Empirical mode decomposition and correlation properties of long daily ozone records

Imre M. Jánosi*
Department of Physics of Complex Systems, Eötvös University, P.O.Box 32, H-1518 Budapest, Hungary

Rolf Müller†
ICG-I, Research Centre Jülich, 52425 Jülich, Germany

(Received 16 November 2004; revised manuscript received 17 March 2005; published 27 May 2005)

Correlations for daily data of total ozone column are investigated by detrended fluctuation analysis (DFA). The removal of annual periodicity does not result in a background-free signal for the tropical station Mauna Loa. In order to identify the remaining quasiperiodic constituent, the relatively new method of empirical mode decomposition (EMD) is tested. We found that the so-called intrinsic mode functions do not represent real signal components of the ozone time series, their amplitude modulation is very sensitive to local changes such as random data removal or smoothing. Tests on synthetic data further corroborate the limitations of decomposing quasiperiodic signals from noise with EMD. Nevertheless the EMD algorithm helps to identify dominating frequencies in the time series, which allows to separate fluctuations from the remaining background. We demonstrate that DFA analysis for the cleaned Mauna Loa record yields scaling comparable to a mid-latitude station.

DOI: 10.1103/PhysRevE.71.056126 PACS number(s): 02.50.−r, 05.40.−a, 89.75.Da, 94.10.Fa

I. INTRODUCTION

Empirical mode decomposition (EMD) is a new and promising method to analyze nonlinear and nonstationary signals [1]. It has been already proven remarkably effective for engineering design [2], chaotic systems [3], geophysical signal processing [4], meteorological data sets [5], or medical research [6]. The technique is essentially defined by an algorithm for adaptively representing signals as sums of zero-mean amplitude- and frequency-modulated components called intrinsic mode functions (IMF). Adding all the IMFs (more than 8–10 are rarely needed) together with the residual slow trend reconstructs the original signal without information loss or distortion. In this respect, the method is much more "economic" than the traditional Fourier or wavelet decompositions. However, as we demonstrate here, the interpretation of IMFs is not similarly transparent.

In this work, the EMD procedure is pieced together with detrended fluctuation analysis (DFA), which is another emerging tool to handle nonstationary time series [7]. Several theoretical studies mostly on synthetic data [8–13] revealed that DFA results for signals with different correlation properties and slowly changing backgrounds can be fully explained by the variance superposition principle [10,12]. An important issue in the method is the removal of apparent short range regularities from the data, such as dominant seasonal periodicities in atmospheric variables. This can be especially problematic when this component is not strictly periodic, like in the case of North-Atlantic Oscillations, El Nino Southern Oscillations, or the Quasi-Biennial Oscillations [14].

Here we analyze the correlation properties of daily total ozone data for two measuring stations: Arosa (Switzerland, 46.8 °N, 9.7 °E) and Mauna Loa (Hawaii, 19.5 °N, 155.6 °W). The removal of annual periodicity does not eventuate a background-free signal for the Mauna Loa record, therefore we attempt to deploy the EMD method. We show that the resulting IMFs are not existing seasonal components, the most useful information is in their dominant frequencies. Nevertheless this information yields to a successful detrending and to a detection of possible long-range correlations in the fluctuations.

The Dobson spectrophotometer has been developed as the first instrument to determine the total amount of ozone in a column from the surface to the edge of the atmosphere, referred to as "total ozone." 1 Dobson unit (DU) is defined to be a 0.01 mm layer of pure ozone at standard temperature and pressure, typical atmospheric values are 200–400 DU. Regular measurements begun in the mid-1920s [15]; here we analyze daily total ozone measurements recorded from 1926 in Arosa, and from 1963 in Mauna Loa.

An essential point is that the total ozone records are not continuous, there is no data for approximately 36% of the days. This is inevitable at the given spectroscopic method (direct sun observations in UV range), because clouds hinder precise measurements.

