]. In situ time series of temperature and currents are then necessary to understand their characteristics and the way they propagate and modify their shape to the coast.

The wavelet transform [2] is widely used as a timefrequency analysis technique to deal with nonstationary signals. Unlike Empirical Mode Decomposition (EMD), wavelets based methods are not adapted to be used when the system is nonlinear or when the time series are unevenly spaced. Therefore, an EMD based on the Hilbert-Huang Transform (HHT) [3] has been applied to the data to analyze the variability. The aim is to identify the presence of tidal internal waves in the different records and to study their correlations.

The paper is structured in the following six sections. After the description of the background and objectives (section I), and of the data set (section II), time series are analysed to understand the involved physical processes according to their time scale by applying the Hilbert-Huang transform method, an advanced spectral analysis (section III). To study the cross correlation between the time series, section IV presents the analysis of the recently published Time Dependent Intrinsic Correlation (TDIC) [4]. Comparisons between EMD and wavelets are provided in the discussion section V. Finally, section VI draws some conclusions.

II. PRESENTATION OF THE EXPERIMENTAL DATABASE

Automatic measurements in 40 m depth waters of the North coast of Réunion island (located in the Indian Ocean 700 km east of Madagascar) are considered. Using Acoustic Doppler Current Profilers (ADCP), both bottom temperatures and currents are recorded every ten minutes from 21st July 2011 to 19th January 2012. To date, this experimental database has not been published. In this paper, the authors consider four temperature time series, shown in Fig. 1, measured at four different stations in the island. In the following, the time series are denoted by 1 \( T_{\text{emp}} \), 2 \( T_{\text{emp}} \), 3 \( T_{\text{emp}} \) and 4 \( T_{\text{emp}} \).

These series are nonstationary. Indeed, we have also applied Augmented Dickey-Fuller (ADF) tests [5] for testing of unit root and stationarity. The unit-root tests exclude that the series are pure random walk processes (first model) or random walk processes with a drift (second model). Nevertheless, these series are nonstationary and have a deterministic trend (third model).
Furthermore, the four temperature measures indicate a periodic component associated to the tide together with stochastic fluctuations. Spectral analysis is a widely used technique to describe the cyclic components of the time series. However, since the recorded bottom temperature series are nonstationary, standard Fourier spectral analysis is inappropriate. In the next section, we perform adequate power spectral analysis of these data.

III. ADVANCED SPECTRAL METHODS

The classical spectral estimates perform well when the system is linear and when data are periodic or stationary. In Earth sciences, however, time series are often unevenly spaced and nonstationary. To accommodate the variety of data generated by nonlinear and nonstationary processes, Huang et al. developed a new adaptive time series analysis method designated by the NASA as the Hilbert-Huang Transform (HHT) [3] and introduced hereafter.

A. A brief description of Hilbert-Huang transform

The HHT consists of the combination of the EMD and the Hilbert spectral analysis (HSA). The key part of this approach is that any complicated dataset can be decomposed with the EMD method into a finite and small number of Intrinsic Mode Functions (IMFs), which represent different scales of the original time series and physically meaningful modes. An IMF is defined as a function having the same number of extrema and zero-crossings. It has also symmetric upper and lower envelopes defined by the local maxima and minima respectively. Due to a dyadic filter bank property of the EMD algorithm [6-8], usually in practice, the number of IMFs modes is less than \(\log_2(N)\), where \(N\) is the length of the data set.

EMD and the associated Hilbert spectral analysis have already been applied in marine sciences. For example, Dätig and Schlurmann [9] applied HHT to show excellent correspondence between simulated and recorded nonlinear waves. Schmitt et al. [10] applied the HHT method to characterize the scale invariance of velocity fluctuations in the surf zone. Ying et al. [11] applied the method and identified three kinds of low-frequency waves using some observations in the coastal water of the East China Sea. The EMD scheme has also been used in studying sea level rise [12].

The EMD algorithm has been applied to the temperature data sets recorded at the same dates. For the four time series, similar IMFs modes have been detected. The analysis for \(\text{Temp}_1\) is presented below.

B. Empirical mode decomposition and Hilbert spectrum results

After EMD decomposition of \(\text{Temp}_1\), 12 IMFs modes are obtained plus the residual, as shown in Fig. 2. The time scale is increasing with the number of the IMF mode, the first IMF thus corresponding to the highest frequency. Several tidal waves are identified through the IMFs: diurnal (IMF_1, 24 hours such as K1), semidiurnal (IMF_2, 12 hours such as M2), third diurnal (IMF_3, 8 hours such as M3), fourth diurnal (IMF_4, 6 hours such as M4), spring and neap tides (IMF_10, 15 days), monthly waves (IMF_11), and seasonal waves (IMF_2).

