Physics Letters A 380 (2016) 1426–1431

Contents lists available at ScienceDirect

Physics Letters A

www.elsevier.com/locate/pla

Multivariate refined composite multiscale entropy analysis

Anne Humeau-Heurtier

University of Angers, LARIS – Laboratoire Angevin de Recherche en Ingénierie des Systèmes, 62 avenue Notre-Dame du Lac, 49000 Angers, France

ARTICLE INFO

Article history:
Received 18 May 2015
Received in revised form 12 February 2016
Accepted 18 February 2016
Available online 23 February 2016
Communicated by A.P. Fordy

Keywords:
Complexity
Nonlinear dynamics
Entropy
Fractal
Multivariate embedding
Multiscale entropy

ABSTRACT

Multiscale entropy (MSE) has become a prevailing method to quantify signals complexity. MSE relies on sample entropy. However, MSE may yield imprecise complexity estimation at large scales, because sample entropy does not give precise estimation of entropy when short signals are processed. A refined composite multiscale entropy (RCMSE) has therefore recently been proposed. Nevertheless, RCMSE is for univariate signals only. The simultaneous analysis of multi-channel (multivariate) data often over-performs studies based on univariate signals. We therefore introduce an extension of RCMSE to multivariate data. Applications of multivariate RCMSE to simulated processes reveal its better performances over the standard multivariate MSE.

Several entropy measures have been proposed to assess the regularity of times series. Among them, we can cite the sample entropy [1]. Sample entropy is equal to the negative of the natural logarithm of the conditional probability that sequences close to each other for m consecutive data points will also be close to each other when one more point is added to each sequence [1]. However, sample entropy operates on a single scale. Real world data, as physiological data, exhibit high degree of structural richness. Studies based on a single scale are therefore not adapted for real world signals. Analyses on multiple time scales have become necessary. In the 2000s, Costa et al. proposed the multiscale entropy (MSE) to quantify complexity over multiple scales [2,3]. The MSE algorithm is composed of two steps [2,3]: (i) a coarse-graining procedure to derive a set of time series representing the system dynamics on different time scales. The coarse-graining procedure for scale τ is obtained by averaging the samples of the time series inside consecutive but non-overlapping windows of length τ; (ii) the computation of the sample entropy for each coarse-grained time series. MSE has become a prevailing method to quantify the complexity of signals. It has been shown through several studies that MSE is able to underline the general loss of complexity behavior when a living system changes from a healthy state to a pathological state [2,3].

Nevertheless, the coarse-graining procedure used in the MSE algorithm shortens the length of the data that are processed: for an original time series of N samples, the length of the coarse-grained time series at a scale factor τ is N/τ. It has been reported that for an embedding dimension $m = 2$, the sample entropy is significantly independent of the time series length when the number of data points is larger than 750 [4]. For shorter time series, the variance of the entropy estimator may grow very fast with the reduction of the number of data points. Therefore, at large scales, the coarse-grained time series may not be adequately long to obtain an accurate value for the sample entropy. Moreover, for some cases, the sample entropy value may not be defined because no template vectors are matched to one another. These two drawbacks (inaccurate or undefined sample entropy values) lead to problems of accuracy and validity of MSE at large scales. In order to overcome the accuracy concern of MSE, Wu et al. proposed the composite MSE (CMSE) [5]. In the CMSE algorithm, all coarse-grained time series for a scale factor τ are processed to compute their sample entropy (each of the τ coarse-grained time series corresponding to different starting points of the coarse-graining process is used in the CMSE algorithm whereas, in the conventional MSE algorithm, for each scale, only the first coarse-grained time series is taken into account). The CMSE value for a given scale is therefore defined as the mean of several entropy values [5]. Therefore, CMSE estimates entropy more accurately than MSE. Unfortunately, CMSE increases the probability of inducing undefined entropy. This is why a refined CMSE (RCMSE) algorithm has been proposed in 2014, see below [6].

However, MSE, CMSE, and RCMSE are able to process univariate data only. For multivariate time series, the three algorithms treat individual time series separately. This may be satisfactory only if the individual signals are statistically independent or at

E-mail address: anne.humeau@univ-angers.fr.

http://dx.doi.org/10.1016/j.physleta.2016.02.029
0375-9601/

least uncorrelated, which is often not the case when real world signals from a given system are registered simultaneously. To overcome this shortcoming, an extension of the MSE algorithm to multivariate data has been proposed in 2011: the multivariate MSE (MMSE) [8,9]. MMSE is able to operate on any number of data channels and provides a robust relative complexity measure for multivariate data [8,9]. MMSE has been used in studies from different fields [10–13]. However, the same concerns as MSE are found in MMSE. This is why in this work we propose an extension of the RCMSE algorithm to a more general case. To this end, we introduce the multivariate RCMSE (MRCMSE), and evaluate its performances on synthetic multivariate processes.

1. Multivariate refined composite multiscale entropy

1.1. Refined composite multiscale entropy

RCMSE aims at improving the CMSE algorithm because, as mentioned previously, CMSE estimates entropy more accurately than MSE but increases the probability of inducing undefined entropy [6,7].

