Dispersion complexity–entropy curves: An effective method to characterize the structures of nonlinear time series

Runze Jiang; Pengjian Shang

Chaos 34, 033137 (2024)
https://doi.org/10.1063/5.0197167
Dispersion complexity–entropy curves: An effective method to characterize the structures of nonlinear time series

Cite as: Chaos 34, 033137 (2024); doi: 10.1063/5.0197167
Submitted: 11 January 2024 · Accepted: 27 February 2024 · Published Online: 25 March 2024

Runze Jiang and Pengjian Shang

AFFILIATIONS
School of Mathematics and Statistics, Beijing Jiaotong University, Beijing 100044, China

a)Author to whom correspondence should be addressed: 23111517@bjtu.edu.cn
b)Electronic mail: pjshang@bjtu.edu.cn

ABSTRACT

The complexity–entropy curve (CEC) is a valuable tool for characterizing the structure of time series and finds broad application across various research fields. Despite its widespread usage, the original permutation complexity–entropy curve (PCEC), which is founded on permutation entropy (PE), exhibits a notable limitation: its inability to take the means and amplitudes of time series into considerations. This oversight can lead to inaccuracies in differentiating time series. In this paper, drawing inspiration from dispersion entropy (DE), we propose the dispersion complexity–entropy curve (DCEC) to enhance the capability of CEC in uncovering the concealed structures within nonlinear time series. Our approach initiates with simulated data including the logistic map, color noises, and various chaotic systems. The outcomes of our simulated experiments consistently showcase the effectiveness of DCEC in distinguishing nonlinear time series with diverse characteristics. Furthermore, we extend the application of DCEC to real-world data, thereby asserting its practical utility. A novel approach is proposed, wherein DCEC-based feature extraction is combined with multivariate support vector machine for the diagnosis of various types of bearing faults. This combination achieved a high accuracy rate in our experiments. Additionally, we employ DCEC to assess stock indices from different countries and periods, thereby facilitating an analysis of the complexity inherent in financial markets. Our findings reveal significant insights into the dynamic regularities and distinct structures of these indices, offering a novel perspective for analyzing financial time series. Collectively, these applications underscore the potential of DCEC as an effective tool for the nonlinear time series analysis.

Published under an exclusive license by AIP Publishing.

Nonlinear time series widely exists in various fields including engineering, finance, physiology, and physics. In recent years, many researchers have been devoting to put forward effective methods to distinguish nonlinear time series with different structures. While entropy serves as a common measure for analyzing time series, it falls short in accurately measuring their complexity levels. Consequently, complexity–entropy curve (CEC) is proposed combining the concept of Rényi entropy, Tsallis entropy, and statistical complexity measures (SCMs), which provides a novel perspective to visualize the richer information embedded within time series. However, the original permutation complexity–entropy curve (PCEC) method predominantly focuses on permutation patterns, which may overlook the crucial information contained within the means and amplitudes of time series. In this paper, we combine dispersion entropy (DE) and SCM to construct CEC and call it dispersion complexity–entropy curve (DCEC). Both simulated and real-world data are used to demonstrate the efficacy of our enhanced method. The results underscore the effectiveness of the proposed approach in revealing the complexity degree and inherent structures within the time series data.

I. INTRODUCTION

Measuring the complexity of nonlinear time series is a common research issue in the realm of nonlinear dynamics, as it enables researchers to unveil the inherent structures and characteristics of time series. Over the past few decades, numerous methods have been put forward to quantify complexity. Algorithmic complexity, including time complexity and space complexity, stands out as...
a widely employed metric to gauge the efficiency of a procedure. Mandelbrot introduced the concept of fractal dimension as a measure of irregularity, reflecting the effectiveness of complex forms in occupying space. Lyapunov exponent, employed as the criterion to identify the chaotic motion, is defined as the average exponential divergence of adjacent trajectories in phase space. These methods, along with various other technologies, are widely applied in diverse fields such as physics, computer science, and biomedicine. In recent years, novel approaches have emerged for assessing complexity based on Shannon’s information theory. In 2002, Bandt and Pompe introduced permutation entropy (PE) as a means to measure the complexity and irregularity of nonlinear time series, capable of distinguishing periodic, chaotic, and stochastic processes. Subsequently, several enhanced PE formats were proposed such as weight-permutation entropy (WPE) or reverse permutation entropy (RPE). These innovative methods have significantly advanced PE theories, obtaining noteworthy results in both research and practical applications. However, PE exhibits limitations under certain circumstances. Rosso et al. highlighted a notable issue that when calculating PE for a logistic map with complete chaos, the results resemble some stochastic processes like 1/f noise. This similarity can create confusion in distinguishing nonlinear time series. To address this concern, a novel metric called statistical complexity measure (SCM) was introduced. SCM, denoted as C, is defined as the product of entropy and disequilibrium, where disequilibrium is the distance between the probability distribution of the pattern in time series and uniform distribution (usually measured using Euclidean distance or Jenson–Shannon divergence). It is important to note that when the probability distribution of patterns is uniform, the entropy value H reaches its maximum, and C equals zero. Conversely, in cases of complete order within the system, the results are reversed. Therefore, SCM serves as a balanced method that integrates entropy and disequilibrium, offering a means to detect the hidden structure between the two extreme states of nonlinear time series. Rosso et al. used entropy (H) and SCM (C) to construct the plane known as complexity–entropy causality plane. This approach enables a more effective distinction between chaotic and stochastic series, as their respective (H, C) points on the plane occupy distinct positions. While the complexity–entropy causality plane has demonstrated its efficacy in generating numerous positive outcomes, it faces challenges in distinguishing certain chaotic systems from stochastic processes due to the close proximity of their (H, C) points. In response to this issue, researchers have introduced a fractal causality plane that integrates Rényi entropy and Tsallis entropy. Both Rényi entropy and Tsallis entropy are the nonparametric generalizations of the Shannon entropy, which significantly enhancing the capabilities of the traditional entropy method and have been widely applied in many areas such as time–frequency analysis, physics, and picture processing. In 2017, Ribeiro et al. introduced Tsallis complexity–entropy curves, building upon permutation entropy. By adjusting the Tsallis parameter q, changes in the values of H and C lead to the creation of an orbit in the plane. Expanding on this concept, Jauregui et al. proposed the Rényi complexity–entropy curve with the goal of distinguishing between periodic and chaotic systems. Subsequently, several innovative methods have emerged to enhance the original complexity–entropy curves. Mao et al. successfully differentiated stochastic processes from chaotic systems by combining power spectral entropy with Rényi and Tsallis complexity-entropy curves. Peng et al. constructed complexity–entropy curves by utilizing ordinal pattern transition network for time series analysis. And there are also several derivative methods that have been proposed. In our work, we propose the use of dispersion entropy (DE) in conjunction with Rényi and Tsallis complexity–entropy curves, termed dispersion complexity–entropy curves (DCEC), for time series analysis. DE, an enhanced entropy method widely applied in various research domains, overcomes the limitations of previous entropy measures. In comparison to sample entropy (SE), DE significantly reduces computation time, especially for lengthy signals. Additionally, when compared to PE, DE adeptly considers the amplitudes of time series for analysis. Various improved formats, including reverse dispersion entropy (RDE) and fluctuation-based dispersion entropy (FDE), further enhance the versatility of DE. To demonstrate the effectiveness of DCEC, we conduct experiments using simulated data. Additionally, we extend the curves from a two-dimensional plane to a three-dimensional space, improving clarity in revealing relationships and distinctions among different nonlinear time series. Furthermore, we apply our method to real-world data to diagnose faulty bearings and analyze the structures of stock markets, confirming the effectiveness of our approach in uncovering the inner structure of nonlinear time series.

