



Time dependent intrinsic correlation analysis of temperature and dissolved oxygen time series using empirical mode decomposition

Yongxiang Huang^a, François G. Schmitt^{b,*}

^a Shanghai Institute of Applied Mathematics and Mechanics, Shanghai Key Laboratory of Mechanics in Energy Engineering, Shanghai University, Shanghai 200072, People's Republic of China

^b CNRS and University of Lille 1, Laboratory of Oceanology and Geosciences, UMR 8187 LOG, 62930 Wimereux, France

ARTICLE INFO

Article history:

Received 5 December 2012

Received in revised form 22 May 2013

Accepted 17 June 2013

Available online 25 June 2013

Keywords:

Coastal oceanic time series

Oceanic temperature

Oceanic dissolved oxygen

Empirical mode decomposition

Hilbert spectral analysis

Cross correlation

ABSTRACT

In the marine environment, many fields have fluctuations over a large range of different spatial and temporal scales. These quantities can be nonlinear and non-stationary, and often interact with each other. A good method to study the multiple scale dynamics of such time series, and their correlations, is needed. In this paper an application of an empirical mode decomposition based time dependent intrinsic correlation, of two coastal oceanic time series, temperature and dissolved oxygen (saturation percentage) is presented. The two time series are recorded every 20 min for 7 years, from 2004 to 2011. The application of the empirical mode decomposition on such time series is illustrated, and the power spectra of the time series are estimated using the Hilbert transform (Hilbert spectral analysis). Power-law regimes are found with slopes of 1.33 for dissolved oxygen and 1.68 for temperature at high frequencies (between 1.2 and 12 h) with both close to 1.9 for lower frequencies (time scales from 2 to 100 days). Moreover, the time evolution and scale dependence of cross correlations between both series are considered. The trends are perfectly anti-correlated. The modes of mean year 3 and 1 year have also negative correlation, whereas higher frequency modes have a much smaller correlation. The estimation of time-dependent intrinsic correlations helps to show patterns of correlations at different scales, for different modes.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Generally in geosciences, but especially in the marine environment, many fields have fluctuations over a large range of spatial and temporal scales. To study their dynamics and estimate their variations at all scales, high frequency measurements are needed (Chang and Dickey, 2001; Chavez et al., 1997; Dickey, 1991). Here a time series obtained from automatic measurements in a moored buoy station in coastal waters of Boulogne-sur-mer (eastern English Channel, France) is considered, recorded every 20 min from 2004 to 2011 (Zongo and Schmitt, 2011; Zongo et al., 2011). This fixed buoy station can record various biogeochemical parameters simultaneously. Here, mainly to illustrate the application of a new method for multi-scale data analysis there is a focus on two parameters: temperature, due to its obvious importance, influenced by the dynamics, by meteorology, and as a link with ecosystem forcing, and dissolved oxygen time series, due to its important role in biological processes, and also for the probable growing importance of this parameter to assess the quality of coastal waters, in the framework of European directives (Best et al., 2007).

These physical and biogeochemical time series are nonlinear, and non-stationary, and may possess interactions at different scales. In order to consider their multi-scale dynamic properties and explore their correlations at different scales, the empirical mode decomposition (EMD) framework is applied here (Huang et al., 1998).

EMD and the associated Hilbert spectral analysis (resp. Hilbert–Huang Transform, HHT) have already been applied in marine sciences (Dätig and Schlurmann, 2004; Hwang et al., 2003; Schmitt et al., 2009; Veltcheva and Soares, 2004). For example, Hwang et al. (2003) applied the HHT method to ocean wave data. They found that the HHT method detects more energy in lower frequencies, leading to a lower average frequency in HHT spectra than using the Fourier framework. Dätig and Schlurmann (2004) showed that the HHT method can be used to study nonlinear waves using instantaneous frequencies. Schmitt et al. (2009) applied the HHT method to characterize the scale invariance of velocity fluctuations in the surf zone. They observed that the scale invariance holds for almost two decades of time scales.

In the following, the methodology is presented and then its application is illustrated on the chosen time series. Section 2 presents the Hilbert–Huang Transform method and the fairly recent time dependent intrinsic correlation; Section 3 presents the data base; Section 4 presents the analysis of the intrinsic correlation and Section 5 draws the main conclusion of this paper.

* Corresponding author. Tel.: +33 321992935; fax: +33 321992901.
E-mail addresses: yongxianghuang@gmail.com (Y. Huang), francois.schmitt@univ-lille1.fr (F.G. Schmitt).

2. Hilbert–Huang transform and time dependent intrinsic correlation

In this section, the Hilbert–Huang transform and the empirical mode decomposition based time dependent intrinsic correlation are presented. These time series analysis techniques have been applied, since their introduction in 1998 (Huang et al., 1998), in several thousand different studies in natural and applied sciences. Here the main idea is recalled but the method is not presented in too much detail.

The HHT consists of two steps. The first step is the so-called ‘empirical mode decomposition’, which separates a multi-scale time series into a sum of intrinsic mode functions without a priori basis assumption (Flandrin and Gonçalves, 2004; Huang et al., 1998). In the second step, the Hilbert spectral analysis is applied to each mode function to extract the time-frequency information. The so-called Hilbert spectrum and the corresponding Hilbert marginal spectrum are then introduced to characterize the time-frequency distribution of a given time series (Chen et al., 2010; Huang, 2009; Huang et al., 1998, 2008).

2.1. Empirical mode decomposition

Empirical mode decomposition is a fully adaptive technique to study the nonlinear and non-stationary properties of time series (Flandrin and Gonçalves, 2004; Huang et al., 1998, 1999; Huang et al., 2011). The main idea of EMD is to locally separate a given multi-scale signal into a sum of a local trend and a local detail, respectively, for a low frequency part and a high frequency part (Rilling et al., 2003). The latter is called the intrinsic mode function (IMF), and the former is called the residual. The procedure is repeated to the residual, considered as a new times series, extracting a new IMF using a spline function, and obtaining a new residual until no more IMF can be extracted (Flandrin et al., 2004; Huang et al., 1998, 1999; Rilling et al., 2003). The EMD method then expresses a multi-scale time series as the sum of a finite number of IMFs and a final residual (Flandrin et al., 2004; Huang et al., 1998).

To be an IMF, an approximation to the so-called mono-component signal, it must satisfy the following two conditions: (i) the difference between the number of local extrema and the number of zero-crossings must be zero or one; and (ii) the running mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero (Huang et al., 1998, 1999; Rilling et al., 2003). The so-called empirical mode decomposition algorithm is then proposed to decompose a signal into IMFs (Huang et al., 1998, 1999; Rilling et al., 2003):

1. identify the local extrema of the signal $x(t)$;
2. construct upper envelope $e_{\max}(t)$ by using the local maxima through a cubic spline interpolation (other interpolations are also possible). Construct a lower envelope $e_{\min}(t)$ by using the local minima;
3. define the mean value $m_1(t) = (e_{\max}(t) + e_{\min}(t))/2$;
4. remove the mean value from the signal, providing the local detail $h_1(t) = x(t) - m_1(t)$;
5. check if the component $h_1(t)$ satisfies the above conditions to be an IMF. If yes, take it as the first IMF $C_1(t) = h_1(t)$. This IMF mode is then removed from the original signal and the first residual, $r_1(t) = x(t) - C_1(t)$ is taken as the new series in step 1. If $h_1(t)$ is not an IMF, a procedure called the “sifting process” is applied as many times as necessary to obtain an IMF (not detailed here).

By construction, the number of extrema decreases when going from one residual to the next; the above algorithm ends when the residual has only one extrema, or is constant, and in this case no more IMF can be extracted; the complete decomposition is then

achieved in a finite number of steps. The analyzed signal $x(t)$ is finally written as the sum of mode time series $C_i(t)$ and the residual $r_n(t)$:

$$x(t) = \sum_{i=1}^N C_i(t) + r_n(t). \quad (1)$$

Based on a dyadic filter bank property of the EMD algorithm, the number of IMF modes is estimated as

$$N \leq \log_2(L) \quad (2)$$

where L is the length of the data in points (Flandrin et al., 2004; Huang et al., 2008; Wu and Huang, 2004). Unlike Fourier based methodologies, e.g., Fourier analysis and wavelet transform, this method does not define the basis a priori (Flandrin and Gonçalves, 2004; Huang et al., 1998, 1999). It thus possesses full adaptability and is very suitable for non-stationary and nonlinear time series analysis (Huang et al., 1998, 1999).