II. DFA ANALYSIS

Correlations are important in the dynamics and chemistry of the atmosphere, and serve as testbed for numerical models. As a first step of DFA analysis for atmospheric records, the annual cycle is removed from the raw data O_i by computing the ozone anomaly series $x_i = O_i - \langle O_i \rangle_d$, where $i = 1, \ldots, N$, and $\langle \cdot \rangle_d$ denotes the long-time average for the given calendar day. Next, the anomaly series is integrated to obtain the so-called profile $y_j = \sum_{i=1}^{j} x_i$. The profile is divided into time segments of equal length n, and the local trend is fitted by a polynomial of order p in each segment. The fluc-

*Electronic address: janosi@lecso.elte.hu
†Electronic address: ro.mueller@fz-juelich.de

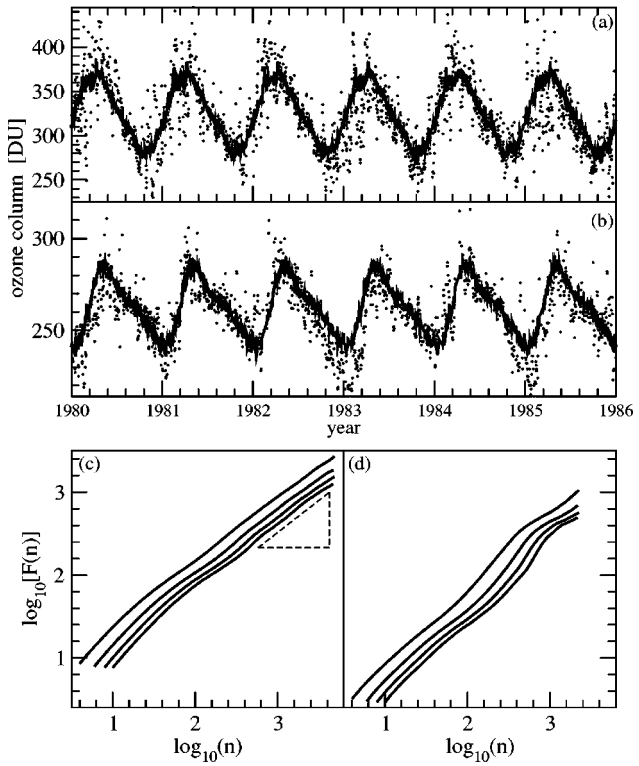

FIG. 1. Part of the daily total ozone data for (a) Arosa and (b) Mauna Loa, solid lines indicate the long-time daily averages. (c) and (d): DFA1–DFA4 curves for Arosa and Mauna Loa in the same scales. An apparent vertical shift is the consequence of lower fluctuation amplitudes for Mauna Loa. The dashed line has a slope of 0.75.

tuation strength for a given segment is determined as the root mean squared deviation from the local trend, and an average $F_p(n)$ is formed over the different segments. A power-law relationship between $F_p(n)$ and n indicates scaling with an exponent δ (DFA p exponent):

$$F_p(n) \sim n^{\delta}. \tag{1}$$

The numerical value for δ is established when the asymptotic slope of the DFA curves (in log-log scales) does not change by increasing p. Long-memory (persistent) processes are characterized by DFA exponents $\delta > 0.5$, uncorrelated time series (e.g., white noise) obey $\delta = 0.5$, antipersistent signals have $\delta < 0.5$.

Note that missing intervals do not distort DFA results when the data are positively correlated ($\delta > 0.5$), as was pointed out by Chen *et al.* [12]. They randomly cut out up to 50% of synthetic data of known scaling exponents and stitched together the remaining parts. We used the same procedure for the discontinuous ozone records.

Figures 1(a) and 1(b) show part of the raw data together with the empirical annual cycles obtained by averaging for each calendar day. The overall course for one year is very similar for both stations, apart from the different magnitudes (the ozone level and its variance are systematically higher in Arosa by 25%–30%). DFA curves for both records after removing the annual periodicities are plotted in Figs. 1(c) and 1(d). Scaling can be established for Arosa with an exponent value $\delta \approx 0.75$, however the apparent kinks impede fitting for the Mauna Loa data. Similarly breaking slopes indicate the presence of an oscillating background [10,16]. However, neither visual inspection nor standard Fourier analysis did reveal a pure monofrequency component. In order to unfold further details and attempt to separate clean fluctuations, EMD analysis was performed for the records.

