Nonparametric segmentation of nonstationary time series

S. Camargo,[1] S. M. Duarte Queirós,[2] and C. Anteneodo[1,3]
[1]*Departamento de Física, PUC-Rio, Rio de Janeiro, Brazil*
[2]*Istituto dei Sistemi Complessi–CNR, Roma, Italy*
[3]*National Institute of Science and Technology for Complex Systems, Rio de Janeiro, Brazil*
(Received 29 April 2011; revised manuscript received 9 August 2011; published 12 October 2011)

The nonstationary evolution of observable quantities in complex systems can frequently be described as a juxtaposition of quasistationary spells. Given that standard theoretical and data analysis approaches usually rely on the assumption of stationarity, it is important to detect in real time series intervals holding that property. With that aim, we introduce a segmentation algorithm based on a fully nonparametric approach. We illustrate its applicability through the analysis of real time series presenting diverse degrees of nonstationarity, thus showing that this segmentation procedure generalizes and allows one to uncover features unresolved by previous proposals based on the discrepancy of low order statistical moments only.

DOI: 10.1103/PhysRevE.84.046702 PACS number(s): 05.10.−a, 05.45.Tp, 05.40.−a, 89.75.−k

I. INTRODUCTION

Complex systems are seldom in equilibrium or even in stationary states; however, their evolution can in many cases be thought of as being composed of spells of quasistationarity in which time-varying pseudoparameters can be considered unchanged. Examples of such framework can be found in finance [1], biology [2], physics [3,4], and physiology [4,5], just to mention a few areas. By identifying stationary segments, one can apply standard techniques, e.g., extracting stochastic equations (Kramers-Moyal coefficients) from data [6], overcoming the difficulties of nonstationary treatments [5,7]. As another application, a proper segmentation is important to assess the scenario of mixed statistics [8] based on the idea of local equilibrium, typically applied by considering a fixed characteristic scale (window length). In general, segmentation provides a useful portrait of the local statistical properties for modeling nonstationary systems.

In order to identify such quasistationary patches, algorithms based on standard statistical methodology have been introduced. Explicitly, they lean on moving along the series a pointer to detect the position that maximizes a given quantifier of the statistical discrepancy between the segments on both sides of the pointer. Among others [9], worthy of mention are the algorithms based on the Student's t statistic (used to test the significance of the null hypothesis of equal means) [10,11] or on the Jensen-Shannon divergence in the case of symbolic sequences [12].

Despite the interesting results provided by these methods [12–19], limitations hampering the performance can be found in every one of them. On the one hand, in the statistical moments criteria there is the problem of boiling down the existence of nonstationarity to the change of preestablished local quantities. For instance, even if the time series presents fluctuations in the variance, the t test may give us the indication that the series is stationary. Although it could be improved through the unequal variance t-test statistic or also through an F test, it will still rely on assumptions over the moments and on the validity of the central limit theorem. On the other hand, entropy-based methods are more fitted to symbolic sequences, while information is lost if discretizing a real valued series by means of thresholds. Moreover, the segmentation stopping criteria can be deemed arbitrary. Its proposed improvement by means of the Bayesian information criterion can be disputed as well since such a criterion often favors minimalist modeling [20]. With the aim of surmounting those difficulties, we introduce a *fully* nonparametric segmentation approach by using the Kolmogorov-Smirnov (KS) statistic, D_{KS}, which measures the maximal distance between the cumulative distributions of two samples, as an estimate of the discrepancy between segments. Note that it allows one to test whether two samples come from the same distribution with no need to specify which is the common distribution.

II. KS-SEGMENTATION ALGORITHM

Our algorithm (named KS segmentation) works as follows. Given a segment of a time series, $\{x_i, i_1 \leqslant i \leqslant i_n\}$, a sliding pointer, at $i = i_p$, is moved in order to compare the two fragments $S_L \equiv \{x_{i_1}, \ldots, x_{i_p}\}$ and $S_R \equiv \{x_{i_p+1}, \ldots, x_{i_n}\}$. The position i_p of the pointer is moved so that the sizes of the two segments ($n_L = i_p - i_1 + 1$ and $n_R = i_n - i_p$) are at least unitary. Then, one selects the position $i_{\max}$ that maximizes the KS statistic $D \equiv D_{\mathrm{KS}}(1/n_L + 1/n_R)^{-1/2}$, between the two patches S_L and S_R.