2.2. Hilbert spectral analysis

To characterize the time-frequency distribution of the IMF mode, a complementary analysis technique namely Hilbert spectral analysis (HSA) is then applied to each IMF mode to extract the local frequency information (Huang, 2009; Huang et al., 1998, 1999, 2011; Long et al., 1995). In this complementary step, the Hilbert transform is used to construct the analytical signal, i.e.,

$$\tilde{C}(t) = C(t) + j \frac{1}{\pi} P \int_{-\infty}^{+\infty} \frac{C(t')}{t-t'} dt' \quad (3)$$

in which P is the Cauchy principle value (Cohen, 1995; Flandrin, 1998; Huang et al., 1998; Long et al., 1995). The above equation can be rewritten as

$$\tilde{C}(t) = \mathcal{A}(t) \exp(j\theta(t)) \quad (4)$$

in which $\mathcal{A}(t) = |\tilde{C}(t)|$ is the modulus and $\theta(t) = \arctan(\text{IM}(\tilde{C}(t))/\text{RE}(\tilde{C}(t)))$ is the instantaneous phase function (Cohen, 1995; Flandrin, 1998; Huang, 2009; Huang et al., 1998, 2011; Long et al., 1995). The instantaneous frequency is then defined as

$$\omega(t) = \frac{1}{2\pi} \frac{d\theta(t)}{dt}. \quad (5)$$

Note that the instantaneous frequency ω is very local since the Hilbert transform is a singularity transform and the differential operator is used to define the frequency ω (Huang et al., 1998, 1999; Huang, 2009). It was found experimentally that the Hilbert-based methodology is free with the Heisenberg–Gabor uncertainty and can be used to describe nonlinear distortions by using an intrawave-frequency-modulation mechanism, in which the frequency can be varied with time in one period (Huang, 2005; Huang et al., 1998, 2011). Therefore, the method is free with high-order harmonic components, which are required in Fourier-based methods to capture the non-stationary and nonlinear characteristics of the data (Huang et al., 1998, 1999; Huang, 2005; Huang et al., 2011).

Note that several methods exist, that could be applied to estimate the instantaneous frequency, e.g., direct quadrature, and teager energy operator; see more detail in (Huang et al., 2009a). In practice, the Hilbert method already provides a good estimation of ω in a statistical sense (Huang et al., 2010, 2011).

A Hilbert spectrum $H(\omega, t) = \mathcal{A}^2(\omega, t)$ can be designed to represent the energy of the original signal as a function of frequency ω and time. It can be taken as the best local fit to $x(t)$ using an amplitude and phase varying trigonometric function (Huang, 2005). This corresponds to a

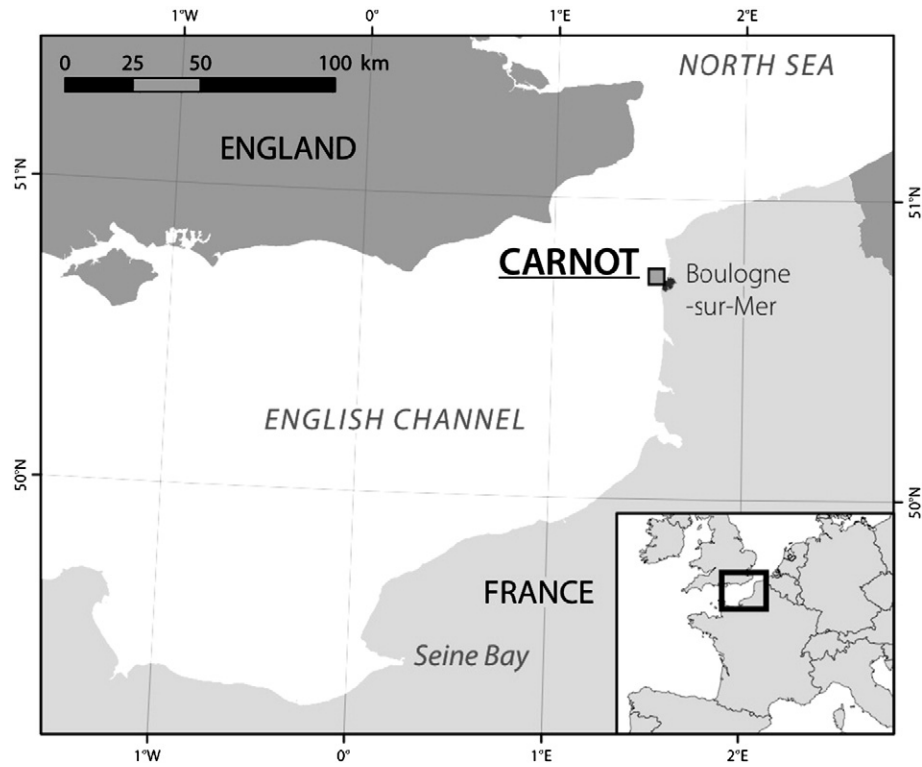


Fig. 1. A map showing the measurement location in Boulogne-sur-mer's coastal waters (France) at position 50.7404 N, 1.5676 W, in the eastern English Channel.

high resolution fit both in physical space and frequency space (Huang et al., 1998; Long et al., 1995). Then, the Hilbert marginal $h(\omega)$ can be defined as

$$h(\omega) = \frac{1}{T} \int_T H(\omega, t) dt \quad (6)$$

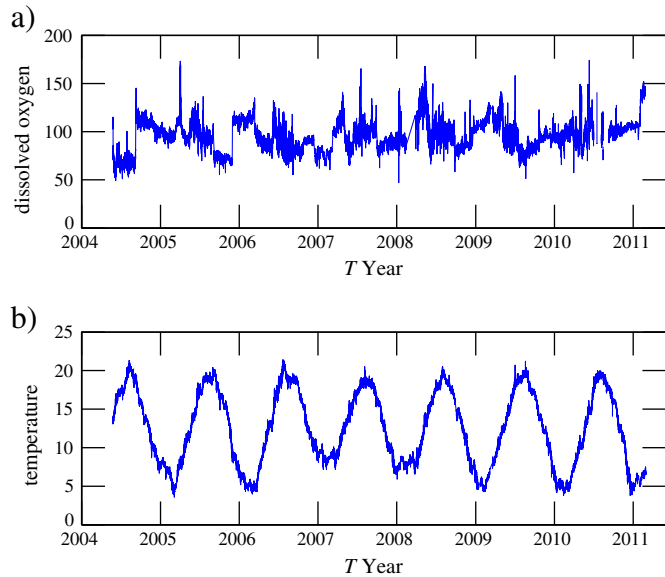


Fig. 2. The observed dissolved oxygen a) and the temperature b) on the time period 25th May 2004 to 1st March 2011. Note that the sampling interval is not always 20 min, due to missing data. The mean period is 24 min for both data sets. A moderate warming winter is observed in 2007 and 2008. The observed temperature during the winter is 2 °C higher than usual. Note that a strong riding-wave phenomenon is observed from January to April 2008. The observed riding-wave will cause the mode mixing problem in the EMD decomposition.

in which T is the time period to calculate the spectrum. This is comparable to the power spectrum in Fourier analysis, but it has a quite different definition and the resulting curve would be expected to be different from the Fourier $E(f)$ curve if the times series is nonlinear and non-stationary (Huang et al., 1998). In fact, here the definition of instantaneous frequency is different from the one in the Fourier frame and the interpretation and detailed physical meaning of $h(\omega)$ is still to be fully characterized.

Another way to estimate $h(\omega)$ is based on the joint probability density function (pdf) $p(\mathcal{A}, \omega)$ of amplitude \mathcal{A} and the instantaneous

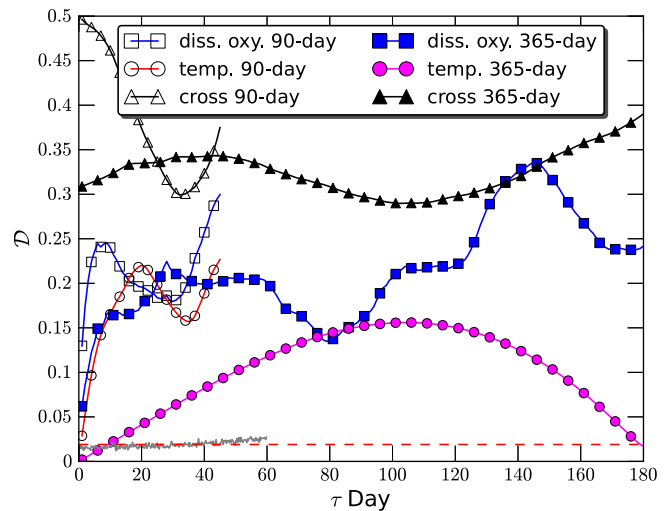


Fig. 3. Statistical test of the degree of stationarity \mathcal{D} for the autocorrelation and cross-correlation of two data sets with a sliding window $t_w = 90$ days and $t_w = 365$ days. The test for white noise is illustrated as a horizontal dashed line, showing the validation of the definition of the degree of stationary. Note that the measured \mathcal{D} is strongly dependent on the windows size t_w . The curves indicate that none of the series are stationary but that the temperature data set is more stationary than the dissolved oxygen one.

