Using empirical mode decomposition to correlate paleoclimatic time-series

J. Solé¹, A. Turiel², and J. E. Llebot¹

¹Department of Physics, Universitat Autonoma de Barcelona. Campus de la UAB 08193 Bellaterra (Cerdanyola del Vallès), Catalunya, Spain
²Institute of Marine Sciences, CSIC. Pg. Marítim de la Barceloneta 37-49, 08003 Barcelona, Catalunya, Spain

Received: 23 October 2006 – Revised: 15 March 2007 – Accepted: 3 April 2007 – Published: 17 April 2007

Abstract. Determination of the timing and duration of paleoclimatic events is a challenging task. Classical techniques for time-series analysis rely too strongly on having a constant sampling rate, which poorly adapts to the uneven time recording of paleoclimatic variables; new, more flexible methods issued from Non-Linear Physics are hence required. In this paper, we have used Huang’s Empirical Mode Decomposition (EMD) for the analysis of paleoclimatic series. We have studied three different time series of temperature proxies, characterizing oscillation patterns by using EMD. To measure the degree of temporal correlation of two variables, we have developed a method that relates couples of modes from different series by calculating the instantaneous phase differences among the associated modes. We observed that when two modes exhibited a constant phase difference, their frequencies were nearly equal to that of Milankovich cycles. Our results show that EMD is a good methodology not only for synchronization of different records but also for determination of the different local frequencies in each time series. Some of the obtained modes may be interpreted as the result of global forcing mechanisms.

1 Introduction

The progress over the past decades in the extraction and processing of geological records of paleoclimatic data has lead to a relative increase, both in number and in time scope, of the data about past, distant dates of the planet (Cronin, 1999). However, to link the timing of significant events according to the different series with the desired accuracy presents great difficulties (Rahmstorf, 2003). These difficulties arise from the diversity of locations at which the probes were obtained, the differences in the physical and chemical processes giving rise to the concentration of the proxy species, and the inherent uncertainties in the determination of the precise time pace of geological deposition of the proxy elements (Saltzman, 2002). For the same reasons, any hypothesis directly based on the comparison of experimental data about the mechanisms underlying the evolution of the local temperature at different places on Earth can always be blamed as non-conclusive, due to this lack of confidence on the experimental record and the problems with its precise timing.

To solve the timing problem, the different series are usually put into correspondence by more or less manual methods based on the visual assessment of some precise time instants or events with strong signature in all the series (Bond et al., 1993; Morgan et al., 2002; Landais et al., 2004; Knutti et al., 2004; Cruz et al., 2005; Pahnke and Zahn, 2005): typical milestones for this manual correspondence are rapid transitions from stadial to inter-stadial periods, and some other events, like Heinrich events (Heinrich, 1988; Broecker, 1994), for those cases in which they induce a signature strong enough in the compared records. Therefore, such kind of assessment leads to a qualitative use of the data. In addition, it implies the necessity of performing an expertised re-analysis of the probes to understand the differences in event duration according to the different records, typically due to the changing environmental conditions influencing the fixation of the proxy to the substrate (Hemming, 2004).

In this context, it would be convenient to have a post-processing method which could help manipulating data in a more objective, automatized way. This method should be able to deal with data in a multiple resolution fashion, so enabling to distinguish among different processes involved by the proxy evolution. One goal would be to separate those (faster) processes specific to the particularities of the chosen proxy from the (slower) processes common to all proxies and which can thus confidently be taken as a manifestation of planetary (or at least non-local) changes. In addition, the method should help in clearly marking the starting and finish
instants of relevant events in a non-conventional way, so they could be used to unify the time reference for the different time series. Besides, the method should rely in a flexible scheme for describing oscillatory events, in which the associated time frequency could eventually evolve with time.

During the last decade a method satisfying all the requirements above has been devised. Empirical Mode Decomposition (EMD) (Huang et al., 1998, 2003; Wu and Huang, 2003) is a useful objective method for studying time series. Recently the EMD method has been employed over many different datasets such as cardiorespiratory synchronization (Wu and Hu, 2006), ozone records (Jánosi and Müller, 2006), precipitation variability related with El Nino (El-Askary et al., 2004), analysis of North Atlantic oscillation (Hu and Wu, 2004), analysis of solar insolation (Lin and Wang, 2006), space-time rainfall analysis (Sinclair and Peagram, 2005), ice-cover analysis (Gloersen and Huang, 2003) and analysis of temperature under global warming (Molla et al., 2006). The applicability of EMD to non-linear and non-stationary time series make it specially interesting for the study of series in which the stationarity can not be assured. EMD decomposes a given time series in oscillatory Intrinsic Mode Functions (IMF), determined by well-defined oscillatory characteristics of given resolution but time-depending frequency. For this reason, each IMF is guessed to be the effect of a different physical cause with different characteristic time. In this paper, we propose to employ EMD for the study of paleoclimatic series, putting the IMFs arising from the different series in correspondence, and so be able to interpret these IMFs as effects of a common physical cause.