Once found, the position $i_{\max}$ of the maximal distance D, $D^{\max}$, one checks the statistical significance (at a chosen significance level $\alpha = 1 - P_0$) of a potentially relevant cut at $i_{\max}$ by comparison with the result that would be obtained was the sequence random [10]. The potential cut ticks the first stage if $D^{\max}$ exceeds its critical value, $D^{\max}_{\mathrm{crit}}(n)$, for the selected significance level (see Fig. 7 and further technical details in the Appendix). Before final acceptance of the cut, one can still require a minimal size (number of points) ℓ_0, namely, $i_{\max} - i_1 + 1$, $i_n - i_{\max} \geqslant \ell_0$. The procedure is then recursively applied starting from the full series $\{x_i, 1 \leqslant i \leqslant N\}$, where N is the total number of data points, until no segmentable patches are left. The search for $D^{\max}$ within a given segment $\{i_1, \ldots, i_n\}$ during the iterations, as well as in the determination of the critical curves, is performed for $i_1 \leqslant i_p \leqslant i_n - 1$. The outcome of this segmentation procedure when applied to the paradigmatic nonstationary time series is hereafter presented.

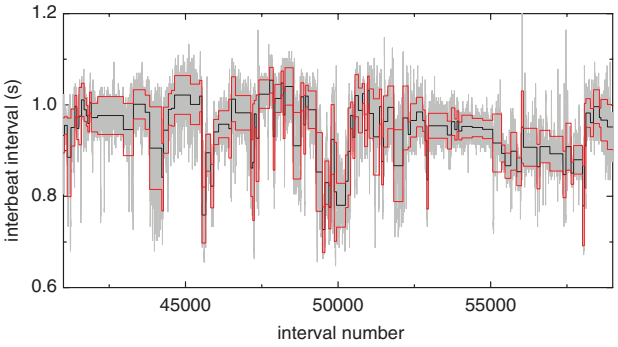

FIG. 1. (Color online) Fragment of a heart rate time series for a healthy individual (n5nn.txt) (light gray lines). The mean plus or minus the standard deviation of the segments resulting from the KS segmentation, with $\ell_0 = 50$ and $P_0 = 0.95$, are displayed.

III. APPLICATIONS

We first survey the segmentation of heart-rate (RR) series, which motivated the introduction of the segmentation algorithm based on the discrepancy of the means (mean-based algorithm) [10] and that have also been suggested as a common focus to solve the controversy over the potential chaoticity of normal heart rate [21]. Namely, we revisit the study of interbeat time series from healthy individuals (NOR) and patients with congestive heart failure (CHF) [22]. Time series [tagged n1nn-n5nn and c1nn-c5nn for NOR and CHF, respectively] are about 24 h long and had their outliers removed. The segmentation outcome is depicted in Fig. 1.

We computed the first moments for each resulting segment. The variance is deemed not constant throughout segments, but it is dispersed over more than one decade, as illustrated in Fig. 2, for all segment sizes. The local variance is larger for the healthy subjects. Therefore, equal variance cannot be assumed as in previous analysis of heart-rate series [10]. Furthermore, our finding implies that if one keeps such a simpler analysis, at least the effective degrees of freedom in the t test should be obtained by the Welch-Satterthwaite equation. It is worth noting that despite the tendency for the variance to be larger in

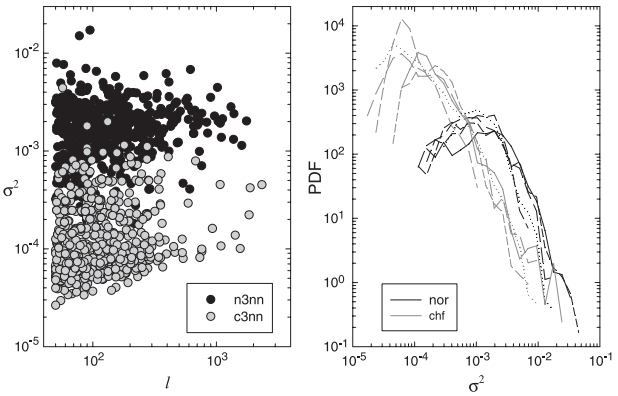

FIG. 2. Local variance vs segment length for a representative individual of each group (left). Probability distribution function (PDF) of the variance for all individuals of the NOR (black lines) and CHF (gray lines) groups (right panel).

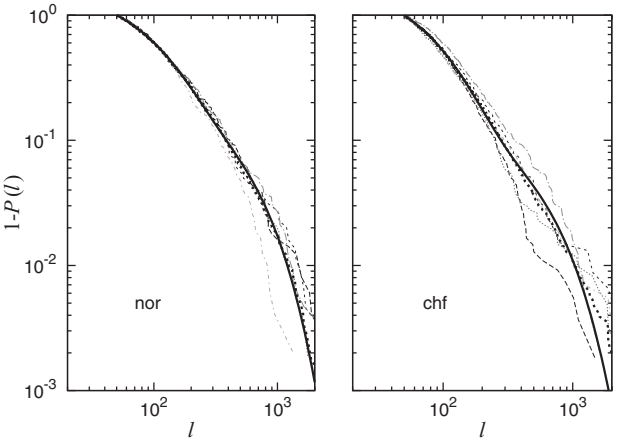

FIG. 3. Cumulative distribution of segment sizes for NOR (left panel) and CHF (right panel) individuals. In each panel, the thin lines correspond to each individual of the group (five samples of 24 h data), the dark dotted line to the entire group, and the dark full line to fits to $Ae^{-(l-\ell_0)/L_1} + (1-A)e^{-(l-\ell_0)/L_2}$ with amplitude and characteristic lengths $(A, L_1, L_2) = (0.78, 78, 372)$ and $(0.86, 64, 373)$, respectively.

the NOR group, the quotient for consecutive segments is very similarly distributed in both groups, with a slow power-law decay (with exponent close to -3) (not shown).