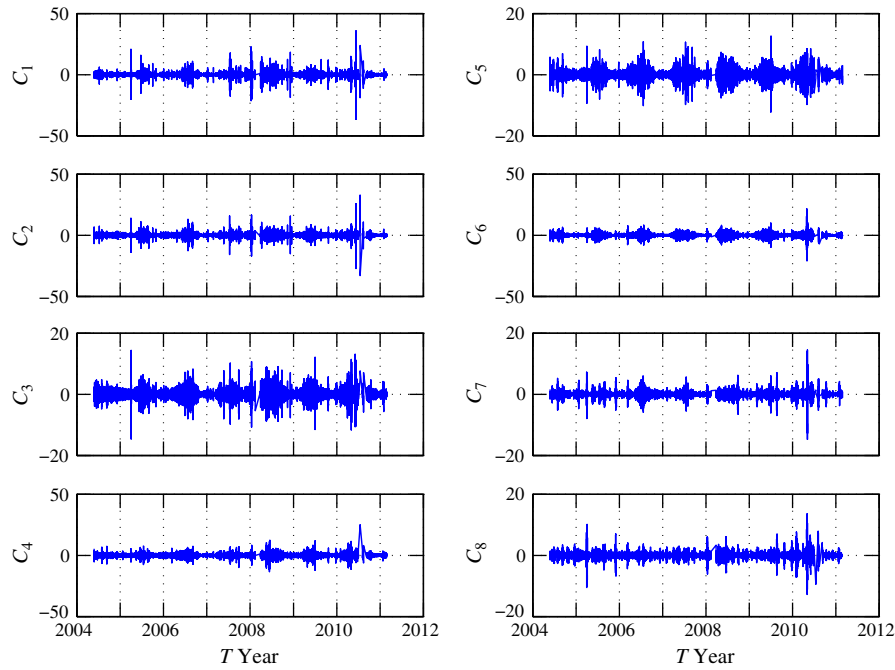


Fig. 4. The first-eight IMF modes for dissolved oxygen. The time scale is increasing with the mode index n .

frequency ω from all IMF modes (Huang et al., 2008; Huang, 2009; Huang et al., 2010, 2011). It is written as

$$h(\omega) = \int_{-\infty}^{+\infty} p(\mathcal{A}, \omega) \mathcal{A}^2 d\mathcal{A}. \quad (7)$$

The combination of the EMD and HSA is then called Hilbert–Huang Transform (Huang, 2005). It has been applied successfully in various research fields to characterize the energy–time–frequency in different topics (Chen et al., 2010; Coughlin and Tung, 2004; Echeverria et al., 2001; Huang et al., 2008, 2009b; János and Müller, 2005; Loutridis, 2004; Molla et al., 2006a,b; Schmitt et al., 2009; Zhu et al., 1997), to cite a few.

2.3. Time dependent intrinsic correlation

The classical global expression for the correlation, applied to non-stationary time series, may distort the true cross correlation information between time series and provide unphysical interpretation (Chen et al., 2010; Hoover, 2003; Rodo and Rodriguez-Arias, 2006). An alternative way, consistent with the possible non-stationarity of the time series, is to estimate the cross correlation coefficient by using a sliding window or a scale dependent correlation technique (Papadimitriou et al., 2006; Rodo and Rodriguez-Arias, 2006). But the main problem of these techniques is to determine how large the window should be (Chen et al., 2010). Time series from natural sciences possess fluctuations in a whole range of scales, and the data characterization must take into account this scale question (Cohen,

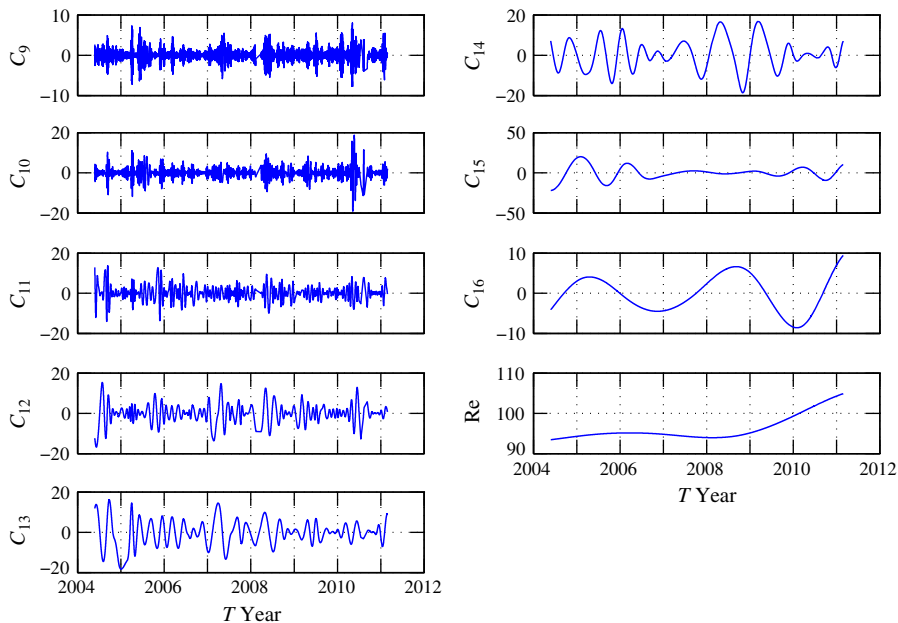


Fig. 5. The last-eight IMF modes for dissolved oxygen. The time scale is increasing with the mode index n .

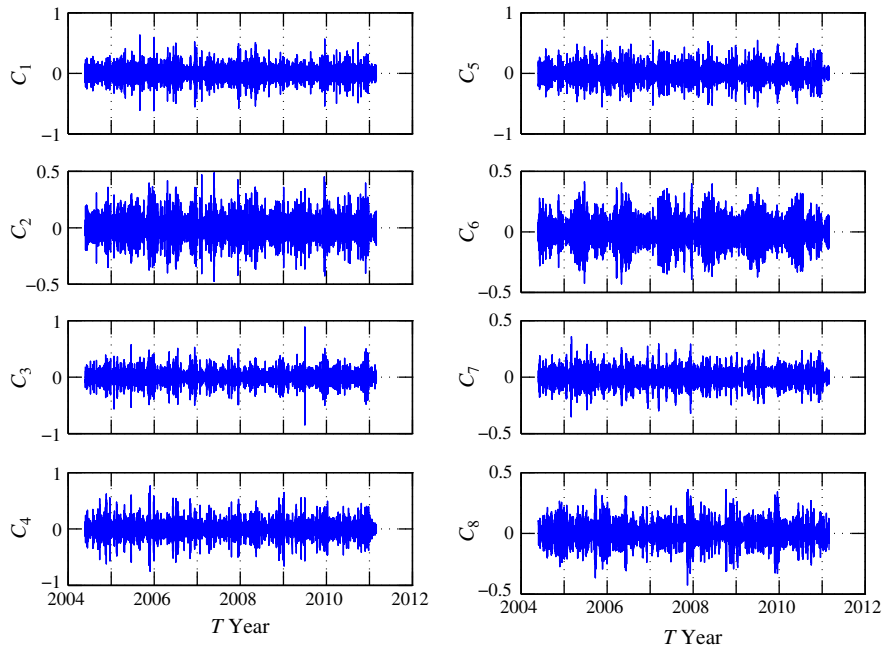


Fig. 6. The first-eighth IMF modes for temperature. The time scale is increasing with the mode index n .

1995; Flandrin, 1998; Huang et al., 1998). The estimation of the cross correlation between time series, in a multi-scale framework, may use a window based on the local characteristic scale given by the data itself.

Recently, Chen et al. (2010) proposed to use the EMD to estimate an adaptive window in order to calculate a so-called time-dependent intrinsic correlation (TDIC). The TDIC of each pair of IMFs is defined as follows:

$$R_{ij}(t_k^n | t_w^n) = \text{Corr}(C_{1,i}(t_w^n) C_{2,j}(t_w^n)) \text{ at any } t_k \quad (8)$$

where Corr is the cross correlation coefficient of two time series and $t_w^n = [t_k - nt_d/2 : t_k + nt_d/2]$ is the sliding window. The minimum sliding window size for the local correlation estimation is chosen as $t_d = \max(T_{1i}(t_k), T_{2j}(t_k))$, where T_{1i} and T_{2j} are the instantaneous periods $T = \omega^{-1}$, and n is any positive real number (Chen et al., 2010). The efficiency of this approach to characterize the relation between two time series for several problems has been shown in (Chen et al., 2010). Here we use this new method (with $n \geq 1$) to estimate cross correlations between coastal marine time series, where the latter can be considered as typical nonlinear and non-stationary data.