For the present study, we have used three paleoclimatic series of ice-core temperature proxies: Grip (Dansgaard et al., 1993), Vostok (Petit et al., 1999), Epica (EPICA community members, 2004). Each proxy series are expressed in different units, related to the concentration of an appropriate radiative isotope with known decay properties. Thus, we are not interested in studying effects associated to the amplitude (which has dimensions and depends on the physical coupling constants between the proxy and the temperature) but those associated to the local frequency, which are directly related to the timing of the series.

So that, we first proceed to extract the IMFs associated to the different temperature-proxy series. We then just keep the information about the local frequencies of the IMFs to evaluate the correlations among the modes from different series. We proceed to compare the data series in pairs, trying to characterize a general, “universal” oscillating pattern; we intend to avoid high-frequency, noisy components in the signal and to quantify the main features of the oscillation.

Fig. 1. Time-series of temperature proxies. From top to bottom for: Grip ($\delta D$ in ‰), Vostok and EPICA ($\delta D$ in ‰). Time is expressed in kyr.
Materials and methods

A modified version of EMD method, developed in Huang et al. (1998), has been used to produce a linear decomposition of series in non-linear modes, the so-called Intrinsic Mode Functions (IMF). IMFs represent partial Hilbert transforms of the signal, and so they possess special properties such as the smoothness in both frequency and amplitude modulation. The EMD method is based on the existence of local time-scales of the data, what gives a precise meaning to local frequencies and allows to remove spurious harmonics (typical in Fourier analysis) in the representation of non-linear and non-stationary signals (Coughlin and Tung, 2004). The essence of EMD is to empirically identify oscillatory modes in the data by means of their local extrema. Such a decomposition is based in three assumptions (Huang et al., 1998, 2003; Wu and Huang, 2003): (1) the signal has at least two extrema—one maximum and one minimum; (2) the characteristic local time scale is defined by the time interval between two consecutive extrema; and (3) if the data were totally free of extrema but contained only inflection points, then the signal can be obtained by the integration of the components.

For this work we have used a MatLab implementation of the algorithm, provided by Patrick Flandrin and collaborators (Rilling et al., 2003; Flandrin et al., 2004). The code can be retrieved at the following URL: http://perso.ens-lyon.fr/patrick.flandrin/emd.html.

The series under study are represented in Fig. 1. They are GRIP 100 ky before present (BP) series, Vostok 400 ky BP series, and Epica 741 ky BP series (all downloaded at http://www.ncdc.noaa.gov/paleo/data.html). GRIP series describes the time evolution of the differences in concentration of the oxygen isotope ($^{18}$O) (Saltzman, 2002) $\delta^{18}$O (expressed in parts per thousand, ‰), which is directly proportional to the zonal temperature (Dahl-Jensen et al., 1998; Landais et al., 2004). These data have uneven time resolution, ranging from 1.3 years for the most recent records to 172 years for the oldest one (Dansgaard et al., 1993; Landais et al., 2003). Vostok time series (Petit et al., 1999) describes the time evolution of the differences in concentration of deuterium (D) (Saltzman, 2002) $\delta$D (expressed in parts per thousand, ‰) in a time span of 422 766 years. The time resolution of the data is 17 years for the earliest records and 664 yr in the most distant ones. Epica (EPICA community members, 2004) time series describes the time evolution of the concentration of $\delta$D (expressed in parts per thousand, ‰) in a time span of 741 kyr BP. The time resolution is 68.5 yr for the earliest records and 5421 yr for the most distant ones. Previous to the extraction of the IMF’s we pre-process the time series by calculating a uniform sampling surrogate of lower resolution. This surrogate is obtained by sampling the different series with boxes of 200 years for Grip, 500 years for Vostok and 3000 years for Epica, associating to each box a representative: the average of the points contained in that box interval. Such an average-downsampling acts much as a low-pass filter and diminishes fluctuations. Uniformly sub-sampling the series is not really necessary, as EMD should also work on the non-uniform series; however, this pre-process is convenient in order to give more stability to the procedure of IMF extraction, so reducing the number of iterations necessary to obtain each IMF. In fact, this uniform sampling is necessary to compare the IMFs coming from different time series.