The complementary cumulative distributions of segment sizes for NOR and CHF individuals are displayed in Fig. 3. For both groups, the plots can be described by a double exponential $Ae^{-(l-\ell_0)/L_1} + (1-A)e^{-(l-\ell_0)/L_2}$ with characteristic lengths $L_1 \simeq 70$ and $L_2 \simeq 370$. Hence there is no indication of a scale-free behavior, as suggested by previous segmentation analysis through the mean-based algorithm [10].

Our next example concerns the scenario of mixed statistics that has been applied to the study of fluid turbulence [3,23]. Besides turbulence, its relevance is highlighted by the fact that several models for finance have been inspired by this physical problem [24]. Succinctly, the mixed approach corresponds to a conjecture where one has a classical Boltzmann-Gibbs statistics, conditioned to given temperature ($T \sim \beta^{-1} \sim \sigma^2$), which signals the existence of local equilibrium, that is associated with certain distribution $P(\beta)$. Nonetheless, up to now, the approaches to the problem have considered the existence of a *single scale* of local equilibrium, which can be seen as a first step for an outright description [24,25]. Endowed with our segmentation algorithm, we are in a position to evaluate the distribution of local time scales, τ, of the β factor and verify the local equilibrium assumption in this type of system. As an example, we consider a series of wind velocities (one month of measurements at a 30 s acquisition interval) [26]. We reduced the strong daily periodicity by scaling the data by the average at each time of the day. Then velocity $\vec{v}$ is dimensionless.

Accordingly, in a segment within which local equilibrium holds, the velocity distribution is defined by $p(\vec{v}|\beta) = \frac{\beta}{2\pi}\exp[-\beta\frac{|\vec{v}|^2}{2}]$, from which the distribution of the speed, $v \equiv |\vec{v}|$, is $p(v|\beta) = \beta v \exp[-\beta\frac{v^2}{2}]$. Along these lines, we apply our segmentation procedure pitching at detecting the time intervals where the local equilibrium approximation is

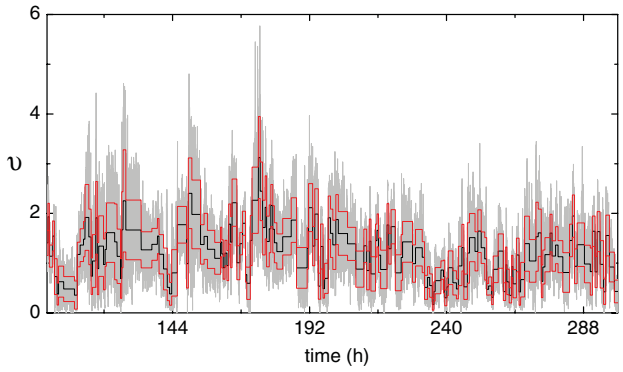

FIG. 4. (Color online) Time series of $v = |\vec{v}|$ (dimensionless). The mean plus or minus the standard deviation of the segments resulting from the KS segmentation are displayed (for $\ell_0 = 50$, $P_0 = 0.95$).

valid. The result of the segmentation is depicted in Fig. 4. The complementary cumulative distribution of segments decays more slowly than exponentially (plausibly a stretched or double exponential, the latter with characteristic times of 32 and 93 min). The distribution of segment lengths has mean (standard deviation) approximately equal to 129 (91) points [corresponding to ∼64 (∼45) min.]. One observes in Fig. 5 that the two-dimensional (2D) Maxwell distribution fails in describing the distribution of velocities, because the local variance is dispersed (inset of Fig. 5). Then we considered the mixing $p(v) = \int_0^\infty d\sigma^2\, p(v|\sigma^2)p(\sigma^2)$, where $p(v|\sigma^2)$ is the Maxwell distribution defined above, substituting $\beta = (4-\pi)/(2\sigma^2)$, given that the (conditioned) raw moments are $\langle v^k\rangle_\beta = (2/\beta)^{k/2}\Gamma[1+k/2]$. One observes that the mixed