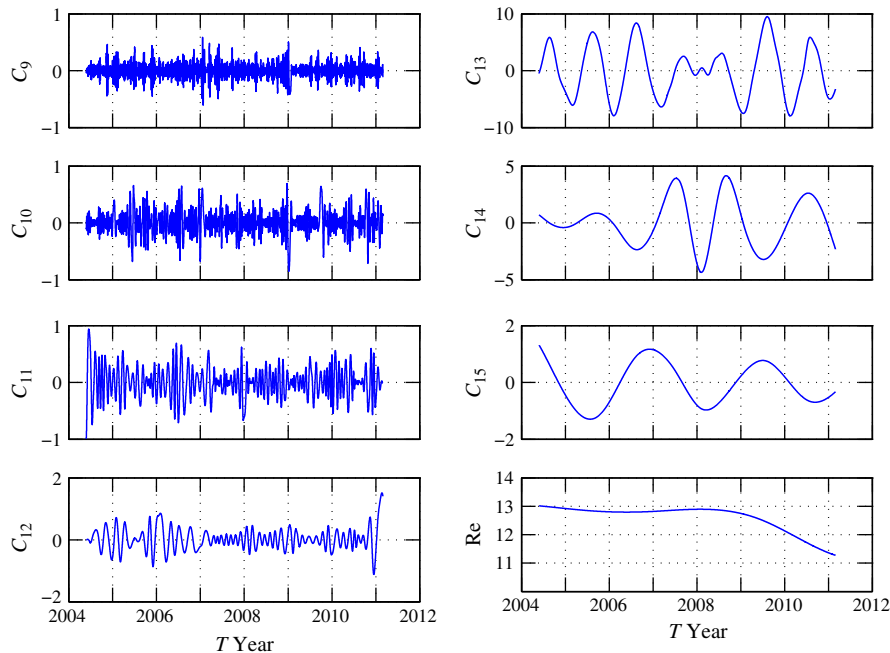


Fig. 7. The last-seventh IMF modes together with the trend term for temperature. The time scale is increasing with the mode index n . Note that there is a serious mode mixing problem for the 13th IMF mode around 2008. The mean period of this mode is around 1 year. The mode mixing is caused by a riding-wave phenomenon from January 2008 to April with a typical time scale of 3 months, see Fig. 2b.

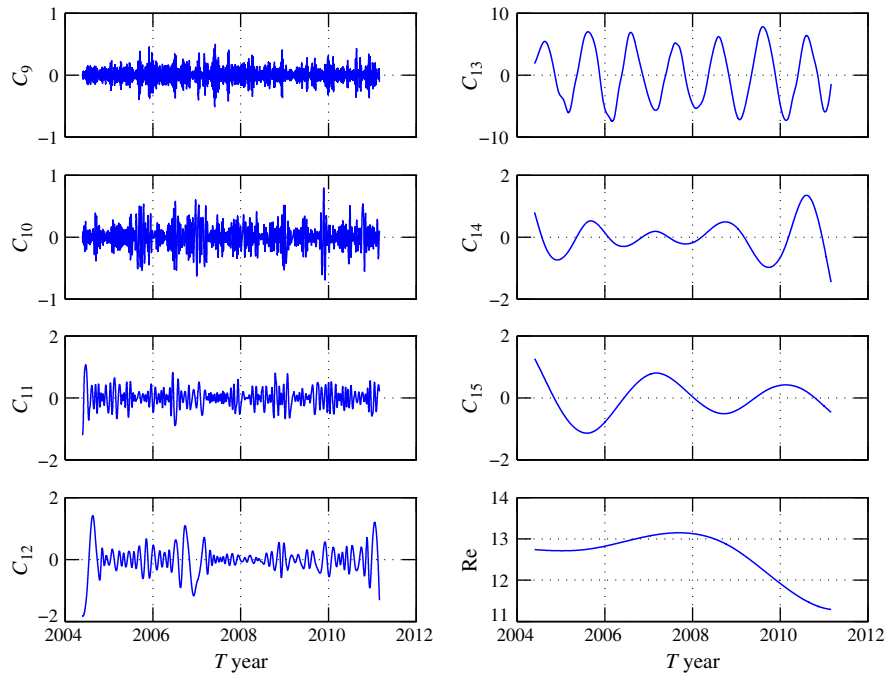


Fig. 8. The last-seventh IMF modes together with the trend term for temperature by removing the riding-wave. A better annual cycle, i.e., C_{13} , is obtained. Note that the trend is not influenced by removing the riding-wave.

3. Presentation of the experimental database

The time series analyzed here belong to the MAREL network (Automatic monitoring network for littoral environment, Ifremer, France¹). Such a system is based on the deployment of moored buoys equipped with physico-chemical measuring devices, working in continuous and autonomous conditions (Blain et al., 2004; Woerther, 1998). These stations use automatic systems for seawater analysis and real time data transmission. They record several parameters, such as temperature, salinity, dissolved oxygen, pH, and turbidity, with a fixed time resolution. The measuring station used here, the MAREL Carnot station, is situated in the eastern English Channel in the coastal waters of Boulogne-sur-mer (France) at position 50.7404 N, 1.5676 W (Fig. 1) and records data with a 20 min resolution (Zongo et al., 2011). Water depth at this position varies between 5 and 11 m; the measurements are done using a floating system inserted in a tube, 1.5 m below the surface. Statistical and scaling properties of MAREL automatic measurements in the English Channel have been studied previously, considering temperature time series (Dur et al., 2007) and turbidity, oxygen and pH (Schmitt et al., 2008) in the Seine estuary, and pH fluctuations in both the Seine estuary and Boulogne-sur-mer waters (Zongo and Schmitt, 2011).

Among other parameters, the measuring buoy provides temperature, salinity and dissolved oxygen (DO) concentration data, as well as percentage of DO relative to the dissolved oxygen at equilibrium (SaO_2) for the same temperature and salinity. The DO time series is directly linked to temperature due to the solubility equation given below, whereas SaO_2 is free from such direct influence. In the following, the latter dissolved oxygen series is considered. The oxygen saturation percentage equation is given by Garcia and Gordon, as a nonlinear function of the form (Garcia and Gordon, 1992):

$$SaO_2 = \frac{100DO}{\exp(P_1(T) + SP_2(T) + C_0S^2)} \quad (9)$$

where the denominator is the nonlinear fit expressing the oxygen solubility, S and T are salinity and temperature and P_1 and P_2 two

polynomial developments (P_1 of degree five, and P_2 of degree three). Their coefficients are given in Garcia and Gordon (1992). SaO_2 data have thus no dimension and values exceeding 100% correspond to supersaturation while values below 100% correspond to undersaturation associated to oxygen depletion; values lower than 50% are found for anoxic waters which can be dangerous for many species.

Due to occasional failures in the measuring system, and to regular maintenance, the time series sampling frequency is not always 20 min; there are frequently missing data and the mean sampling period is 24 min for both data sets. A total of 145,587 data points for dissolved oxygen collected from 25th May 2004 to 1st March 2011 and 155,825 data points for sea temperature collected from

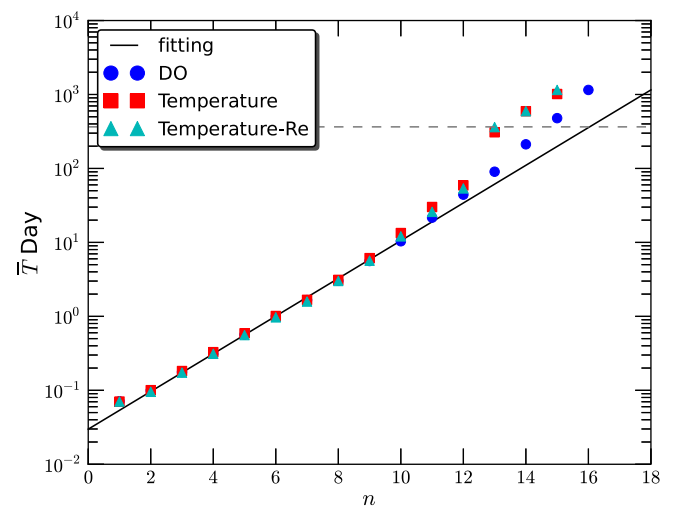


Fig. 9. Mean period of IMF modes obtained for dissolved oxygen (\circ), temperature with (\square) and without (Δ) riding-wave, respectively. The annual cycle is indicated by a dashed line. A better annual cycle for the temperature is retrieved by removing the riding-wave. An exponential law is observed for both data sets with an exponent 1.78 ± 0.04 , indicating a quasi dyadic-like filter bank of the EMD algorithm. Note that the first twelve IMF modes for both data sets approximately have the same mean period.

¹ <http://www.ifremer.fr/marel>.

24th March 2004 to 1st March 2011 have been analyzed. Fig. 2 shows the two data sets. The temperature data displays a strong annual cycle. These two data sets have been analyzed on the same time period 25th May 2004 to 1st March 2011. There is a warm winter in 2007 and 2008, showing also a weak oscillation of the temperature of the sea water.

The degree of stationary which was introduced by (Chen et al., 2010) is first calculated, i.e.,

$$\mathcal{D}(\tau) = \text{std}(R(t_k^n, \tau)) \quad (10)$$

in which $R(t_k^n, \tau)$ is an autocorrelation function with a sliding window t_k^n and time delay τ . The value of \mathcal{D} depends on the window size and the time delay τ . It varies from 0 to 1 and characterizes the deviation from a stationary process (Chen et al., 2010): 0 for exactly stationary processes and the larger the value, the more it is non-stationary. Fig. 3 shows the measured degree of stationary \mathcal{D} for the autocorrelation and cross-correlation with a sliding window size of $t_w = 90$ days and $t_w = 365$ days. To show the validation of Eq. (10), a test for white noise is performed, which is illustrated by a dashed line. As expected a small value of \mathcal{D} is found for white noise. These results applied on two experimental time series indicate that both are non-stationary and that the temperature is more stationary than the dissolved oxygen.

