Physica A 421 (2015) 583–593

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

EMD based refined composite multiscale entropy analysis of complex signals

Jing Wang *, Pengjian Shang, Jianan Xia, Wenbin Shi

Department of Mathematics, School of Science, Beijing Jiaotong University, Beijing 100044, PR China

H I G H L I G H T S

- Trends superimposed in signals greatly affect its complexity measurement.
- EMD–RCMSE method gives good complexity measurement of signals with various trends.
- Traffic signals are more complex than the results calculated from traditional MSE.

A R T I C L E I N F O

Article history:
Received 20 October 2014
Received in revised form 1 December 2014
Available online 9 December 2014

Keywords:
Complexity
Multiscale entropy
Empirical mode decomposition
Traffic signal
Periodical trend

A B S T R A C T

Multiscale entropy (MSE) is an effective method to measure the complexity of signals from complex systems, which has been applied to various fields successfully. However, MSE may yield an inaccurate estimate of entropy and induce undefined entropy as the coarse-graining procedure reduces the length of data considerably at large scales. Refined composite multiscale entropy (RCMSE) is then developed to solve this problem. However, trends superimposed in signals may significantly affect the complexity measurement. Thus we introduce an empirical mode decomposition based RCMSE, called EMD–RCMSE to first eliminate such effects of trends and then measure the complexity of signals. It is validated from simulated signals and has good estimation. In addition, this method is also applied to study the complexity of traffic signals and we obtain some interesting results: (1) Traffic signals are more complex than the results showing from RCMSE/MSE; (2) Weekday and weekend patterns (different combination of trends) greatly affect the results; (3) Complexity indices change with time at each day, due to the degree of human activities.

1. Introduction

Complex systems are regulated by interacting mechanisms that operate across multiple spatial and temporal scales [1]. Signals generated from these systems usually exhibit complex fluctuations, unpredictable perturbation, together with high levels of nonlinearity and nonstationarity but contain rich information about the underlying dynamics. This may challenge the understanding and studies for such systems. However, there have been successful attempts by researchers applying various methods to analyze signals for complex systems, e.g. physiological systems [2–5], traffic systems [6–12], financial systems [13–17], geographical and meteorologic systems [18–21]. These findings help us to unveil the rich structures of complex systems and inspire further understanding of their underlying mechanisms.

* Corresponding author.
E-mail address: 05271053@bjtu.edu.cn (J. Wang).

http://dx.doi.org/10.1016/j.physa.2014.12.001
0378-4371/

How to measure the complexity of output signals from complex systems has received a considerable amount of attention in recent years. Numerous techniques have been developed [22,23], among which the multiscale entropy (MSE) proposed by Costa et al. [1,24,25] has been applied in many fields successfully. The MSE algorithm calculates sample entropy (SampEn) over a range of scales to represent the complexity of a time series. It resolves the contradiction between the lower entropy and higher complexity of $1/f$ noise compared with white noise. The procedure for calculation of MSE can be summarized in two steps as follows: (1) construct coarse-grained time series according to a scale factor τ; (2) calculate the SampEn of each coarse-grained time series. For large scale τ, the coarse-grained time series may not be adequately long to obtain an accurate SampEn. Moreover, the SampEn, in some cases, is undefined as no template vectors are matched to one another. Inaccurate or undefined SampEns lead to the reduction of reliability of MSE algorithm [26]. Wu et al. [26] developed refined composite multiscale entropy (RCMSE) to solve these problems and demonstrated that RCMSE can increase the accuracy of entropy estimation and reduce the probability of inducing undefined entropy. In addition, the RCMSE algorithm is suggested for use when the coarse-grained time series is shorter than 750 [26].