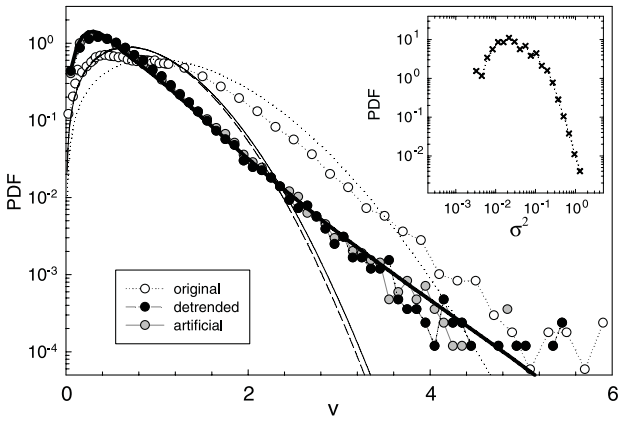

FIG. 5. Distribution of velocity v (circles). Data were detrended by subtracting the excess average with respect to the one given by the Maxwell distribution with the same local variance (black circles). Distribution of an artificial series with the same segments as the real one, with values of v independently drawn from 2D-Maxwell distributions with the local variance of the real series (gray circles). Thin lines correspond to the respective 2D-Maxwell distributions, with the variance of the whole series, and the full line to the mixing (numerically summed up) of the 2D Maxwell with the distribution of the local variance obtained from the segmentation process, shown in the inset.

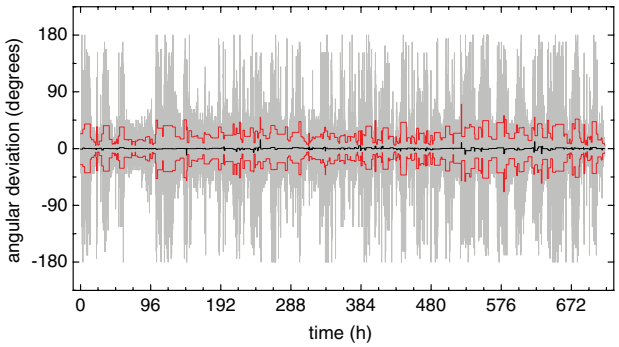

FIG. 6. (Color online) Segmentation of angular deviation at 30 s lag. The mean plus or minus the standard deviation of the segments resulting from the KS segmentation are displayed.

distribution is in good accord with the data distribution, once local trends are removed. The mixed distribution also agrees with the histogram built from artificial series obtained by juxtaposition of sequences of Maxwellian random numbers, with the same length and local variance as the real ones.

Afterward, we computed the local variance that is proportional to the inverse β factor. Its distribution, presented in the inset of Fig. 5, is responsible for the deviation of $p(v)$ from the 2D-Maxwell distribution (main frame of Fig. 5).

In addition to the speed, v, we also looked at angle variations (at 30 s lag). While a more quantitative analysis on this matter is to be addressed in future work, here we would like to call attention to the fact that, even if the average value is constantly close to zero, changes in the local variance are detected by the present method, as depicted in Fig. 6, while no segmentation occurs with the procedure based on the discrepancy of the means.

IV. FINAL REMARKS

In this paper we have presented a segmentation method which aims at coping with nonstationary signals from widespread physical and nonphysical systems where the local-equilibrium or local-stationarity hypotheses hold. Our method, which is based on the Kolmogorov-Smirnov test, improves previous proposals as soon as it is nonparametric and thus independent of any preassumed order of fluctuation between the stationary segments, resulting in a more flexible and effectual algorithm.

Concerning algorithmic complexity, our algorithm is more efficient than methods based on matrix diagonalization, such as principal component analysis. In comparison with moment-based segmentation proposals, our algorithm requires sorting segments of length n, which increases the complexity in a factor of order $\ln n$, which does not represent a significant larger computational cost, despite the enhanced ability.

The applicability of our proposal was then tested with two-poles-apart signals, heart-rate intervals, and wind velocities, with significant results in each case. In the first case, the nonstationarity portrait is altered with respect to that of previous analysis based on the discrepancy of the means, as soon as the local variance cannot be assumed constant. In the second, the procedure is shown to be useful to detect meaningful windows to compute local statistical quantities.

In general, proper segmentation is helpful in several problems where local stationarity applies.

ACKNOWLEDGMENTS

We are grateful to Pedro Bernaola-Galván for useful exchange of correspondence about previous segmentation proposals. Also acknowledged are Salvo Rizzo and Stefano Ruffo for aiding us in obtaining wind data. We acknowledge Brazilian agencies Faperj (Foundation for Research Support, State of Rio de Janeiro) and CNPq (National Council for Scientific and Technological Development), and the European Commission through the Marie Curie Actions FP7-PEOPLE-2009-IEF (Contract No. 250589) for financial support.