4. Analysis results

4.1. Empirical mode decomposition result

The EMD algorithm is applied to the two observed data sets for the same time period without interpolating the data into a regular time interval, since in the EMD step the information is contained in the local extrema points and can be applied to the time series with irregular time intervals (Flandrin and Gonçalves, 2004; Huang et al., 1998). After the EMD decomposition, there are respectively, 16 and 15 IMF modes with one residual for the dissolved oxygen and temperature. Figs. 4 to 7 show the corresponding IMF modes. Note that the time

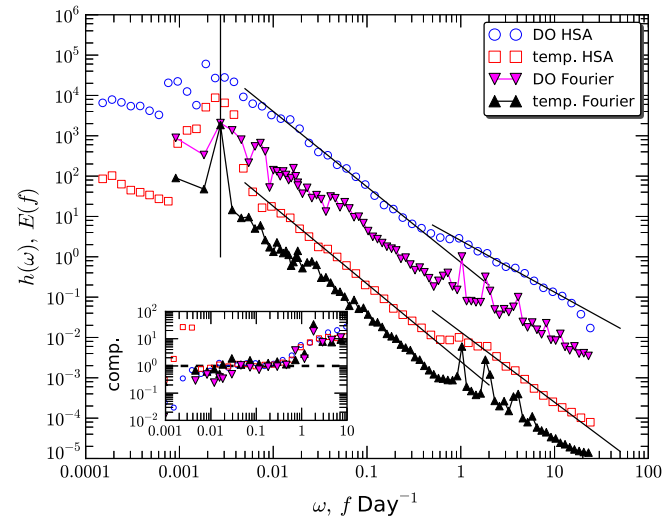


Fig. 10. The Hilbert spectrum of dissolved oxygen (○) and temperature (□), in which the vertical line indicates the annual cycle. Power law behavior is observed for both curves on the range $0.01 < \omega < 0.5$ day and $2 < \omega < 20$ day⁻¹, corresponding to a time scale $2 < T < 100$ days and $1.2 < T < 12$ hours, respectively. The corresponding scaling exponents are 1.93 ± 0.05 and 1.87 ± 0.08 for large scales, and 1.68 ± 0.10 and 1.35 ± 0.10 for small scales respectively. For comparison, the Fourier power spectrum is also shown for both dissolved oxygen (▽) and temperature (△). The inset shows the compensated spectrum $h(\omega)\omega^2C^{-1}$ ($E(f)f^2C^{-1}$) with fitted C to emphasize the observed power laws.

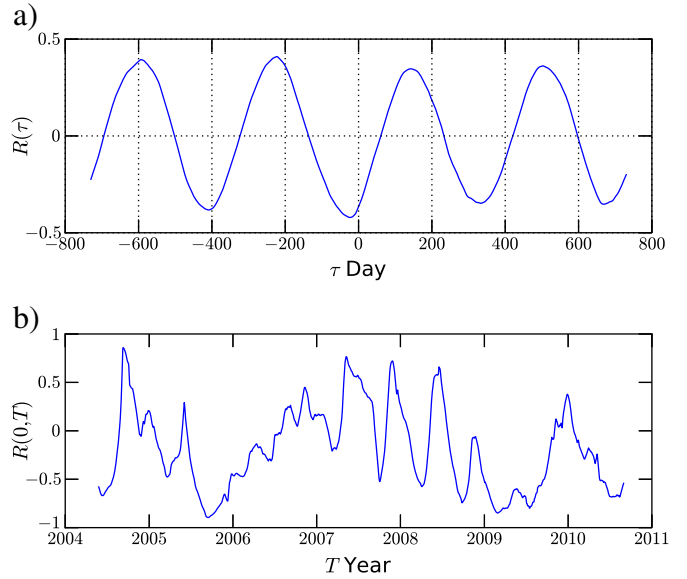


Fig. 11. a) The global cross correlation coefficient $R(\tau)$ between the dissolved oxygen and the temperature. Note that there is a strong annual cycle with a phase difference of 30 days. The global correlation coefficient is $R(0) = -0.37 \pm 0.10$. b) The measured cross correlation coefficient $R(0, t_w)$ with a sliding window $t_w = 180$ days. It shows that the relation between dissolved oxygen and temperature varies from time to time, indicating a multi-scale property of such data sets.

scale is increasing with the index n : the number of the IMF mode. The IMF modes capture the local variation of the time scales. Note that there is a serious mode mixing problem for the 13th IMF mode of the temperature around 2008, see Fig. 7. This is caused by a riding-wave phenomenon in the original temperature data (Huang et al., 1998), see Fig. 2 b. To eliminate the mode mixing problem, a noise assistant method, namely ensemble empirical mode decomposition (EEMD) could be applied to the data (Wu and Huang, 2009). However, practically speaking, the EEMD will introduce additional scales and bring bias to the estimation of the Hilbert spectrum $h(\omega)$. Therefore the riding-wave was removed manually. Fig. 8 shows the last-seventh IMF modes without the riding-wave: it provides a better 13th IMF mode. Note that the trend term is not influenced by the riding-wave. The mean period of each IMF mode is estimated by calculating the local extrema points and zero-crossing points (Huang et al., 2009b; Rilling et al., 2003), i.e.,

$$\bar{T}(n) = \frac{L}{N_{n,\max} + N_{n,\min} + N_{n,0}} \times 4 \quad (11)$$

in which L is the length of the data and n is the index. Fig. 9 shows the measured \bar{T} in a semilog plot. The first 12 IMF modes of the two

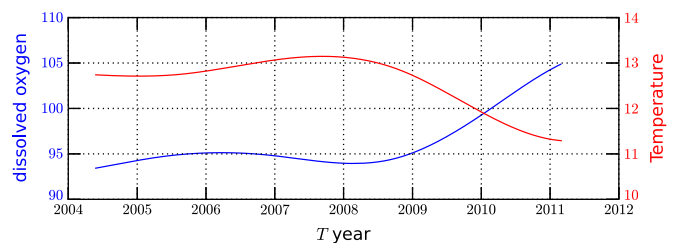


Fig. 12. The trend from EMD for dissolved oxygen and temperature. The direct measurement of the cross correlation is -0.96 , showing a global out-of-phase relation between the temperature and dissolved oxygen variation.

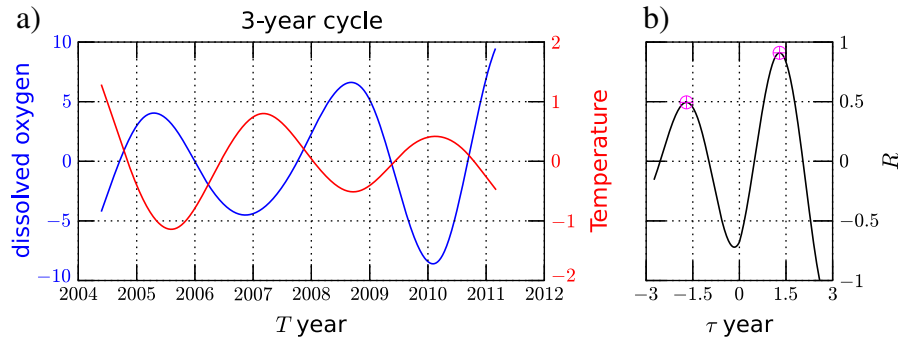


Fig. 13. a) The IMF modes with a 3-year mean period. b) The measured cross correlation. They are negatively correlated with a phase difference of around 2 months. The overall correlation coefficient is $R(0) = -0.68$. The mean period is determined by the distance between the two local extrema points (\oplus).

data sets approximately have the same mean period \bar{T} and follow an exponential law, i.e.,

$$\bar{T}(n) = \alpha \times \gamma^n \quad (12)$$

in which $\alpha \approx 0.035$ and $\gamma \approx 1.78$ obtained by using a least square fitting algorithm. This value is close to 2, which indicates a quasi dyadic filter bank property of the EMD algorithm for these series, as found in other situations (Flandrin et al., 2004; Huang et al., 2008, 2010; Wu and Huang, 2004). In other words, the mean scale of each mode series is 1.78 times the mean scale of the previous one.