III. EMD ANALYSIS

The EMD algorithm (also called "sifting") breaks down a signal into its component IMFs obeying two properties, (i) an IMF has only one extrema between zero crossings (that is the number of local minima and maxima differs at most by 1), and (ii) an IMF has a mean value of zero. For a given time series $O(t)$, let $m_1(t)$ be the mean of its upper and lower envelopes as determined from a cubic-spline interpolation of local maxima and minima. In the second sifting round, the first residual $r_1(t) = O(t) - m_1(t)$ is treated as the data, and $m_{11}(t)$ will be the mean of the two envelopes for r_1, then $r_{11}(t) = r_1(t) - m_{11}(t)$. This procedure is repeated j times until the mean $m_{1j}(t)$ is sufficiently close to zero. At this step $r_{1j}(t)$ is designated as the first IMF (IMF1) containing the shortest period component of the signal. The same sifting procedure is continued with the difference $O(t) - \mathrm{IMF1}$, until the remaining signal is almost zero everywhere. (For the details of algorithmic implementation see the review [1].)

The result for Mauna Loa is shown in Fig. 2 (the last component has a vanishing amplitude). Note that the IMF of the largest magnitude (IMF6, third from the bottom in Fig. 2) has a characteristic period of approximately 1 year. It is apparent, however, that the variability of its envelope is much stronger than that of the original signal (Fig. 2, top). This observation suggests that IMF6 does not represent a real, existing component function.

Since the EMD algorithm provides a fully local decomposition, holes in a time series is expected to influence high frequency IMFs. It is not obvious that missing segments will affect the low frequency IMFs as well. Figure 3 illustrates that an additional random removal of 1%–2% data indeed results in drastic changes in IMFs. Other manipulations altering locally the data, such as smoothing by three-point or five-point running averages, result in similarly strong effect on the amplitude modulation.

This sensitivity to local details can be exploited to identify periodic or quasiperiodic background signals. Figure 4(d) shows the ozone anomaly time series for Mauna Loa, where the empirical annual periodicity [Fig. 3(a)] is already subtracted from the raw data (Fig. 2, top frame). A straight DFA analysis of the very time series gave the curves in Fig. 1(d) indicating residual background oscillations. The sifting procedure of this signal (not shown here in detail) results in 11 significant IMFs, where the variance of the annual component (IMF7 in this case) is strongly reduced, as expected. Figure 4(a) shows the Fourier amplitudes of IMF8, the remaining dominant component. After randomly removing an additional 1% or 2% of data from the series, the component IMF8 suffers from strong changes reflected also in the Fou-

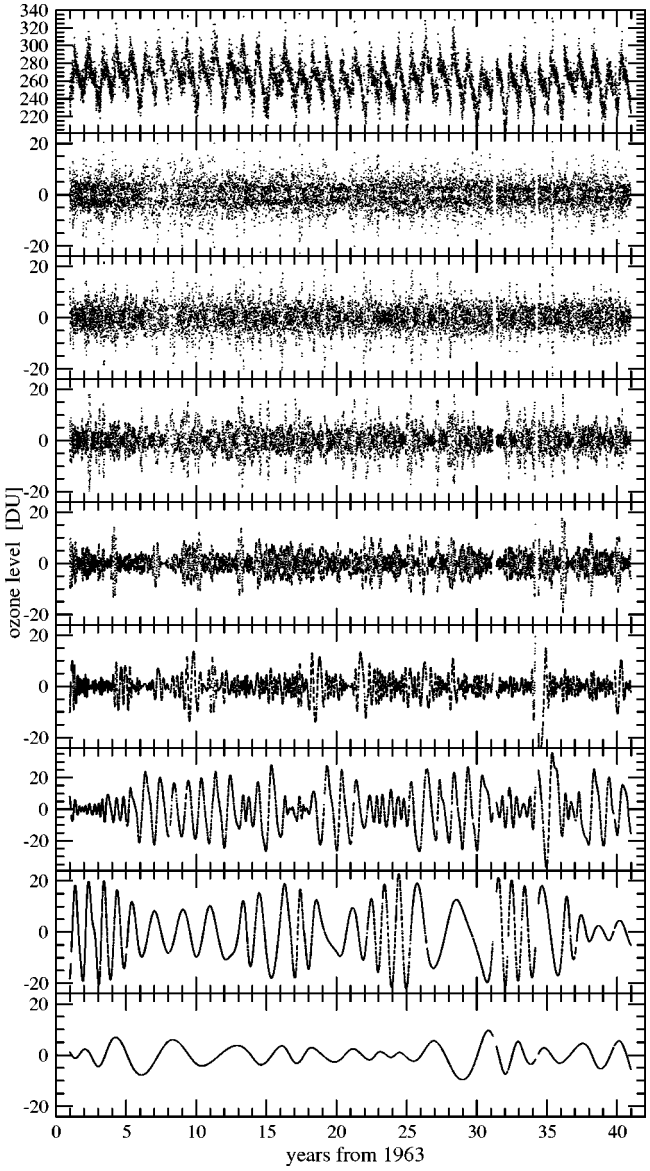

FIG. 2. From top to bottom, daily ozone column in Dobson units measured at the Mauna Loa station, and the first eight IMFs resulted in the EMD sifting process.

rier amplitudes shown in Figs. 4(b) and 4(c). Nevertheless the period of the two main peaks are not changing, they are 627 and 879 days, respectively.