APPENDIX

1. Statistical significance criterion

To obtain the critical curves for significance testing, we determined $D^{\max}$ numerically for a large number ($>10^4$) of sequences of n independent and identically distributed (i.i.d.) Gaussian numbers and built its cumulative distribution. From the cumulative distribution, we obtained the critical values of $D^{\max}(n)$, $D^{\max}_{\rm crit}(n)$, for each given significance level $\alpha = 1 - P_0$. The resulting critical curves are shown in Fig. 7. For the significance tests applied throughout the segmentation procedure, we used the effective form of the critical curves given by the heuristic simple expression

$$D^{\max}_{\rm crit}(n) = a(\ln n - b)^c, \tag{A1}$$

with (a,b,c) equal to (1.41,1.74,0.15), (1.52,1.8,0.14), and (1.72,1.86,0.13) for $P_0 = 0.90$, 0.95, and 0.99, respectively. These curves are more restrictive than the critical ones for the standard two-sample KS test, that for large n tend to 1.22, 1.36, and 1.63, for the above values of P_0, respectively. The employ of the usual KS test would lead to oversegmentation. This is because the two samples under comparison are not any two samples but they arise from a potential cut at maximal D, hence the more restrictive criterium defined above must be used.

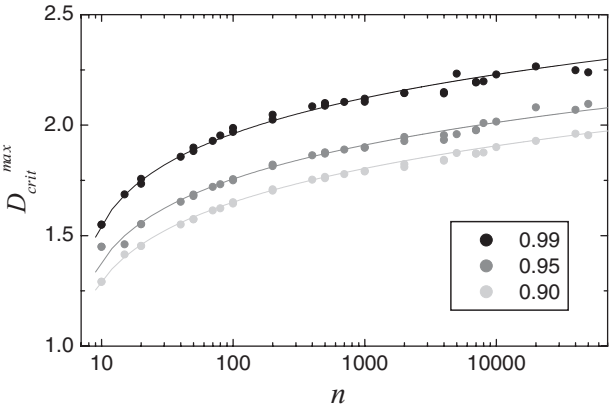

FIG. 7. Critical values of $D^{\max}$ as a function of the sequence length n for series of Gaussian i.i.d. random numbers, at different significance levels P_0 indicated on the figure. Solid lines are fits to the data, used as phenomenological formulas for significance checking.

We noticed that, along a random series, the position $i_{\max}$ for which D is maximal is not uniform but presents a U-shaped distribution. This could set forth a bias in the cutting performance propping up an increase in the number of short segments. Let us mention that in the mean-based algorithm an alike U shape is also present. This issue arises because of the fact that we are considering effective critical curves, only as a function of $n = n_L + n_R$, while the intrinsic ones actually depend both on n_R and n_L. On the other hand, these intrinsic curves would be hardly assessed, then to contour this issue, we tested a redefinition of the standard KS distance by considering $D \equiv D_{\rm KS}(1/n_R + 1/n_L)^{-\gamma}$, with effective γ and observed that the flattest (more uniform) distribution of $i_{\max}$, for any size n, occurs for $\gamma \simeq 0.64$. We compared the implementation of the algorithm for both values of γ (0.5 and 0.64), finding no significant differences in the segmentation portrait both for real and artificial series, as soon as ℓ_0 is not too small ($\ell_0 \gtrsim 10$). Then we kept the standard definition of D, i.e., with $\gamma = 0.5$.

2. Testing artificial series

To check the performance of the algorithm, we analyzed artificial series $\{y_i, 1 \leqslant i \leqslant N\}$, formed by segments of m Gaussian numbers. We set unitary jumps in the means of consecutive segments ($\Delta\bar{y} = 1$) and alternating standard deviation (square root of the variance) σ_1,σ_2, as illustrated in Fig. 8. We varied each standard deviation from 1/10 to 10, then embracing a wide range of values relative to the jump size.

Diagrams of the segmentation results in the plane σ_1,σ_2 are shown in Figs. 9 and 10, for $\ell_0 = 10$ and 50, respectively, at level $P_0 = 0.95$. For each sequence, the percent relative number of cuts with respect to the actual one is displayed. The outcomes of the KS algorithm are shown in the right-hand-side panels and those of the mean-based algorithm are also presented (left-hand-side panels) for comparison. Time series belonging to the middle gray regions are correctly (100%)

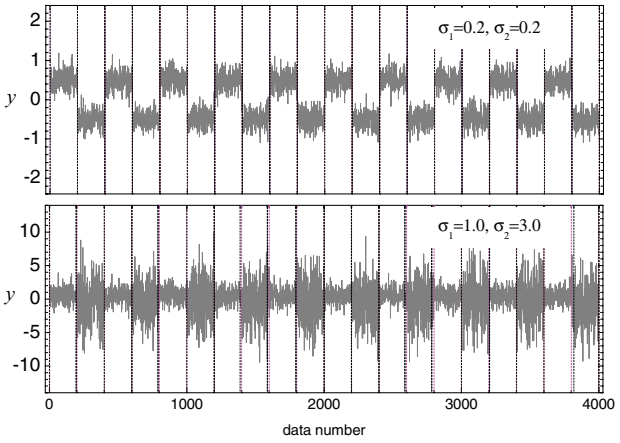

FIG. 8. (Color online) Artificial series (light gray lines) formed by segments of $m = 200$ Gaussian numbers with alternating means $+1/2$, $-1/2$, and standard deviation σ_1,σ_2 (values indicated on each panel). Segmentation was performed by means of the KS algorithm at level $P_0 = 0.95$. The vertical dotted lines indicate the exact (magenta, gray) and calculated (black) borders.