Fig. 2. \(\text{Temp}_1\), the IMFs modes from EMD and the residual.

To analyze the variability, the HHT has also been applied to the temperature time series observed at the four sites. The time series are divided into a series of modes. Unlike the Fourier transform where each cosine or sine component has a constant frequency, each IMF mode has a time-dependent frequency and amplitude. Reconstructing all the modes together describes the distribution of variability as a function of frequency and time (see Fig. 3).

Fig. 3. Variability of \(\text{Temp}_1\) as a function of time and frequency: Hilbert Spectrum of the EMD. Red/Blue colors indicate high/low variability.

Indeed, the HHT allows frequency-modulation and amplitude-modulation simultaneously. Equation (1) enables us to represent the amplitude and the instantaneous frequency in a three-dimensional plot, in which the amplitude is the height in the time-frequency plane. This time-frequency distribution is designated as the Hilbert-Huang spectrum \(H(w, t)\):

\[
H(w, t) = \operatorname{Re}\left[\sum_{n=1}^{N} a_n(t) \exp\left[j \omega_n(t) t\right]\right]
\]  

Where \(a_n(t)\) is the amplitude, \(\omega_n(t)\) is the instantaneous frequency and \(N\) is the number of modes.

In Fig. 3, a first wintry period is observed in July-October (color blue is dominant) with some peaks of temperature at times; followed by the summer in end November-January (red color corresponds to higher energy in the spectrum). Note that Fig. 3 is consistent with the records of Fig. 1, since the maximum recorded temperatures are obtained in December-January. Indeed, during the summer in the Réunion island, a vertical stratification of the water column combined with horizontal oscillatory currents creates bottom temperature fluctuations, as observed in similar islands of the pacific ocean [13]. Such energetic fluctuations can be seen every few days. However, a long blue line can be distinguished in the higher...
values of energy towards the end of November (from 18th to 21st November) and it corresponds to the linear interpolation of missing data at these dates. The loss of all the high frequency is similarly observed for the missing data from 15th to 19th September, where a stressed dark blue line appears in Fig. 3.

As the signal content is now known in the frequency domain, the focus will shift to the application of an EMD based Time Dependent Intrinsic Correlation (TDIC) of any pair of temperature time series, in order to study the possible links between physical observations.

IV. TIME DEPENDENT INTRINSIC CORRELATION

The classical global expression for the correlation (defined as the covariance of two variables divided by the product of the standard deviation of the two variables) assumes that the variables should be stationary and linear. Applied to nonstationary time series, the cross correlation information may be altered and distorted. The limitations of the correlation coefficient are also obvious: it is unable to provide local temporal information, and it cannot distinguish the main cycles from noise when measuring correlation. Many scientists tried to address the problem of nonsense correlations through different ways. The wavelet transform [2] is widely used as a time-frequency analysis technique to deal with nonstationary signals. The choice of the mother wavelet is usually dependent on the type of data to deal with. HHT on the other hand does not require any convolution of the signal with a predefined basis function or mother wavelet. The process of decomposition is totally data-driven. Comparisons of wavelets with the methods presented in this paper have been investigated in other studies. Some results are presented in section V. An alternative is to estimate the correlation coefficient by means of a time-dependent structure. For example, Papadimitriou et al. [14] applied a sliding window to localize the correlation estimations. Rodo and Rodriguez-Aria [15] developed the scale-dependent correlation technique. Although these methods detected the correlation between two nonstationary signals by computing the correlation coefficient in a local sliding window, the main problem is to determine the size of this window. Recently, Chen et al. [4] introduced an approach based on EMD. They proposed to first decompose the nonlinear and nonstationary data into their IMFs, then use the instantaneous periods of the IMFs to determine an adaptive window and finally compute the time dependent intrinsic correlation coefficients. Huang and Schmitt [16] used TDIC to analyze temperature and dissolved oxygen time series obtained from automatic measurements in a moored buoy station in coastal waters of Boulogne-Sur-Mer (France).