For a discrete time series $x = \{x_i\}_{i=1}^N$, the RCMSE algorithm is based on the following three steps [6]

1. the kth coarse-grained time series for a scale factor τ is defined as $y_k^{(\tau)} = \{y_{k,j}\}_{j=1}^{N/\tau}$ where [5]

$$y_{k,j}^{(\tau)} = \frac{1}{\tau} \sum_{i=(j-1)\tau+k}^{j\tau+k-1} x_i, \qquad 1 \le j \le \frac{N}{\tau}, 1 \le k \le \tau \tag{1}$$

2. for each scale factor τ, and for all τ coarse-grained time series, the number of matched vector pairs $n_{k,\tau}^{m+1}$ and $n_{k,\tau}^m$ is computed, where $n_{k,\tau}^m$ represents the total number of m-dimensional matched vector pairs and is computed from the kth coarse-grained time series at a scale factor τ
3. RCMSE is then defined as [6]

$$RCMSE(x, \tau, m, r) = -\ln\left(\frac{\sum_{k=1}^{\tau} n_{k,\tau}^{m+1}}{\sum_{k=1}^{\tau} n_{k,\tau}^{m}}\right). \tag{2}$$

Using the same notation, CMSE is defined as [6]

$$CMSE(x, \tau, m, r) = \frac{1}{\tau}\sum_{k=1}^{\tau}\left(-\ln\frac{n_{k,\tau}^{m+1}}{n_{k,\tau}^{m}}\right). \tag{3}$$

The CMSE value is therefore undefined when one of the values $n_{k,\tau}^{m+1}$ or $n_{k,\tau}^m$ is zero. By opposition, RCMSE value is undefined only when all $n_{k,\tau}^{m+1}$ or $n_{k,\tau}^m$ are zeros. It has been reported that RCMSE outperforms CMSE in validity, accuracy of entropy estimation, independence of data length, and computational efficiency [6]. RCMSE has been used in recent studies [14].

1.2. Multivariate multiscale entropy

MMSE is an extension of the MSE algorithm to multivariate data. MMSE relies on the same steps as MSE [8,9]: (i) a coarse-graining procedure; (ii) a sample entropy computation for each coarse-grained time series. However, due to the multivariate nature of the data processed by MMSE, these two steps are adapted to multivariate signals. Thus, for the coarse-graining procedure, temporal scales are defined by averaging a p-variate time series $\{x_{l,i}\}_{i=1}^N$ ($l = 1, \ldots, p$ is the channel index and N is the number of samples in every channel) over non-overlapping time segments of increasing length. Thus, for a scale factor τ, a coarse-grained multivariate time series is computed as $y_{l,j}^{(\tau)} = \frac{1}{\tau}\sum_{i=(j-1)\tau+1}^{j\tau} x_{l,i}$ where $1 \le j \le N/\tau$, and the channel index l goes from 1 to p. For the entropy computation, the multivariate sample entropy (MSampEn) is used for each coarse-grained multivariate. The MSampEn algorithm is an extension of the univariate sample entropy [1]. For a tolerance level r, MSampEn is calculated as the negative of the natural logarithm of the conditional probability that two composite delay vectors close to each other in a m dimensional space will also be close to each other when the dimensionality is increased by one. The detailed MSampEn algorithm can be found in [8,9].

1.3. Multivariate refined composite multiscale entropy

Based on RCMSE and MSampEn, we define the MRCMSE algorithm as follows:

1. for a p-variate time series $\{x_{l,i}\}_{i=1}^N$, $l = 1, \ldots, p$, where p denotes the number of variates (channels) and N is the number of samples in each variate, and for a scale factor τ, determine the coarse-grained multivariate time series $\{y_{l,k,j}^{(\tau)}\}_{j=1}^{N/\tau}$ as

$$y_{l,k,j}^{(\tau)} = \frac{1}{\tau}\sum_{i=(j-1)\tau+k}^{j\tau+k-1} x_{l,i}, \tag{4}$$

where $1 \le j \le N/\tau$, $1 \le k \le \tau$, $l = 1, \ldots, p$
2. for each coarse-grained multivariate compute $B^m(r)$ and $B^{m+1}(r)$ as defined in Table 1 [8,9]. For the coarse-grained multivariate $\{y_{l,k,j}^{(\tau)}\}_{j=1}^{N/\tau}$, $l = 1, \ldots, p$, these two quantities are denoted as $B_{k,\tau}^m(r)$ and $B_{k,\tau}^{m+1}(r)$, respectively
3. compute $MRCMSE(\tau, M, r, \epsilon, N) = -\ln\left(\frac{\sum_{k=1}^{\tau} B_{k,\tau}^{m+1}(r)}{\sum_{k=1}^{\tau} B_{k,\tau}^{m}(r)}\right)$.

2. Results and discussion

In order to analyze the behavior of MRCMSE on multivariate data, we generated a trivariate time series, where originally all the data channels were realizations of mutually independent white noise [8]. We then gradually decreased the number of variates representing white noise (from 3 to 0) and simultaneously increased the number of data channels representing independent $1/f$ noise (from 0 to 3), as already proposed in [8,9]. The total number of variates was always three. For each kind of trivariate data, 50 independent realizations were simulated and, for each realization, 10 000 samples were generated in each variate. Scales were chosen between 1 and 20. Therefore, the shortest coarse-grained time series had a length of 500 samples. For each channel the embedding dimension m_k was chosen equal to 2 and the threshold r was fixed to 0.15 × (standard deviation of the normalized time series) for each data channel. We recall that a multivariate time series is considered more structurally complex than another if, for most of the scale factors τ, its multivariate entropy values are higher than those of the other time series. When the multivariate entropy values decrease with the scale factor τ, the time series that is processed only contains information at the smallest scales. It is thus not structurally complex. This is the same as what is observed for the univariate MSE where sample entropy values of random white noise (uncorrelated) decrease with the scale factor whereas for $1/f$ noise (long-range correlated), the sample entropy values are constant over multiple scales.

For each above-mentioned simulated trivariate data, MRCMSE and MMSE were determined. The results are presented in Fig. 1. We observe that MRCMSE and MMSE curves are close to each other. Moreover, the higher the number of variates representing $1/f$ noise, the higher the multivariate entropy value, for a given scale factor τ. This is in accordance with what was expected. This behavior is the same for MRCMSE and MMSE.