4.2. Hilbert spectrum

With the IMF modes obtained from EMD algorithm, the HSA can be applied to each IMF mode. In this step, we first interpolate the IMF mode by using a Piecewise Cubic Hermite Interpolating Polynomial (phcip) algorithm into a regular time interval with $dt = 30$ min, which is slightly larger than the mean time interval of 24 min. Then the HSA is applied to extract the joint pdf $p(A, \omega)$. The Hilbert marginal spectrum is then calculated by using Eq. (7). Fig. 10 shows the Hilbert marginal spectra for dissolved oxygen (\circ) and temperature (\square), in which the vertical line indicates the annual cycle. A strong annual cycle is observed for the temperature data, which is consistent with the observation of the original temperature data, see Fig. 2b. It is emphasized here that for the temperature data with and without riding-wave around 2008, the Hilbert spectrum $h(\omega)$ is the same, showing the robustness of the present method. Power law behavior is observed on the frequency range $0.01 < \omega < 0.5 \text{ day}^{-1}$, corresponding to time scales $2 < T < 100$ days (resp. $T = \omega^{-1}$), and $2 < \omega < 20 \text{ day}^{-1}$, corresponding to time scales $1.2 < T < 12$ h. The measured scaling exponents are respectively 1.93 ± 0.05 and 1.87 ± 0.08 for the first scaling range of temperature and dissolved oxygen, and 1.68 ± 0.10 and 1.35 ± 0.10 for the second scaling range. The scaling

exponents for the first scaling range are statistically close to each other. For the second scaling exponent, the temperature one is close to the Kolmogorov value $5/3$ (Frisch, 1995), implying that the variation of the temperature within one day is like a passive scalar. For comparison, the corresponding Fourier power spectrum $E(f)$ is also presented. Both Hilbert and Fourier spectra show strong annual and daily cycles. However, there is no half-day cycle observed in the Hilbert spectrum. It might indicate that the half-day cycle is harmonic from the nonlinear distortion of the daily cycle. Power law behavior is also observed for the Fourier power spectrum on the same range of time scale $2 < T < 100$ days as for the Hilbert framework. The corresponding scaling slope is close to 2. To emphasize the observed power law behavior, the compensated spectra $h(\omega)\omega^2 \text{ C}^{-1}$ (resp. $E(f)f^2 \text{ C}^{-1}$) are shown in the inset. The observed plateau confirms the power law. Note that the observed -2 power law indicates that there might be a cascade process in the marine coastal environment (Thorpe, 2005). This should be investigated further in future studies. Let us note that a shorter portion of this data set has been studied in (Zongo and Schmitt, 2011): for the years 2006 to 2008, with about 50,000 data points, a Fourier analysis has found a power-law slope of 1.66 for scales between 1 year and 2 h, where the fit was done for this whole range; it should be noted that different fits on separated portions of the frequencies give values closer to the Hilbert-based estimates.

4.3. Time dependent intrinsic correlation

The correlation between this two data sets is considered here. Fig. 11 displays the global cross correlation $R(\tau)$ between the dissolved oxygen and the temperature. There is strong annual cycle with a 30 day phase difference. The global correlation coefficient is found to be $R(0) = -0.37 \pm 0.10$. To show the variation of R , the cross correlation coefficient $R(0, t_w)$ is calculated with a sliding window $t_w = 180$ days, which is shown in Fig. 11b. The standard

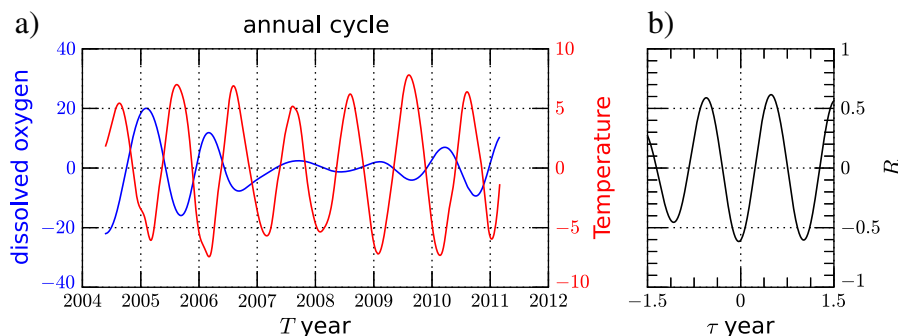


Fig. 14. a) The annual cycle from EMD. b) The measured cross correlation. The overall correlation coefficient is $R(0) = -0.60$ with 14 days phase difference. Note that there is a weak period between 2007 and 2009 for both temperature and dissolved oxygen.

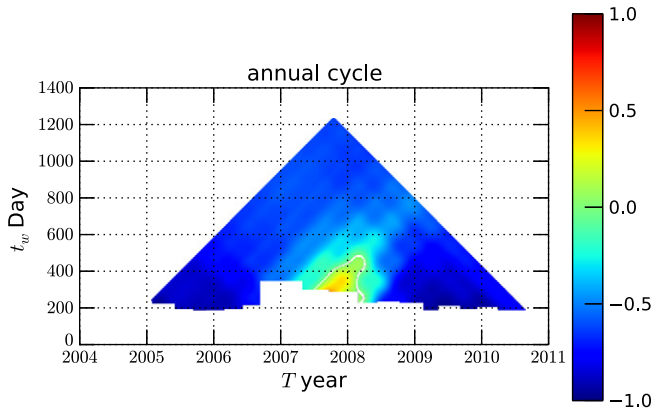


Fig. 15. The measured TDIC for the annual cycle. It shows a transition from negative positive correlation between 2007 and 2008. The overall correlation coefficient is -0.60 , see Fig. 14b. The hole is the R cannot pass the t -test. The horizontal axis is the location of the center of the sliding window.

deviation is found to be 0.40 . Note that the standard deviation as a measurement of the stationarity has been introduced, see Eq. (10). It confirms that the global cross correlation coefficient ignores the local/multi-scale information, which could be recovered by a proper methodology (Chen et al., 2010).

After the EMD decomposition, the data sets are represented in a multi-scale way (Flandrin and Gonçalves, 2004; Huang et al., 1998; Huang, 2009). These are used for the multi-scale correlation analysis. Time scales larger than or equal to one year are now considered. Fig. 12 shows the residual from the EMD algorithm, which has been recognized as the trend of the given data (Moghtaderi et al., 2011; Wu et al., 2007). Fig. 13a) shows the IMF modes with a 3-year mean period, and Fig. 13b) the corresponding cross correlation $R(\tau)$. They show a negative correlation with each other with a phase difference of 2 months. Fig. 14a) shows the annual cycle from EMD, Fig. 14b) the corresponding cross correlation $R(\tau)$. Again they are negatively correlated with each other. The overall correlation coefficient is -0.63 with a 14 day phase difference. An interesting observation is that there is a weak period between 2007 and 2009 for both data sets, indicating that a special event has happened. The measured TDIC is shown in Fig. 15. There is a positive correlation around 2008. The transition from negative (resp. out-of-phase) to positive (in-phase) and again to negative correlation from mid of 2007 to the mid of 2008 might be an effect of the warming from mid 2007 until almost 2009. During this period, the sea water temperature was consequently 2°C higher than usual in the winter time. Therefore, the out-of-phase relation between DO and temperature is affected by this event.

Let us consider scales less than one year. Fig. 16 shows the IMF modes with a mean period of 60 days. They are positively correlated with each other on some portions and negatively correlated on others, showing rich dynamics. However, the overall correlation coefficient is small (resp. -0.02). Fig. 17 displays the measured

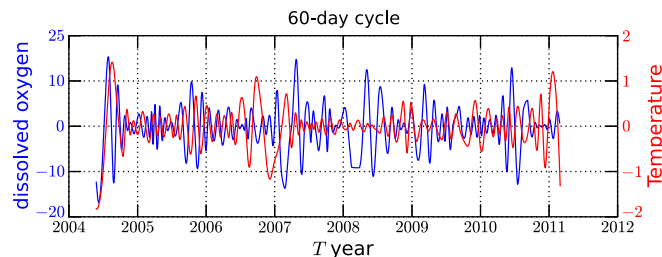


Fig. 16. The 60-day cycle from EMD. The overall correlation coefficient is -0.02 with 15 days phase difference.

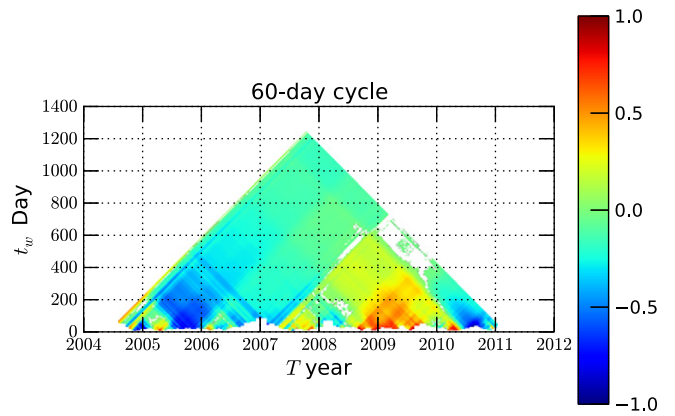


Fig. 17. The measured TDIC for 60-day mean period. There is a more than half year strong correlation pattern around 2009 and 2010. The overall correlation coefficient is -0.02 .

TDIC, which confirms the direct observation of the IMF modes. A strong positive correlation is observed during the time period 2009–2010. Due to the complexity of the coastal environment system, the exact reason for these changes from negative to positive is unclear. It will be addressed in future studies. For time scales smaller than 60 days, the original IMF is not shown here, but only the measured TDIC. Figs. 18–21 show the measured TDIC for mean periods of 30-day to 13-hour respectively. All TDICs show rich patterns at a small sliding window. Let us also note that with the increase of the window size they decorrelate. To show the decorrelation of the TDIC, an absolute mean value of the measured $R_i(t_k^n)$ is defined, i.e.,

$$\bar{R}(t_w) = \langle |R(t_k^n, t_w)| \rangle_t. \quad (13)$$

Fig. 22 shows in a log–log plot of the measured $\bar{R}(t_w)$ for seven time scales considered above. Note that the horizontal axis t_w has been normalized by their mean period \bar{T} . Generally speaking, the curves are decreasing with the increasing of the sliding window size t_w/\bar{T} , showing the decorrelation between dissolved oxygen and temperature on different scales. A power law behavior is observed on $0.4 < t_w/\bar{T} < 4$, i.e.,

$$\bar{R} \sim (t_w/\bar{T})^{-\alpha} \quad (14)$$

with a scaling exponent $\alpha = 0.40 \pm 0.10$ obtained by using a least square fitting algorithm. This indicates that the considered modes decorrelate with each other in a power law way, with the same law.