This observation suggests that a quasiperiodic background signal with two dominating frequencies is present in the data. We have found that the Ansatz

$$x(t) = a_0 + a_1 t + a_2 \cos\left(\frac{2\pi}{1.7166} t + a_3\right) + a_4 \cos\left(\frac{2\pi}{2.4065} t + a_5\right) \quad (2)$$

gives a satisfying estimate with the fitted parameters $a_0 = 5.28$, $a_1 = -0.26$, $a_2 = -4.09$, $a_3 = 1.11$, $a_4 = 4.52$, and $a_5 = 1.19$ (note that the time is measured here in units of years). A detailed analysis of the slow quasiperiodic component is

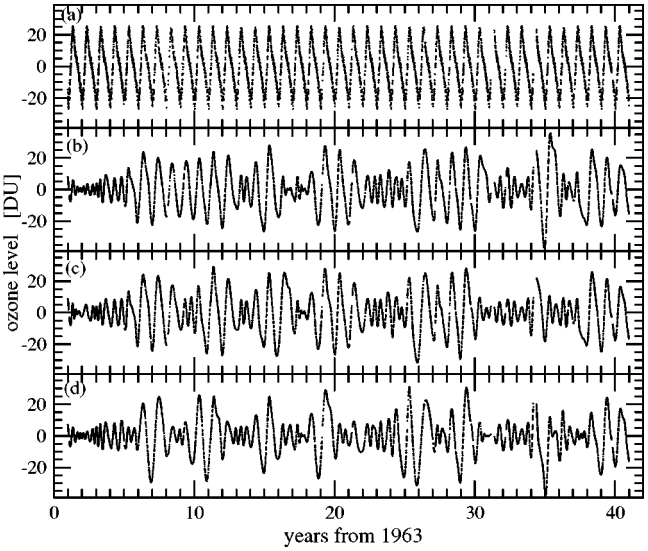

FIG. 3. (a) Empirical annual cycles of the Mauna Loa record (see Fig. 2 top frame). (b) IMF6 for the full record (characteristic frequency is 1 year). IMF6 for the record after a random removal of (c) 1% and (d) 2% of the data (147 and 302 from the original 9310 points covering about 40 years).

beyond the scope of the present report, yet we note that the smeared oscillation of ~30 months period is recognized in many tropical signals (mostly in the stratosphere), and it is known as quasibiennial oscillation (see Ref. [17] and references therein).

The results of DFA analysis for the Mauna Loa ozone anomaly data cleaned by subtracting the background Eq. (2)

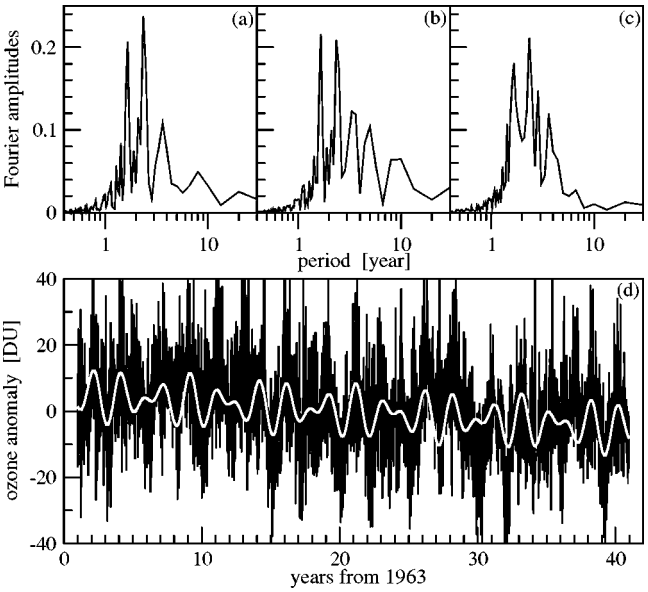

FIG. 4. (a) Fourier amplitudes as a function of periods (note the logarithmic scale) for the IMF8 of the anomaly time series shown below (black line); (b) the same as (a) but 1% of data randomly removed before EMD analysis; (c) the same as (b) but 2% random removal. (d) Ozone anomaly time series for Mauna Loa (black line), and the best quasiperiodic fit obtained by Eq. (2) (thick white line).