Not all the IMFs directly emerging from EMD can be considered as physically significant. As a matter of fact, there must be some degree of noise contaminating the sample. We will assume that series are affected by additive, white noise. When a signal is affected by other types of noise (e.g. red noise), this will appear as an additional IMF component in the decomposition, which will utterly need to be interpreted. As EMD is a linear decomposition, we can assume that, after decomposition, each empirical IMF is the sum of a signal IMF plus a noise IMF. As noise contributes with constant amplitude at all frequencies, noise IMFs have the same, constant amplitude and frequency ranges coincident with those of the associated signal IMFs. As far as noise IMF amplitude is significantly smaller than signal IMF amplitude, we can consider the empirical IMF to be physically meaningful. Contrarily, when noise IMF amplitude is larger than signal IMF amplitude the noise will mask the structure of the signal IMF and hence the empirical IMF must be discarded. We can reasonably assume that the first empirical IMF, which comprises the highest frequencies, is completely corrupted by noise and we take its amplitude level as a reference of the amplitude of noise IMFs. So, we establish the following criterion to consider an empirical IMF physically significant: the amplitude of that IMF must be at least one third of the maximum amplitude of the first (i.e. highest frequency) IMF. Notice that, contrarily to noise, physical signals (for instance, finite variation signals) usually have amplitudes which increase as frequency decreases (what can be observed, for instance, when the power-spectrum is evaluated and a power-law decay is observed). Hence, as we consider following orders in the IMF decomposition frequency decreases and signal IMF amplitude increases, so they eventually become significantly larger than those of noise and the empirical IMF can then considered as reliable. This does not mean that as frequency decreases only physically significant empirical IMFs happen: sometimes, an additional, completely spurious IMF appears, with no signal contribution. These can be recognized as purely noise IMFs precisely because of their small amplitude. Notice also that our criterion to select IMFs has a direct physical interpretation: the square of the amplitude of each IMF gives its energy; hence, small amplitudes mean low energy contribution to the whole signal. In order to keep the total energy of the signal, we accumulate all noise-dominated IMFs and the lowest frequency

\[\text{In fact, this downsampling could be considered as cutting the original ice core probes in pieces of longer extent.}\]
IMF (the trend, as it does not oscillate) in a remaining term. Hence, we produce a reduced IMF decomposition, formed by the physically significant, oscillating IMFs.

Once the reduced set of IMFs for each signal has been obtained, we can calculate the phase difference between two IMFs belonging to two different series, in order to assess their degree of synchronization. Given two time series, \( s^1(t) \) and \( s^2(t) \), their reduced IMF decompositions read:

\[
s^1(t) = \sum_n c_n^1(t); \quad s^2(t) = \sum_m c_m^2(t)
\]

where \( c_n^1(t) \), \( c_m^2(t) \) are the IMFs of each series (we follow the convention of designating the remaining term as the last IMF). In order to evaluate the phase differences of two given IMFs, \( c_n^1(t) \) and \( c_m^2(t) \), we need first to define their local frequencies. This can be done by means of the Hilbert transform of each mode. The Hilbert transform of a function \( f(t) \) (Rudin 1987) creates a complex completion of it, shifting its phase by 90 degrees. Hence, if an IMF \( c_n(t) \) can be expressed in terms of its amplitude \( a_n(t) \) and phase evolution \( \theta_n(t) \) as \( c_n(t) = a_n(t) \cos(\theta_n(t)) \), the Hilbert transform of \( c_n(t) \), denoted by \( Hc_n(t) \), verifies that \( Hc_n(t) = a_n(t) \sin(\theta_n(t)) \).