A. TDIC analysis results

The correlation between two data sets is considered here. Suppose the two time series Temp,1(t) and Temp,2(t) can be represented in terms of their IMFs as Temp,1(t) = ∑(j=1)→(n) hj,1(t) + r,1(t) where hj,1(t) is the jth IMF of Temp,1(t) and r,1(t) are the residues. We find the mean period Temp,1(t) of each hj,1(t) either by calculating the local extrema points and zero crossing points, i.e., Temp,1(t) = 4 × data_length × Nbr_max + Nbr_min + Nbr_0 [16] (Nbr = number) or by considering the Fourier energy weighted mean frequency, i.e., Temp,1(t) = 1/[j X’ j(f) df] where X’ j(f) is the Fourier power spectrum of each IMF mode. Then, at time t_max, the sliding window is given by t_max = t_max – a × [max(Tp1(t)Tp2(t))] / 2 – t_min + a × [max(Tp1(t)Tp2(t))] / 2, where a is any positive number. This window is different from classical sliding windows: it is based on the maximum of two instantaneous periods max(Tp1(t),Tp2(t)) and thus it is adaptive.

The focus will shift now to illustrate the cross correlations between Temp,1 and Temp,2 time series as well as Temp,1 and Temp,4. First, the global cross correlation coefficients show a global in-phase relation between the temperature time series. For both case studies, the time series are highly correlated since the maximum correlation coefficient is 0.93 for Temp,1 and Temp,2 with a phase difference of 40 minutes, and 0.86 for Temp,1 and Temp,4, occurring when Temp,4 is shifted forwards for about 3 hours. Note that the time lag between stations S1 and S2 (resp. S3) corresponds to the 1 m s⁻¹ propagation speed, which is coherent with the characteristics of these stations in the Réunion island since S1 (resp. S2) is 2.5 km (resp. 11 km) away from S1 and the phase velocity is 1 m s⁻¹ between them. This result is also observed among the other stations in the island. When considering only the last two months (21st November to 15th January) the correlation coefficients between Temp,1 and Temp,2 or Temp,4 fall down to 0.73 and 0.3 respectively. This highlights the loss of coherency of internal tide over a limited distance of less than 20 km.

The EMD algorithm is first applied to all the data sets for the same time period. There are 12 IMFs modes with one residual, which has been recognized as the trend of the given data [17]. The method allows the data sets to be represented in a multiscale way [18-19]. These are used for multiscale correlation.

Let us consider the IMFs modes with a mean period of 12 hours. The global correlation coefficient is 0.23 for the corresponding IMFs modes of Temp,1 and Temp,2, with a phase difference of 40 minutes; the same phase difference as for the original time series. This result is important since the considered semidiurnal mode is the most energetic one, and it should reflect the global relationship between Temp,1 and Temp,2 time series. A small correlation coefficient of 0.17
(phase difference of 3 hours) is also obtained for the IMFs modes of \( \text{Temp}_1 \) and \( \text{Temp}_2 \), with a mean period of 12 hours.

Fig. 4a displays the measured TDIC and shows rich patterns at small sliding window. We note a decorrelation of the TDIC with the increase of the window size. Although the global cross correlation is small, the TDIC detects several periods of high correlation between the modes. As an example, a strong positive correlation between the IMFs of \( \text{Temp}_1 \) and \( \text{Temp}_2 \) appears nearly by the beginning of November.

The measured TDIC obtained after EMD decomposition of \( \text{Temp}_1 \) and \( \text{Temp}_2 \) a) for the 12-hour and b) for the 15-day mean period.

To focus on this period, Fig. 5 shows a zoom of the IMFs modes with a mean period of 12 hours. These are positively correlated with each other on some portions and negatively correlated on others; showing rich dynamics. The correlation coefficient between these IMFs is also estimated: it is equal to 0.60 for the considered period between 2nd November and 8th November, 2011. Compared to the global correlation coefficient of the IMFs, the value of 0.60 is higher than 0.23 which is coherent with the direct observation of the TDIC in Fig. 4a.

Moreover, the lower the frequency modes are, the higher the correlation is. To show the time evolution and such scale dependence of cross correlation between the time series, we can consider higher periods such as one day, one week, 15 days, etc. Let us consider the IMFs modes of \( \text{Temp}_1 \) and \( \text{Temp}_2 \) with a mean period of 15 days, which corresponds to the spring-neap tidal cycle. The global correlation coefficient for these IMFs is 0.57. Fig. 4b displays the measured TDIC, which confirms the direct analysis of the IMFs modes. Momentary decorrelations can be observed due to the missing data in the middle of September and by the end of November. Note that the holes in Fig. 4b indicate that the TDIC cannot pass the conducted t-test. This means that the independent-samples t-test has failed.

Finally, the residuals from EMD algorithm for \( \text{Temp}_1 \) and \( \text{Temp}_2 \) are plotted in Fig. 6. They show that the trends of the time series are perfectly correlated with a correlation coefficient of 0.99.