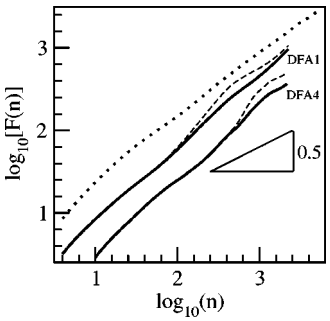

FIG. 5. DFA1 and DFA4 curves for the cleaned ozone anomaly series from Mauna Loa (thick lines), the original results [Fig. 1(d)] are shown as dashed lines. Heavy dotted line is the DFA1 curve for Arosa [Fig. 1(c)].

is shown in Fig. 5. An improvement is apparent, although the DFA curves are not entirely straight on the log-log plot.

IV. DISCUSSION

In order to investigate why the EMD algorithm cannot separate directly a quasiperiodic background component from a time series, we performed several tests on synthetic data sets. The basic signal is given by Eq. (2) without the slow linear shift, that is $a_0=0$, $a_1=0$, the other coefficients retain the fitted values. Noisy synthetic data sets were constructed by amplitude matching, the empirical standard deviation is $\sigma \approx 12.0$ DU for Mauna Loa. (The histogram is almost perfect Gaussian.) Different autocorrelations obeying power-law were modeled with the algorithm developed by Pang, Yu, and Halpin-Healy [18], see also Ref. [16].

The EMD algorithm has been studied in detail for pure noise with and without long-range correlations [19,20]. It was shown that the procedure works as a dyadic filter bank, that is the mean period of a given IMF is approximately 2 times of the previous one. This behavior was reproduced for signals with slow background oscillations and missing segments as well, at least for the first eight mode functions. The aspect here is an apparent breaking at mode number 9. Simple visual inspection is enough to recognize that the remaining mode functions belong to the slow background component.

The effect of noise and missing segments is exactly what we expect by considering the fully local nature of EMD sifting. The quasiperiodic signal is "smeared" from four to five IMFs, and the amplitude modulation is very different for various realizations, determined by particular details. The superposition of the slow modes cannot reproduce the original component without errors [see Fig. 6(d), top curves]. Consequently, the missing features "contaminate" other IMFs, depending on local details, again.

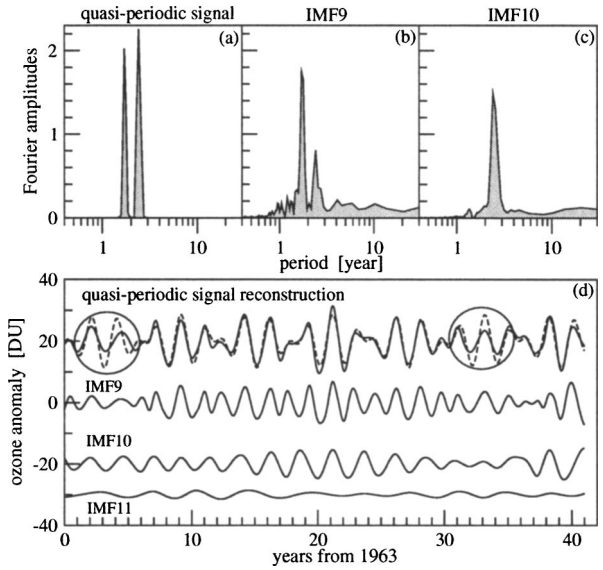

FIG. 6. (a) Fourier spectrum of the quasiperiodic signal given by Eq. (2) with $a_0=0$ and $a_1=0$, also shown with dashed line in (d), top. (b) and (c), Fourier spectra for the ninth and tenth IMF [shown in (d)] of a synthetic data set with amplitude matched correlated noise ($\delta=0.81$), and randomly removed segments of 13.1% from the total length of 14 968. (d) Reconstructed slow background from the superposition of modes 9–12 (heavy line), and three IMFs for the noisy set. For the sake of visualization, the curves are vertically shifted to each other. Ellipses indicate segments of extremely poor reproduction.

