Chaos, Solitons and Fractals 81 (2015) 117–135

Contents lists available at ScienceDirect

Chaos, Solitons and Fractals

Nonlinear Science, and Nonequilibrium and Complex Phenomena

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

Review

Complexity testing techniques for time series data: A comprehensive literature review

Ling Tang[a], Huiling Lv[b], Fengmei Yang[b], Lean Yu[a,*]

[a] School of Economics and Management, Beijing University of Chemical Technology, Beijing 100029, China
[b] School of Science, Beijing University of Chemical Technology, Beijing 100029, China

ARTICLE INFO

Article history:
Received 11 April 2015
Accepted 7 September 2015

Keywords:
Complexity test
Time series data
Literature review
Fractality
Chaos
Entropy

ABSTRACT

Complexity may be one of the most important measurements for analysing time series data; it covers or is at least closely related to different data characteristics within nonlinear system theory. This paper provides a comprehensive literature review examining the complexity testing techniques for time series data. According to different features, the complexity measurements for time series data can be divided into three primary groups, i.e., fractality (mono- or multi-fractality) for self-similarity (or system memorability or long-term persistence), methods derived from nonlinear dynamics (via attractor invariants or diagram descriptions) for attractor properties in phase-space, and entropy (structural or dynamical entropy) for the disorder state of a nonlinear system. These estimations analyse time series dynamics from different perspectives but are closely related to or even dependent on each other at the same time. In particular, a weaker self-similarity, a more complex structure of attractor, and a higher-level disorder state of a system consistently indicate that the observed time series data are at a higher level of complexity. Accordingly, this paper presents a historical tour of the important measures and works for each group, as well as ground-breaking and recent applications and future research directions.

1. Introduction

To investigate and understand time series data, data characteristic testing has created increasing interest in terms of both the theoretical and application perspectives. Typically, characteristics of time series data include nonstationarity (or stationarity) [1], nonlinearity (or linearity) [2], complexity [3], chaos [4], fractality [5], cyclicity [6], seasonality [7], saltation [8], stochasticity [9] and so on. Amongst them, complexity can be viewed as one of the most important measurements in time series data analysis, as it covers or is closely related to other various data characteristics within nonlinear system theory. On the one hand, complexity can provide another vivid description for diverse nonlinear characteristics, e.g., chaoticity, fractality, irregularity, long-range memorability and so on [10,11]. In particular, various nonlinear characteristics attempt to describe the intrinsic structure of the target data dynamics, for which complexity can present a generic quantitative estimation. Generally, higher-level complexity indicates a more intricate, disordered, irregular dynamic system. On the other hand, the complexity characteristic also determines the features of various inner factors (e.g., central trend and cyclical, seasonal and mutable patterns) and their interrelationship [12]. Lower-level complexity indicates the larger impact or governing power of regular factors or inner rules (e.g., central trend and cyclical patterns) on the whole data dynamics [11,12].

Therefore, complexity has become an increasingly dominant estimator in analysing time series dynamics, especially for economic data to evaluate market efficiency [13] and for physical and biological data to capture the hidden rules

* Corresponding author. Tel.: +86 1064438793.
E-mail address: yulean@amss.ac.cn (L. Yu).

http://dx.doi.org/10.1016/j.chaos.2015.09.002

[14,15]. Actually, complex systems are special cases of nonlinear systems, and most natural systems are complex systems [10]. In particular, the term 'complex' describes an intermediate state between a completely regular system and a completely random system [16]. Generally, a lower-level complexity indicates that the observed system is more likely to follow a deterministic process that can be finely captured and predicted, while a higher-level complexity represents less regular rules controlling the data dynamics, which might otherwise be more unpredictable and more difficult to understand [11].

According to the existing literature, numerous complexity testing techniques have been developed for time series data. Based on different features, they can be generally divided into three categories: fractality, methods derived from nonlinear dynamics and entropy [17]. Each category tends to describe the dynamic system of time series data from a different perspective. Although they have different focusses, these three groups are closely related to or even dependent on each other. In particular, a weaker self-similarity (or system memorability) measured via fractality, a more complex structure of attractor via methods derived from nonlinear dynamics, and a higher-level disorder state of the system via entropy consistently indicate that the observed time series data are at a higher-level of complexity.