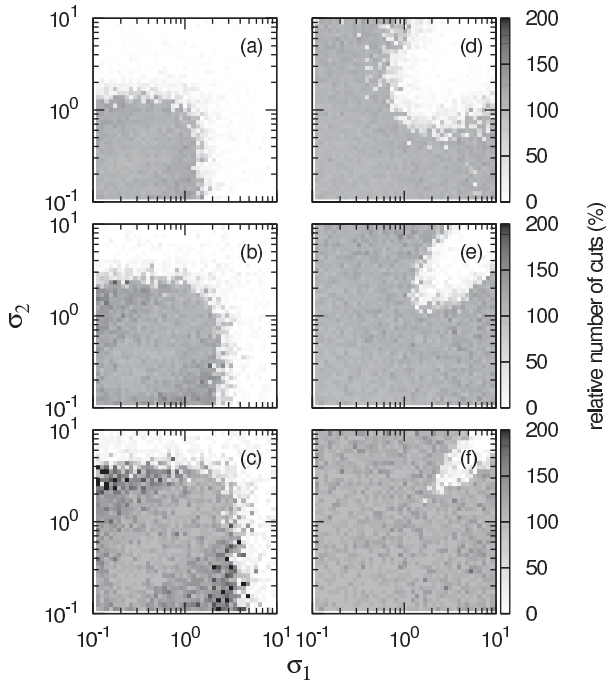

FIG. 9. Segmentation diagram in the parameter plane σ_1,σ_2. The percentual relative number of cuts is represented in a gray scale mapping. Each grid cell corresponds to a different random sequence of size $N = 4000$ and segment sizes $m = 100$ [(a),(d)], 200 [(b),(e)], and 400 [(c),(f)]. Segmentation was performed with the mean-based [(a)–(c)] and KS [(d)–(f)] algorithms, with $\ell_0 = 10$ and $P_0 = 0.95$.

segmented, while those belonging to white regions are typically unsegmentable. Dark cells indicate oversegmentation.

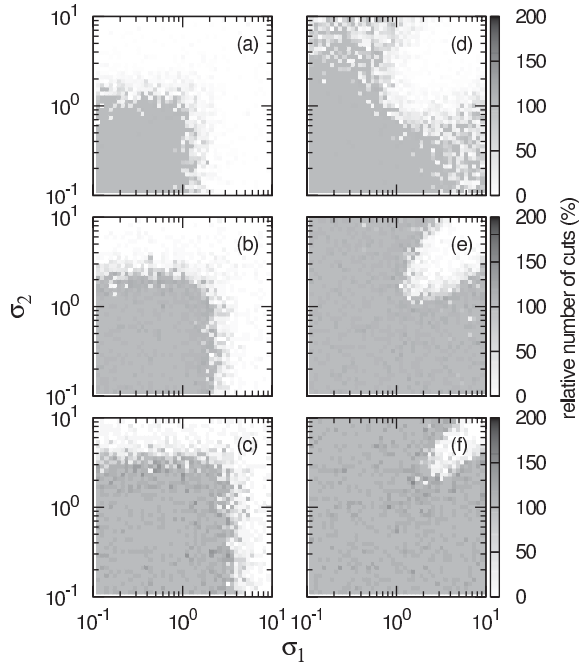

FIG. 10. Segmentation diagram in the parameter plane σ_1,σ_2, as in Fig. 9 but with $\ell_0 = 50$.

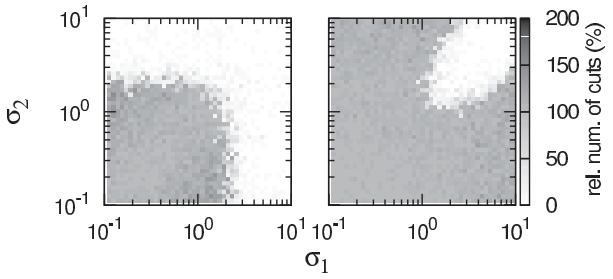

FIG. 11. Segmentation diagram in the parameter plane σ_1,σ_2, as in Figs. 9(b) and 9(e) (i.e., $m = 200$, $\ell_0 = 10$) but with $P_0 = 0.99$.