We can hence use Hilbert transforms to isolate the amplitude \( a_n(t) \) and the phase evolution \( \theta_n(t) \) of each mode \( c_n(t) \), and obtaining from them the phase differences between two given modes. Let us first introduce the set of complex IMFs \( \tilde{c}_n(t) \) and \( \tilde{c}_m(t) \), given by:

\[
\tilde{c}_n^1(t) \equiv c_n^1(t) + Hc_n^1(t) = a_n^1(t) e^{i\theta_n^1(t)}
\]
\[
\tilde{c}_m^2(t) \equiv c_m^2(t) + Hc_m^2(t) = a_m^2(t) e^{i\theta_m^2(t)}
\]

We are interested in the phase velocity, that is, the relative phase speed \( \Delta\theta_{nm}(t) = \theta_{1,n}(t) - \theta_{2,m}(t) \). It can be derived from the following expression:

\[
e^{i \Delta \theta_{nm}(t)} = \frac{\tilde{c}_n^1(t) \cdot (\tilde{c}_m^2(t))^*}{|\tilde{c}_n^1(t)||\tilde{c}_m^2(t)|}.
\]

We define the real projection of the phase shift, \( \delta_{nm}(t) \), as:

\[
\delta_{nm}(t) \equiv \cos(\Delta\theta_{nm}(t)) = \text{Re}(e^{i \Delta \theta_{nm}(t)})
\]

Finally, we will take the variance of \( \delta_{nm}(t) \) (\( \text{var}[\delta_{nm}(t)] \)) as a global indicator of constant phase shift between two given IMFs \( n \) and \( m \). As a matter of fact, \( \text{var}[\delta_{nm}(t)] = 0 \) means constant phase shift, while \( \text{var}[\delta_{nm}(t)] \geq 0.5 \) for modes with very different local frequencies. We fix a compromise threshold of 0.3 for \( \text{var}[\delta_{nm}(t)] \) to detect couples of modes having constant phase shift over a reasonable long time span. The inter-series phase shift of modes is a good estimator of the degree of closeness of the modes and gives a useful information on the oscillating pattern of the couple.
2 Results

We have decomposed the three time series of data into their IMFs. The decomposition separates modes from those with local high frequencies to the ones with low local frequencies. The lowest frequency mode (comprising the empirical lowest frequency mode plus the physically non-significant modes, as explained before) is always an almost DC component, which does not arrive to create a full oscillation so it can be interpreted as the series trend. As discussed before, we expect that low frequency IMFs can be associated to physical forcings common to all the series, while high-frequency IMFs will likely depend on noise and proxy-specific processes. As observed in the figures, local periods are not constant but for each fixed IMF they are contained within different, non-overlapping ranges. The IMFs of GRIP series are shown in the Fig. 2, with oscillation patterns periods of about 100 kyr (last component), 30–50 kyr (6th mode), ∼15 kyr (5th mode), and ∼5 kyr (4th mode). The IMFs of Vostok series are presented in the Fig. 3; there the local periods range from 80–120 kyr observed in the 6th mode, 40–50 kyr in the 5th mode, and 20–30 kyr in the 4th mode. In Epica (Fig. 4) the IMFs exhibit local periods ranging from 80–120 kyr for the 4th mode, 40–50 kyr for the 3rd mode, and 20–30 kyr for the 2nd mode.

The next step has been to calculate the variances of $\delta_{mn}(t)$
The results of the different series decomposition show to compare points of two different series, except in the case of one series according to the other series sampling by using errors, it would be possible to re-define the time sampling well-founded theoretical roots, some problems limit its performance in practice (Huang et al., 1998, 2003). Two are the main limitations. The first one is the difficulty to perform practically all the period, as shown in Fig. 5c-2. The phase ranges between the 120 kyr to 80 kyr period.

3 Discussion

Using an objective method as the Empirical Mode Decomposition (EMD), we have separated different oscillation patterns composing the selected three paleoclimatic time series. The EMD method allows to relate and to compare different time series, despite precise timings, just by considering their common characteristics. A word of warning on EMD should be introduced here. Although the EMD method has well-founded theoretical roots, some problems limit its performance in practice (Huang et al., 1998, 2003). Two are the main limitations. The first one is the difficulty to perform...
a clean separation in IMFs when their local frequencies are too close. This partially happens in some of our IMFs and for that reason the correlations between individual IMFs are a bit lowered, as a fraction of the IMF energy of a given mode is sometimes attributed to another close IMF. The second problem with EMD has to do with domain boundaries of the time series: the two endings (beginning and end of the time span) disturb the local frequency for around half a cycle, and this effect also lowers the mutual correlations, specially when two series with very different time sampling are compared. Recently, these two difficulties have been discussed and different ways to solve them have been proposed (Huang et al., 2003). In spite of these problems, the method has proved to be rather robust and so it is guaranteed that, at least for the longer features, EMD will produce a decomposition reasonable enough to give an interpretation of the results. The main advantage of using EMD lies in the fact that the decomposition is performed in the time domain instead of the frequency domain typical of Fourier methods. This time domain decomposition allows to obtain an instantaneous frequency at each time instant and then to locally compare the instant frequencies of two points from two different time series. Those instantaneous frequencies allow for instance, to identify the starting and finishing instants of relevant events. Frequency domain decomposition methods derived from Fourier analysis, on the contrary, do not allow to associate a characteristic frequency at each point in the time-series and consequently to compare points of two different series, except in the case where the two series are stationary.