The existence of trends in signals generated by complex systems is so common that it is almost unavoidable [27]. For example, the density of air due to gravity has a trend at different altitudes [28]; the air temperature in different geographic locations and the water flow of rivers have a periodic trend due to seasonal changes [29–32]; the traffic speed signals have periodic trends due to frequent human activities, like daily repeated trend [33]. Such trends superimposed in signals may have effects on the complexity, and also may change results of traditional multiscale entropy analysis. Although the overall fluctuation of signals is of great importance, which may guide the prediction of its future behavior, local subtle fluctuations are also crucial for the analysis of signals. Such seemingly subtle variations in signals often carry richer information which may play a big role in the construction for the whole complexity of complex systems. Therefore, it is necessary to focus on the study of complexity of intrinsic dynamics of signals after filtering out the trends. In this study, we investigate the complexity of signals via applying an empirical mode decomposition technique (EMD) [34] based RCMSE algorithm, called EMD–RCMSE, which can effectively eliminate the effects of trends on complexity.

The structure of the paper is as follows. In Section 2, we briefly review SampEn, MSE and EMD techniques. Then the EMD–RCMSE algorithm is introduced in this section. In Section 3, some simulated signals are generated to validate the effectiveness of EMD–RCMSE algorithm. Section 4 is devoted to the detailed complexity analysis of EMD–RCMSE on traffic signals. Finally, a conclusion is presented in Section 5.

2. Methods

In this section, the theoretical backgrounds of sample entropy (SampEn), refined composite multiscale entropy (RCMSE) and empirical mode decomposition (EMD) algorithm are briefly reviewed. The concept of EMD–RCMSE algorithm is also introduced.

2.1. Refined composite multiscale entropy

Sample entropy (SampEn) provides an improved evaluation of time series regularity and should be a useful tool in studies of the dynamics of different systems [35]. We first briefly review the SampEn algorithm. Let $X = \{x_1, x_2, \dots, x_N\}$ represent a time series of length N. The SampEn algorithm can be summarized as follows: (1) Construct template vectors with dimension m by using $x_i^m = \{x_i, x_{i+1}, \dots, x_{i+m-1}\}(1 \le i \le N-m)$. (2) A match occurs when the distance between two template vectors (x_i^m, x_j^m) is smaller than a predefined tolerance r. The distance between two vectors is calculated by using the infinity norm: $d_{ij}^m = \|x_i^m - x_j^m\|_\infty$ $(1 \le i, j \le N-m, j \ne i)$. (3) Then (x_i^m, x_j^m) is defined as an m-dimensional matched vector pair if $d_{ij}^m \le r$. Let n_m represent the total number of m-dimensional matched vector pairs. (4) Repeat Steps 1–3 for $m = m+1$, and n^{m+1} is obtained to represent the total number of $(m+1)$-dimensional matched vector pairs. (5) The SampEn is defined as the logarithm of the ratio of n^{m+1} to n^m; that is $SampEn(x, m, r) = -\ln(n^{m+1}/n^m)$.

For a short time series, the SampEn algorithm may cause the following problems [26]: (a) the SampEn often yields an inaccurate estimation; and (b) n^m or n^{m+1} in above procedures may be zero, thus inducing an undefined SampEn. To obtain a reasonable SampEn, the length of the time series is suggested to be in the range of 10^m–30^m [36]. Regarding $m = 2$, Costa et al. [37] suggested that the time series should be longer than 750 data points. Then we introduce the RCMSE algorithm [26] to solve such problems, which consists of the following procedures:

(1) At a scale factor of τ, the original time series is divided into non-overlapping windows of length τ and the data points inside each window are averaged. As shown in Fig. 1, τ coarse-grained time series are divided from the original time series for a scale factor τ. The kth coarse-grained time series $y_k^\tau = \{y_{k,1}^{(\tau)}, y_{k,2}^{(\tau)}, \dots, y_{k,p}^{(\tau)}\}$ is defined as follows:

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

Above coarse-graining procedure is repeated to obtain the coarse-grained time series on different time scales.

(2) At each scale factor τ, the number of matched vector pairs, $n_{k,\tau}^{m+1}$ and $n_{k,\tau}^m$, is calculated for all τ coarse-grained series.