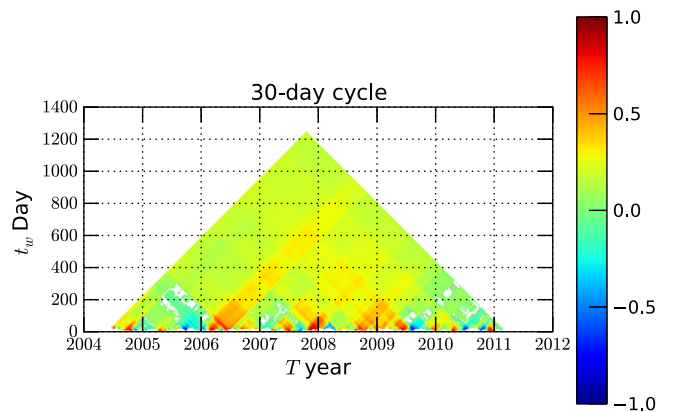


Fig. 18. The measured TDIC for 30-day mean period on the range May 2004–January 2008. There are more than half a year strong correlation patterns between January 2006 and June 2006. The overall correlation coefficient is 0.17 .

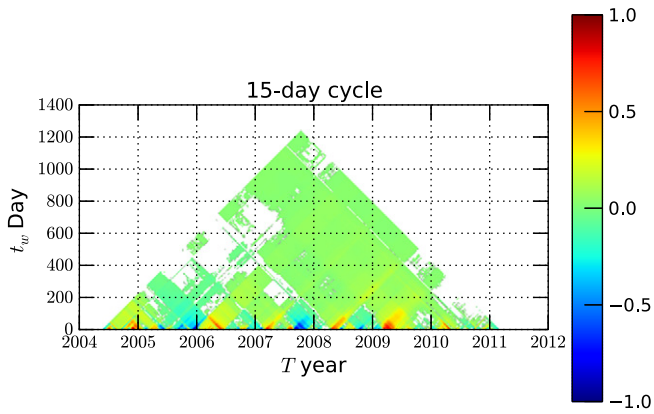


Fig. 19. The measured TDIC for the 15-day mean period. The overall correlation coefficient is 0.10.

5. Conclusion

Marine environmental time series can be quite complex, especially with the influence of turbulent stochastic fluctuations, multi-scale dynamics, and deterministic forcings. Such time series are as a result often nonstationary and nonlinear. It has been shown elsewhere that the EMD method, together with Hilbert spectral analysis, is less influenced by deterministic forcing than other methods (Huang et al., 2010, 2011). Here this approach was applied to automatic temperature and oxygen data. Power law spectra were found using Hilbert spectral analysis. Both series have similar spectra over the range from 2 to 100 days, whereas there is a difference in slopes for high frequency power law regimes. The time evolution and scale dependence of cross correlations between both series were considered. The decomposition into modes helped to estimate how correlations vary among scales. It was found that the trends are perfectly anti-correlated, and that the modes of mean year 3 years and 1 year have also negative correlation, whereas higher frequency modes have a much smaller correlation. A new methodology was also applied, time-dependent intrinsic correlations: this showed the patterns of correlations at different scales, for different modes.

Such analysis may help to identify some key mechanisms. Here a strong anti-correlation was found for modes of mean time scale of 3 years, with a phase shift of 2 months. An explanation of such a result could be that, at these scales, higher temperatures may favor larger phytoplankton growth rate, and hence, with a time delay, a lower percentage of oxygen. Of course such relations are not always

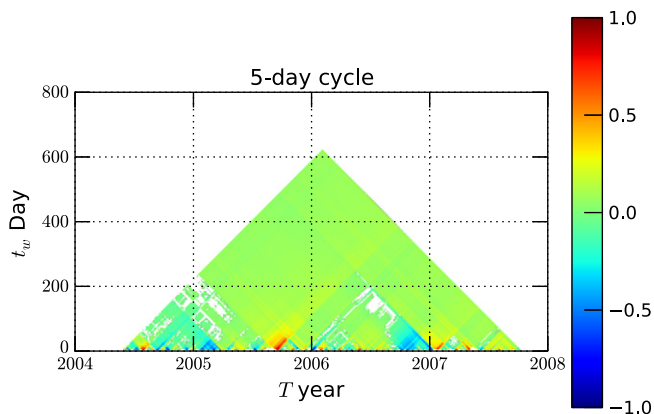


Fig. 20. The measured TDIC for the 5-day mean period. The overall correlation coefficient is 0.1.

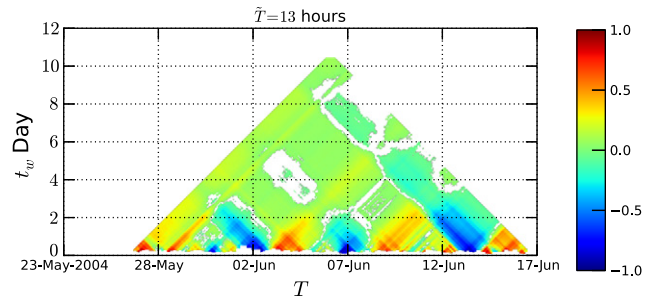


Fig. 21. The measured TDIC for the 13-day mean period. The overall correlation coefficient is 0.1. The holes indicate that the measured TDIC can not pass the t -test.

true, and the color maps of TDIC reveal the variations of the strength of such relationship.

While Fourier space methods can produce in some respects similar results for the power spectra, cross-correlations through co-spectra, and some color maps of correlations in frequency space, they do not provide time-frequency information like the EMD based method used here. Furthermore, the TDIC tested here is a new method for correlation analysis that could be applied on other time series from the environmental and oceanic sciences, since in these fields time series are typically complex, with fluctuations at all scales, and nonlinear and non stationary features.

Acknowledgments

This work is sponsored in part by the National Natural Science Foundation of China (Nos. 11072139, 11032007 and 11202122), and 'Pu Jiang' Project of Shanghai (No. 12PJ1403500) and the Shanghai Program for Innovative Research Team in Universities. The EMD MATLAB codes used in this study are written by Dr Gabriel Rilling and Prof. Patrick Flandrin from the laboratoire de Physique, CNRS & ENS Lyon (France): <http://perso.ens-lyon.fr/patrick.flandrin/emd.html>. We thank Ifremer, and especially Alain Lefebvre and Michel Repecaud for their work in recording and validating the Marel data. We thank the I. Puillat and the other organizers of the Time Series Brest conference for a very nice meeting, and the reviewers for useful comments and suggestions. We thank also Denis Marin (LOG, Dunkerque, France) for the realization of Fig. 1.

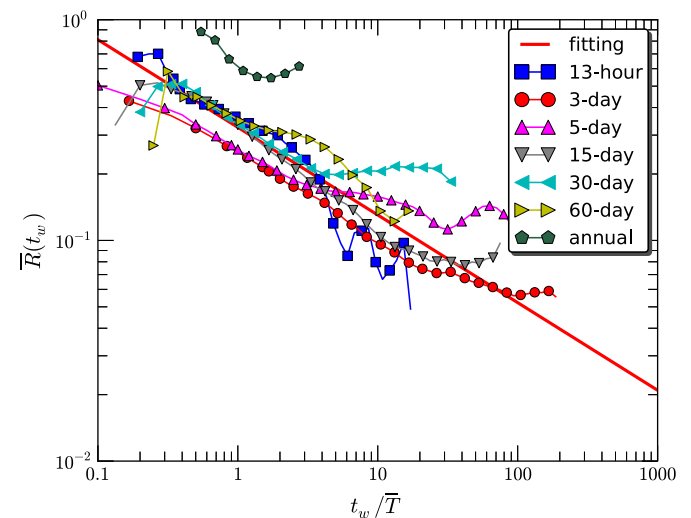


Fig. 22. The measured absolute mean value $\bar{R}(t_w)$ for different time scales \bar{T} . For display clarity, the sliding window t_w is normalized by their own mean period \bar{T} . Power law behavior is observed on the $0.4 < t_w/\bar{T} < 4$ with a scaling exponent 0.40 ± 0.01 .