V. DISCUSSION

While EMD, Fourier, and wavelets are all used to decompose signals, EMD is fundamentally different from the other two. With both the Fourier and Wavelet transforms, one selects a set of basis signal components and then calculates the parameters for each of these signals. Choosing the wrong basis function can greatly increase the number of terms required to fit the time series. For example, the Fourier transform specifically uses sinusoidal-basis functions (and calculates the amplitude and phase offset for each), resulting in the production of numerous (possibly infinite) harmonics when a non sinusoidal signal is processed; while the wavelet transform uses other more complex and orthogonal wave-forms.

On the other hand, the EMD method does not define a basis a priori and makes no assumptions a priori about the composition of the signal. Rather, each IMF obtained by the sifting process will be a single periodic oscillator, but otherwise cannot be predicted before it is empirically observed from the signal. Furthermore, the number of IMFs cannot be predicted before the decomposition. These two disadvantages can make EMD difficult to work with under certain circumstances. Compared to other spectral analysis methods, the EMD is also computationally expensive, especially when the time series is long and has a large frequency distribution. Nevertheless, since EMD makes no assumptions about signal, the results might be more meaningful. Also, since the IMFs can change over time, EMD makes no assumptions about the stationarity of the signal (or the signal components) and is therefore better suited to nonlinear signals than either Fourier or Wavelets. This makes EMD particularly attractive when analyzing signals from complex systems.

Wavelet powerspectrum versus IMFs for \( \text{Temp}_3 \). The thick black contour designates the 5% significance level against red noise.

In Fig. 7, we have superimposed the mean period of the IMFs on the wavelet spectra of \( \text{Temp}_3 \). A noticeable similarity between the two methods can be observed although a poor low-frequency resolution is discerned for the Morlet wavelet spectra (no thick black contour beyond 1 cycle per day). The disadvantage of the wavelet power spectral analysis,
However, is the requirement of evenly-spaced data. The effects of the linear interpolation can be seen in Fig. 7: Two long blue lines are observed in mid-September and mid-November. They correspond to the loss of the high frequency at these dates.

Table I summarizes some tidal waves obtained after EMD decomposition of the Temp$_{3}$ (similar to the results presented for Temp$_{1}$ in section III. B). Table I also shows the global cross correlation coefficients between the IMFs for Temp$_{1}$ with respect to the IMFs for the three other time series. Note how the correlation increases with the mean period of the IMFs, i.e. for lower frequency modes.

![Table I](https://example.com/table.png)

The contribution of each IMF to the total energy is measured by the variance. For Temp$_{2}$, the biggest contribution comes from the seasonal wave IMF$_{12}$ with over 47% of the total energy. For the high frequency components the semidiurnal wave IMF$_{1}$ and the diurnal wave IMF$_{2}$ account for nearly 9% and 6% of the total energy, respectively. The physical meaning for IMF$_{1}$ and IMF$_{2}$ components are not immediately clear. Fortunately, their contributions are too small: only 0.06% and 1.29% of the total energy respectively.

While the Fourier expansion would require tens of modes to represent the whole data, the EMD method decomposes the time series into only 12 IMFs plus the residual. When all the IMFs are added back successively, we notice that all the energy is recovered, as shown in all the cases in Huang et al. [3].

### VI. Conclusion

Marine environmental time series are typically noisy, complex and strongly nonstationary. Few time-frequency decomposition methods are adapted to analyze such series. In this paper, the authors consider the HHT as an adaptive method to study their multiple scale dynamics. From the temperature observations at four stations, a group of tidal waves are detected using the EMD method. The physical meaning for each IMF to the total energy is measured by the variance. For Temp$_{2}$, the biggest contribution comes from the seasonal wave IMF$_{12}$ with over 47% of the total energy. For the high frequency components the semidiurnal wave IMF$_{1}$ and the diurnal wave IMF$_{2}$ account for nearly 9% and 6% of the total energy, respectively. The physical meaning for IMF$_{1}$ and IMF$_{2}$ components are not immediately clear. Fortunately, their contributions are too small: only 0.06% and 1.29% of the total energy respectively.

The authors would like to thank the Région Bretagne for financial support of the post-doctoral fellowship (SAD MASTOC n°8296). They also thank the Région Réunion for financial support brought to the NortekMed group for the acquisition of the data, provided to IFREMER within the framework of the HydroTeck project.

### Acknowledgments

The authors would like to thank the Région Bretagne for financial support of the post-doctoral fellowship (SAD MASTOC n°8296). They also thank the Région Réunion for financial support brought to the NortekMed group for the acquisition of the data, provided to IFREMER within the framework of the HydroTeck project.

### References