The good news is that the presence of fundamental incommensurate frequencies is very robust in the decomposition of all synthetic data with different amount of missing segments (0–42%) and level of noise correlation ($0.5 \leqslant \delta \leqslant 0.85$) we tested. This finding supports our original observation that the Fourier spectra of intrinsic mode functions can provide very useful information about possible physical decomposition of complex measured signals.

ACKNOWLEDGMENTS

The authors thank J. Staehelin for discussions. The Mauna Loa data were obtained from www.cmdl.noaa.gov/ozwv/dobson/select.html and the Arosa data from www.iac.ethz.ch/en/research/chemie/tpeter/totozon.html. This work was supported by a German-Hungarian Academic Collaboration Grant (TéT D-30/02), and the Hungarian Science Foundation (OTKA) under Grant No. TS044839.

[1] N. E. Huang *et al.*, Proc. R. Soc. London, Ser. A **454**, 903 (1998).

[2] J. J. Miau *et al.*, AIAA J. **42**, 1388 (2004); Y. K. Wen and P. Gu, J. Eng. Mech. **130**, 942 (2004); J. Chen and Y. L. Xu, Struct. Des. Tall Build. **13**, 145 (2004).

[3] Y. C. Lai and N. Ye, Int. J. Bifurcation Chaos Appl. Sci. Eng. **13**, 1383 (2003); J. Jamsek *et al.*, Phys. Rev. E **68**, 016201 (2003); A. Hutt, A. Daffertshofer, and U. Steinmetz, *ibid.* **68**,

036219 (2003).
[4] P. A. Hwang, N. E. Huang, and D. W. Wang, Appl. Ocean. Res. **25**, 187 (2003); R. R. Zhang *et al.*, Int. J. Non-Linear Mech. **39**, 1501 (2004); J. R. Gemmrich and D. M. Farmer, J. Phys. Oceanogr. **34**, 1067 (2004); A. D. Veltcheva and C. G. Soares, Appl. Ocean. Res. **26**, 1 (2004).
[5] M. D. Alexandrov, Geophys. Res. Lett. **31**, L04118 (2004); D. G. Duffy, J. Atmos. Ocean. Technol. **21**, 599 (2004); H. El-Askary *et al.*, Adv. Space Res. **33**, 338 (2004).
[6] J. C. Echeverria *et al.*, Med. Biol. Eng. Comput. **39**, 471 (2001); D. A. T. Cummings *et al.*, Nature (London) **427**, 344 (2004); E. P. S. Neto *et al.*, Methods Inf. Med. **43**, 60 (2004).
[7] C. K. Peng *et al.*, Phys. Rev. E **49**, 1685 (1994); Chaos **5**, 82 (1995); H. E. Stanley *et al.*, Physica A **200**, 4 (1996).
[8] P. Talkner and R. O. Weber, Phys. Rev. E **62**, 150 (2000).
[9] C. Heneghan and G. McDarby, Phys. Rev. E **62**, 6103 (2000).
[10] K. Hu *et al.*, Phys. Rev. E **64**, 011114 (2001).
[11] J. W. Kantelhardt *et al.*, Physica A **295**, 441 (2001).
[12] Z. Chen *et al.*, Phys. Rev. E **65**, 041107 (2002).
[13] Z. Chen *et al.*, Phys. Rev. E **71**, 011104 (2005).
[14] G. Wang and D. Schimel, Annu. Rev. Environ. Resour. **28**, 1 (2003); M. P. Baldwin *et al.*, J. Geophys. Res., [Atmos.] **39**, 179 (2001).
[15] G. M. B. Dobson, Appl. Opt. **7**, 387 (1968).
[16] A. Király and I. M. Jánosi, Phys. Rev. E **65**, 051102 (2002).
[17] R. P. Kane, Y. Sahai, and C. Casiccia, J. Geophys. Res., [Planets] **103**, 8477 (1998).
[18] N. N. Pang, Y. K. Yu, and T. Halpin-Healy, Phys. Rev. E **52**, 3224 (1995).
[19] P. Flandrin, G. Rilling, and P. Goncalves, IEEE Signal Process. Lett. **11**, 112 (2004).
[20] Z. Wu and N. E. Huang, Proc. R. Soc. London, Ser. A **460**, 1597 (2004).