The results of the different series decomposition show common, regular oscillation patterns (for medium and low frequencies). Correlating series by pairs we have observed the presence of two common oscillation patterns (for Grip, Vostok and Epica), one with local periods in the range from 80 to 120 ky, and other ranging from 30 to 50 ky. This coincidences could be interpreted as the effect of a non-local common forcing mechanism, but the weak time dependence of the local phase (as can be seen in the coupled modes of three series) and the growing period of oscillation with the time, seems to indicate that the effect on the temperature proxies is strongly non-linear. This seems to suggest that although a periodic non-local forcing mechanism as Milankovitch forcing mechanism (Milankovitch, 1941) can act over well-defined periods, the system response may show non-linear relations, like power-law responses (Huybers and Curry, 2006). Between all the non-linear coupling mechanisms that can produce such patterns, the stochastic resonance is cited in the literature as a plausible phenomena that can trigger abrupt changes or climate oscillations (Gammaitoni et al., 1998; P. Vélez-Belchí et al., 2001; Alley et al., 2001; Ganopolski and Rahmstorf, 2002; Solé et al., 2007). It has also been argued that a possible cause mechanism that forces the change stadial-interstadial can be found in the change of the Atlantic thermohaline circulation (Clark et al., 2002; Rahmstorf, 2002; Stocker, 2002; Alley et al., 2003). Other hypotheses for that change in such a non-constant phase oscillation can be found in a tidal forcing mechanism (Keeling and Whorf, 1997, 2000).

We have seen that the series of temperature proxies studied here have their own particularities and different origins but are subjected to the same climatic forcings, and consequently their dynamic behavior present common aspects. The use of EMD allows us to compare different series, so giving a powerful method to extract the common behavior of the series. Using this method we can extract precise information about local mechanisms in different scales and to decide if the supposed driving forcings are the main influencing mechanism in the time-scales considered or the system has a more complex behavior in such time periods. This can be a first step for to study from a objective point of view the common patterns of given paleoclimatic time series, but more detailed analysis, as well as improvements in the EMD methodology, should be undertaken in the future in order to give more strength to the hypotheses presented here. The introduction of additional temperature proxies would also help to extract more meaningful intrinsic modes.

Another different branch of research concerns the adjustment of time-schedules: if two IMFs are suspected to be coincident but they show some phase difference due to timing errors, it would be possible to re-define the time sampling of one series according to the other series sampling by using an invertible function \( t \rightarrow f(t) \) so that phase shifts vanish, i.e., if we trust the time schedule for \( \delta_{nm,f}(t) \) we look for \( f(t) \) such that \( c_{m,f}^2(t) = c_{nm,f}^2(f(t)) \) verifies \( \delta_{nm,f}(t) = 0 \). The ability of IMFs to separate common forcings from particular relaxation cycles will enable to refer the time labeling of the different series to a fixed reference, enhancing the degree of confidence about event synchronization. Further research should be entertained in this direction.

Acknowledgements. AT is supported by Ramon y Cajal contract from the Spanish Ministry of Education. JELL acknowledges partial financial support of contract REN2003-00185.

Edited by: L. Ferraris
Reviewed by: two referees

References


www.nat-hazards-earth-syst-sci.net/7/299/2007/


Cruz, F., Burns, S., karmann, I., Sharp, W., Vuille, M., Cardoso, A., Ferrari, J., Dias, P. S., and Jr, O.: Insolation-driven changes in atmospheric circulation over the past 116,000 years in subtropical Brazil, Nature, 434, 63–66, 2005.


Milankovitch, M.: Canon of insolation and ice-age problem, Royal Serb Acad, 1941.