The mean-based algorithm performs proper segmentation only if the standard deviations are, at most, of the order of the size of the jumps and works well, within the chosen confidence level, around the diagonal ($\sigma_1 \simeq \sigma_2$), as expected. Meanwhile, the KS algorithm is able to segment series in a larger region of parameter space. Segmentation fails when the standard deviations of consecutive segments are not significantly different and are larger than the jump size ($\sigma_1 \simeq \sigma_2 > \Delta\bar{y}$). In both procedures, for larger segment sizes n, the segmentable domain enlarges, but more false cutting points arise. By setting a larger value of ℓ_0, small segments are discarded and the number of false cuts is reduced, as can be seen by comparison of Figs. 9 and 10 which just differ in the value of ℓ_0. Also, false cutting points could be reduced by increasing the value of P_0 (compare the second row of Fig. 9 with Fig. 11, which only differ in the value of P_0).

Moreover, one could still improve the algorithm by adding a further final step, which is the following one. Before accepting a cut, check through the standard KS test the significance of the discrepancy between the right-hand-side portion and its right neighboring segment (born in the previous generation) as well as the left-hand-side portion with its left neighbor, as has been proposed for the mean-based segmentation [10]. For the analyzed series, this step does not introduce a significant improvement (see Fig. 12 to be compared with the second row of Fig. 9), however, if, depending on the analyzed series, one observes oversegmentation, then that step could be straightforwardly added.

Let us comment that as one approaches the frontier of the segmentable region, although segments are recognized, the position of the boundaries of the segments gets more imprecise, the effect of which is reduced by increasing the confidence level.

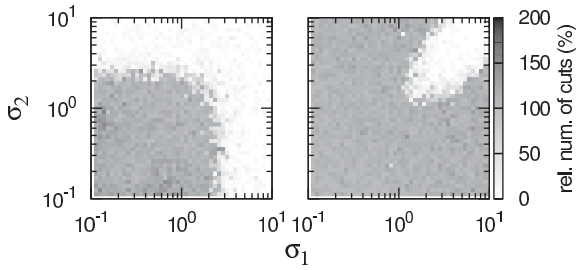

FIG. 12. Segmentation diagram in the parameter plane σ_1,σ_2, as in Figs. 9(b) and 9(e) (i.e., $m = 200$, $\ell_0 = 10$, and $P_0 = 0.95$), with the additional neighbor segment check.

The above diagrams depict the scope of the algorithm for a particular class of artificial series but provide a feeling of its range of applicability and limitations. They also manifest the importance of our method in enlarging the domain of segmentation. There is an infinity of other tests, e.g., with diverse variabilities of means and variances, correlations, and other statistics discrepancies, that could be performed. Also, tests restricted to the comparison of the outcoming statistics could be carried out [11]. However, when applying this or other algorithms, it may be convenient to perform an ad hoc test, depending on the particular statistical characteristics of the analyzed series.

3. Robustness

For the analyzed series, we checked the robustness of the results with respect to the significance level ($P_0 = 0.90, 0.95, 0.99$). The smaller P_0, the larger the tendency to allow small segments, while larger ones are almost unaffected. However, this effect does not significantly change the statistics of segment sizes, as illustrated for heart-rate series in Fig. 13 (right panel). The impact of ℓ_0 was also checked (left panel of Fig. 13). Slopes do not significantly change, except for small values of ℓ_0, as expected, since smaller fragments are allowed with decreasing ℓ_0. Notwithstanding, the probability density of larger segments is not significantly altered. Let us note that the statistics on segments may be affected by the increase of small segments if the studied quantity is correlated with the size. However, this does not seem to happen in the analyzed cases.

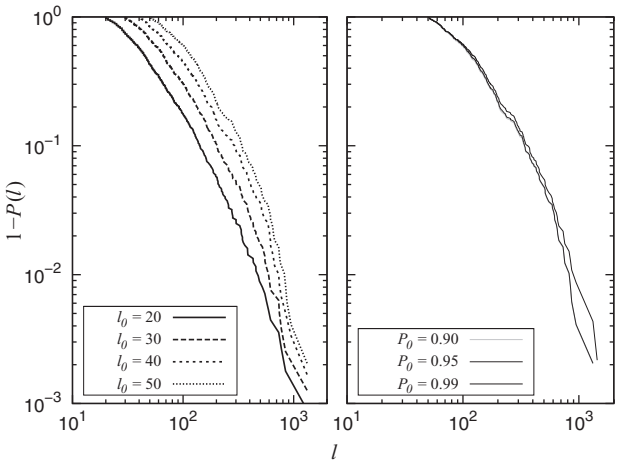

FIG. 13. Left panel: Complementary cumulative distribution of segment sizes obtained for different values of the minimal length ℓ_0 indicated on the figure. Data correspond to a normal individual (n5nn.txt). Right panel: Complementary cumulative distribution of segment sizes at different significance levels indicated on the figure.

We also checked that the same segmentation patches are typically recovered even when analyzing small fragments of the whole series. In fact, significant statistical jumps even for small segments are recognized, prompting a segmentation point.