Regarding the literature reviews on complexity testing techniques for time series data, there are some important studies. For example, Lopes and Betrouni [18] focussed on the fractality estimation for medical signal data and divided the existing tools into mono- and multi-fractal techniques. Sun et al. [19] provided a survey of the popular methods for estimating fractal dimension and their applications to remote sensing problems. Kantelhardt [20] gave an introduction to mono- and multi-fractality analysis methods for stationary and non-stationary time series data. Zanin et al. [21] analysed the theoretical foundations of PE, as well as recent applications to economic markets and biomedical systems. Alcaraz and Rieta [22] provided a review of sample entropy (SampEn) in the context of non-invasive analysis of atrial fibrillation. Marwan et al. [23] presented a comprehensive review of recurrence-based methods and their applications.

However, the existing literature reviews of complexity testing techniques for time series data were insufficient, and improvements can be made from two main perspectives. First, most literature reviews only concentrate on one group of complexity testing techniques while ignoring some other important types. For example, in Refs. [18–20], only the methods for fractality exploration were considered; Zanin et al. [21] only concentrated on PE, and Alcaraz et al. [22] only considered SampEn; Marwan et al. [23] only focussed on recurrence-based methods. However, according to existing methods, three main groups are included in complexity analysis, i.e., fractality, methods derived from nonlinear dynamics and entropy, which describe data dynamics from different perspectives. That is, some important testing techniques were ignored in the existing literature reviews. Second, some studies presented a survey only within a certain application field. For example, Lopes and Betrouni [18] focussed on medical signal data, Sun et al. [19] focussed on remote sensing problems, and Zanin et al. [21] focussed on economic markets and biomedical systems. However, because most natural systems belong to complex systems, complexity testing has been widely applied to various research fields for different data. To the best of our knowledge, there are few studies on literature reviews of complexity testing techniques that address not only different types of methods but also various application fields. Given such a background, this paper tries to fill in this gap and provide a comprehensive literature review on complexity testing techniques for time series data in terms of measures and important works in each group of fractality, methods derived from nonlinear dynamics and entropy explorations, as well as respective ground-breaking and recent applications and future research directions.

Table 1
The abbreviations for different complexity testing techniques for time series data.

Abbreviation	Full meaning
ApEn	Approximate entropy
BMA	Backward moving average
CMA	Centred moving average
DFA	Detrended fluctuation analysis
DMA	Detrended moving average
FuzzyEn	Fuzzy entropy
MF-DFA	Multifractal detrended fluctuation analysis
MF-DMA	Multifractal detrended moving average analysis
PE	Permutation entropy
R/S	Rescaled range
RP	Recurrence plot
RQA	Recurrence qualification analysis
SampEn	Sample entropy
SE	Spectral entropy
SPMF	Standard partition function multi-fractal formalism
WE	Wavelet entropy
WTMM	Wavelet transform modulus maxima

The main aim of this paper is to provide a comprehensive literature review of the complexity testing techniques for time series data. The remaining part of the paper is organised as follows. The analytical framework for complexity testing techniques is given in Section 2. Sections 3–5 present a historical overview of the measures and important works in fractality, methods derived from nonlinear dynamics and entropy, together with respective ground-breaking and recent applications and future research directions. Section 6 summarises the main ideas of the three groups of complexity testing. Finally, Section 7 presents the paper's conclusions.

2. Analytical framework for literature review

According to the literature, numerous techniques have been developed and used to estimate the complexity of time series data. Based on different features, these methods can be divided into three main categories, i.e., fractality, methods derived from nonlinear dynamics and entropy. Accordingly, the abbreviations for different measures are listed in Table 1, and the analytical framework of the literature review is illustrated in Fig. 1.

Fractality theory mainly focusses on self-similarity (or system memorability or long-term persistence) by analysing the metric scaling behaviour of a data system, which mainly includes the techniques of R/S analysis [2], DFA [24–28], DMA [29], SPMF [30], WTMM [31], MF-DFA [32] and MF-DMA [33].