References

- Best, M., Wither, A., Coates, S., 2007. Dissolved oxygen as a physico-chemical supporting element in the water framework directive. *Mar. Pollut. Bull.* 55, 53–64.
- Blain, S., Guillou, J., Treguer, P., Woerther, P., Delauney, L., Follenfant, E., Gontier, O., Hamon, M., Leilide, B., Masson, A., Tartub, C., Vuillemin, R., 2004. High frequency monitoring of the coastal environment using the marel buoy. *J. Environ. Monit.* 6, 569–575.
- Chang, G., Dickey, T., 2001. Optical and physical variability on timescales from minutes to the seasonal cycle on the new England shelf: July 1996 to June 1997. *J. Geophys. Res.* 106, 9435–9453.
- Chavez, et al., 1997. Moorings and drifters for real-time interdisciplinary oceanography. *J. Atmos. Ocean. Technol.* 14, 1199–1211.
- Chen, X., Wu, Z., Huang, N.E., 2010. The time-dependent intrinsic correlation based on the empirical mode decomposition. *Adv. Adapt. Data Anal.* 2, 233–265.
- Cohen, L., 1995. Time-frequency Analysis. Prentice Hall PTR Englewood Cliffs, NJ.
- Coughlin, K.T., Tung, K.K., 2004. 11-Year solar cycle in the stratosphere extracted by the empirical mode decomposition method. *Adv. Space Res.* 34 (2), 323–329.
- Dätig, M., Schlurmann, T., 2004. Performance and limitations of the Hilbert–Huang transformation (hht) with an application to irregular water waves. *Ocean Eng.* 31 (14), 1783–1834.
- Dickey, T., 1991. The emergence of concurrent high resolution physical and bio-optical measurements in the upper ocean and their applications. *Rev. Geophys.* 29, 383–413.
- Dur, G., Schmitt, F.G., Souissi, S., 2007. Analysis of high frequency temperature time series in the seine estuary from the marel autonomous monitoring buoy. *Hydrobiologia* 588.
- Echeverria, J.C., Crowe, J.A., Woolfson, M.S., Hayes-Gill, B.R., 2001. Application of empirical mode decomposition to heart rate variability analysis. *Med. Biol. Eng. Comput.* 39 (4), 471–479.
- Flandrin, P., 1998. Time–frequency/time–scale Analysis. Academic Press.
- Flandrin, P., Gonçalves, P., 2004. Empirical mode decompositions as data-driven wavelet-like expansions. *Int. J. Wavelets, Multires. Info. Proc.* 2 (4), 477–496.
- Flandrin, P., Rilling, G., Gonçalves, P., 2004. Empirical mode decomposition as a filter bank. *IEEE Signal Proc. Lett.* 11 (2), 112–114.
- Frisch, U., 1995. Turbulence: The Legacy of A. N. Kolmogorov. Cambridge University Press.
- Garcia, H.E., Gordon, L.I., 1992. Oxygen solubility in seawater: better fitting equations. *Limnol. Oceanogr.* 37, 1307–1312.
- Hoover, K., 2003. non-stationary time series cointegration and the principle of the common cause. *British J. Philos. Sci.* 54, 527–551.
- Huang, N.E., 2005. Hilbert–Huang transform and its applications. World Scientific, Ch. 1. Introduction to the Hilbert–Huang Transform and Its Related Mathematical Problems, pp. 1–26.
- Huang, Y., 2009. Arbitrary-order Hilbert Spectral Analysis: Definition and Application to Fully Developed Turbulence and Environmental Time Series. (Ph.D. thesis) Université des Sciences et Technologies de Lille – Lille 1, France & Shanghai University, China.
- Huang, N.E., Shen, Z., Long, S.R., Wu, M.C., Shih, H.H., Zheng, Q., Yen, N., Tung, C.C., Liu, H.H., 1998. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. London, Ser. A* 454 (1971), 903–995.
- Huang, N.E., Shen, Z., Long, S.R., 1999. A new view of nonlinear water waves: the Hilbert spectrum. *Annu. Rev. Fluid Mech.* 31 (1), 417–457.
- Huang, Y., Schmitt, F., Lu, Z., Liu, Y., 2008. An amplitude–frequency study of turbulent scaling intermittency using Hilbert spectral analysis. *Europhys. Lett.* 84, 40010.
- Huang, N.E., Wu, Z., Long, S., Arnold, K., Chen, X., Blank, K., 2009a. On instantaneous frequency. *Adv. Adapt. Data Anal.* 1 (2), 177–229.
- Huang, Y., Schmitt, F., Lu, Z., Liu, Y., 2009b. Analysis of daily river flow fluctuations using empirical mode decomposition and arbitrary order Hilbert spectral analysis. *J. Hydrol.* 373, 103–111.
- Huang, Y., Schmitt, F., Lu, Z., Fougairolles, P., Gagne, Y., Liu, Y., 2010. Second-order structure function in fully developed turbulence. *Phys. Rev. E* 82 (2), 026319.
- Huang, Y., Schmitt, F.G., Hermand, J.-P., Gagne, Y., Lu, Z., Liu, Y., 2011. Arbitrary-order Hilbert spectral analysis for time series possessing scaling statistics: comparison study with detrended fluctuation analysis and wavelet leaders. *Phys. Rev. E* 84 (1), 016208.
- Hwang, P.A., Huang, N.E., Wang, D.W., 2003. A note on analyzing nonlinear and non-stationary ocean wave data. *Appl. Ocean Res.* 25 (4), 187–193.
- Jánosi, I., Müller, R., 2005. Empirical mode decomposition and correlation properties of long daily ozone records. *Phys. Rev. E* 71 (5), 56126.
- Long, S.R., Huang, N.E., Tung, C.C., Wu, M.L., Lin, R.Q., Mollo-Christensen, E., Yuan, Y., 1995. The Hilbert techniques: an alternate approach for non-steady time series analysis. *IEEE Geosci. Remote Sens. Soc. Lett.* 3, 6–11.
- Loutridis, S.J., 2004. Damage detection in gear systems using empirical mode decomposition. *Eng. Struct.* 26 (12), 1833–1841.
- Moghtaderi, A., Borgnat, P., Flandrin, P., 2011. Trend filtering: empirical mode decompositions versus I_1 and Hodrick–Prescott. *Adv. Adapt. Data Anal.* 3 (01n02), 41–61.
- Molla, M.K.I., Rahman, M.S., Sumi, A., Banik, P., 2006a. Empirical mode decomposition analysis of climate changes with special reference to rainfall data. *Discret. Dyn. Nat. Soc.* <http://dx.doi.org/10.1155/DDNS/2006/45348> (Article ID 45348, 17 pages).
- Molla, M.K.I., Sumi, A., Rahman, M.S., 2006b. Analysis of temperature change under global warming impact using empirical mode decomposition. *Int. J. Inf. Technol.* 3 (2), 131–139.
- Papadimitriou, S., Sun, J., Yu, P.S., 2006. Local correlation tracking in time series. *Proc. Sixth Int. Conf. Data Mining* 456–465.
- Rilling, G., Flandrin, P., Gonçalves, P., 2003. On empirical mode decomposition and its algorithms. *IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing*.
- Rodo, X., Rodriguez-Arias, M.A., 2006. A new method to detect transitory signatures and local time/space variability structures in the climate system: the scale-dependent correlation analysis. *Clim. Dyn.* 27, 441–458.
- Schmitt, F.G., Dur, G., Souissi, S., Brizard Zongo, S., 2008. Statistical properties of turbidity, oxygen and pH fluctuations in the Seine river estuary (France). *Physica A* 387 (26), 6613–6623.
- Schmitt, F.G., Huang, Y., Lu, Z., Liu, Y., Fernandez, N., 2009. Analysis of velocity fluctuations and their intermittency properties in the surf zone using empirical mode decomposition. *J. Mar. Syst.* 77, 473–481.
- Thorpe, S.A., 2005. The Turbulent Ocean. Cambridge University Press.
- Veltcheva, A.D., Soares, C.G., 2004. Identification of the components of wave spectra by the Hilbert Huang transform method. *Appl. Ocean Res.* 26 (1–2), 1–12.
- Woerther, P., 1998. Marel: Mesures automatisées en réseau pour l'environnement littoral. Léau, Les industrie, Les nuisances 217, 67–72.
- Wu, Z., Huang, N.E., 2004. A study of the characteristics of white noise using the empirical mode decomposition method. *Proc. R. Soc. Lond. Ser. A* 460, 1597–1611.
- Wu, Z., Huang, N.E., 2009. Ensemble empirical mode decomposition: a noise-assisted data analysis method. *Adv. Adapt. Data Anal.* 1, 1–41.
- Wu, Z., Huang, N.E., Long, S.R., Peng, C., 2007. On the trend, detrending, and variability of nonlinear and non-stationary time series. *PNAS* 104 (38), 14889.
- Zhu, X., Shen, Z., Eckermann, S.D., Bittner, M., Hirota, I., Yee, J.H., 1997. Gravity wave characteristics in the middle atmosphere derived from the empirical mode decomposition method. *J. Geophys. Res.* 102, 16545–16561.
- Zongo, S.B., Schmitt, F.G., 2011. Scaling properties of pH fluctuations in coastal waters of the English channel: pH as a turbulent active scalars. *Nonlinear Process. Geophys.* 18, 829–839.
- Zongo, S., Schmitt, F., Lefebvre, A., 2011. Observations biogéochimiques des eaux côtières à boulogne-sur-mer à haute fréquence: les mesures automatiques de la bouée marel. In: Schmitt, F.G. (Ed.), *Du naturalisme à l'écologie*. Presses de l'UOF, pp. 253–266.