[1] J. P. Bouchaud and M. Potters, *Theory of Financial Risks: From Statistical Physics to Risk Management* (Cambridge University Press, Cambridgde, 2000); N. Shephard, J. Econometrics **60**, 181 (1994); O. Elerian, S. Chib, and N. Shephard, Econometrica **69**, 1519 (2001).

[2] M. J. Sippl, J. Mol. Biol. **213**, 859 (1990); T. L. Shearer, M. J. van Oppen, S. L. Romano, and G. Worheide, Mol. Ecol. **11**, 2475 (2002).

[3] N. Mordant, A. M. Crawford, and E. Bodenschatz, Phys. Rev. Lett. **93**, 214501 (2004); A. M. Reynolds, New J. Phys. **7**, 58 (2005).

[4] G. Falkovich, K. Gawedzki, and M. Vergassola, Rev. Mod. Phys. **73**, 913 (2001); J. E. Kutzbach and P. J. Guetter, J. Atmos. Sci. **43**, 1726 (1986); C. S. Kochanek, Astrophys. J. **466**, 638 (1996).

[5] J. Kirchner, W. Meyer, M. Elsholz, and B. Hensel, Phys. Rev. E **76**, 021110 (2007).

[6] S. Siegert, R. Friedrich, and J. Peinke, Phys. Lett. A **243**, 275 (1998); C. Anteneodo and S. M. Duarte Queirós, Phys. Rev. E **82**, 041122 (2010).

[7] A. M. van Mourik, A. Daffertshofer, and P. J. Beek, Phys. Lett. A **351**, 13 (2006).

[8] C. Beck and E. G. D. Cohen, Physica A **322**, 267 (2003).

[9] I. Berkes, R. Gabrys, L. Horváth, and P. Kokoszka, J. R. Statist. Soc. B **71**, 927 (2009); R. Killick, P. Fearnhead, and I. A. Eckley, e-print arXiv:1101.1438v1, and references therein.

[10] P. Bernaola-Galván, P. Ch. Ivanov, L. A. Nunes Amaral, and H. E. Stanley, Phys. Rev. Lett. **87**, 168105 (2001).

[11] K. Fukuda, H. E. Stanley, and L. A. Nunes Amaral, Phys. Rev. E **69**, 021108 (2004).

[12] P. Carpena, J. L. Oliver, M. Hackenberg, A. V. Coronado, G. Barturen, and P. Bernaola-Galván, Phys. Rev. E **83**, 031908 (2011).

[13] G. Vaglica, F. Lillo, E. Moro, and R. N. Mantegna, Phys. Rev. E **77**, 036110 (2008).

[14] I. Reyes-Ramírez and L. Guzmán-Vargas, Eur. Phys. Lett. **89**, 38008 (2010).

[15] P. C. Ivanov, Z. Chen, K. Hu, and H. E. Stanley, Phys. A **344**, 685 (2004).

[16] B. Tóth, F. Lillo, and J. D. Farmer, Eur. Phys. J. B **78**, 235 (2010).

[17] P. Bernaola-Galván, R. Román-Roldán, and J. L. Oliver, Phys. Rev. E **53**, 5181 (1996).

[18] P. Bernaola-Galván, I. Grosse, P. Carpena, J. L. Oliver, R. Román-Roldán, and H. E. Stanley, Phys. Rev. Lett. **85**, 1342 (2000).

[19] I. Grosse, P. Bernaola-Galván, P. Carpena, R. Román-Roldán, J. Oliver, and H. E. Stanley, Phys. Rev. E **65**, 041905 (2002).

[20] W. Li, Phys. Rev. Lett. **86**, 5815 (2001); G. Shafer, J. Am. Stat. Assoc. **77**, 325 (1982); A. E. Raftery, Biometrika **83**, 251 (1995).

[21] L. Glass, Chaos **19**, 028501 (2009).
[22] See http://www.physionet.org/challenge/chaos/.
[23] C. Beck, Europhys. Lett. **64**, 151 (2003); Phys. Rev. Lett. **98**, 064502 (2007).
[24] S. Ghashghaie *et al*., Nature (London) **381**, 767 (1996); R. N. Mantegna and H. E. Stanley, *ibid*. **383**, 587 (1996); E. Van der Straeten and C. Beck, Phys. Rev. E **80**, 036108 (2009).
[25] C. Beck, E. G. D. Cohen, and H. L. Swinney, Phys. Rev. E **72**, 056133 (2005); S. M. Duarte Queirós, Physica A **385**, 191 (2007).
[26] Data recorded at Firenze Airport (Italy) with a 30 s sampling rate (during 1–30 September 2009) from an anemometer (identified RWY23, at the top of a 10-m-high pole −40 m above sea level, located at one end of the runway), were gently provided by ENAV S.p.A. Via Salaria, 716-00138 Roma–U.A.A.V. Firenze.