The organization of this article is as follows. Section II introduces the definition of DE, SCM, Rényi entropy, and Tsallis entropy. We combine these concepts and propose our approach to formulate the DCEC. In Sec. III, we perform experiments using simulated data, which include the logistic map, colored noises, and various chaotic systems, to demonstrate the effectiveness of DCEC. In Sec. IV, we present the application in fault bearings diagnosis and financial time series by DCEC. Finally, the conclusions are given in Sec. V.

II. METHODS

A. Dispersion entropy

Dispersion entropy (DE) is a state-of-art method to detect the dynamic feature of different nonlinear time series. It not only reduces the computation time but also takes the amplitudes of the series into consideration. For a given time series $X: \{x_i, 1 \leq i \leq N\}$ with length $N$, the procedure of DE is as follows:

Step 1: Mapping the elements of the time series $X: \{x_i, 1 \leq i \leq N\}$ into $c$ classes, where $c$ is the number of symbolization ruled by researchers. In this procedure, the normal cumulative distribution function (NDCF) is used to map the original series $X$ into $Y: \{y_i, 1 \leq i \leq N\}$ from 0 to 1. The NDCF function is

$$y_i = \frac{1}{\sigma \sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2\sigma^2}} dx,$$

where $\mu$ and $\sigma$ represent the mean value and standard deviation of the original time series, respectively. Then, we use the linear method to assign each $y_i$ to an integer from 1 to $c$ by the equation $z_i = \text{round}(c \cdot y_i + 0.5)$. In this way, we get the transformed elements $z_i^*: i = 1, 2, \ldots, N$. 

Chaos 34, 033137 (2024); doi: 10.1063/5.0197167
Published under an exclusive license by AIP Publishing
Step 2: Choose the embedding dimension $m$ and time delay $t$ and then a series of embedding vector $z^{m|t} = \{z_1^{m|t}, z_2^{m|t}, \ldots, z_{N-(m-1)t}^{m|t}\}$, $i = 1, 2, \ldots, N - (m - 1)t$ are obtained. Each series $z^{m|t}$ is matched with a dispersion pattern $\pi_{v_1v_2...v_m}$, where $z_i = v_1, z_i^{m|t} = v_1$, and $z_{v_i(v_i+1)}^{m|t} = v_{m-1}$. The number of possible dispersion patterns is equal to $c^m$ since each element in $z_i^{m|t}$ is an integer from 1 to $c$.

Step 3: For each of the $c^m$ possible patterns, the frequency can be calculated by
\[
p(\pi_{v_1v_2...v_m}) = \text{Number } \{i \mid i \leq N - (m-1)t, z_i^{m|t} \text{ has type } \pi_{v_1v_2...v_m}\}. \tag{2}
\]
In this way, each of the $z_i^{m|t}$, $i = 1, 2, \ldots, N -(m-1)t$, can be matched with a pattern from $c^m$ types.

Step 4: Finally, the definition based on Shannon entropy is applied to calculate the DE of the time series when the embedding dimension is $m$, the time delay is $t$, and the number of classes is $c$,
\[
DE(X, m, c, t) = - \sum_{i=1}^{c} p(\pi_{v_1v_2...v_m}) \cdot \ln( p(\pi_{v_1v_2...v_m}) ) \tag{3}
\]
It is evident from the definition of DE that for all the possible probability distributions, the maximum value of DE ($\ln c^m$) is achieved when each pattern occurs with equal frequency. Thus, the normalized distribution entropy (NDE) can be calculated as
\[
NDE(X, c, m, t) = DE(X, c, m, t)/\ln(c^m). \tag{4}
\]

### B. Statistical complexity measure

The statistical complexity measure (SCM), introduced by Martin et al., aims to assess the complexity of various nonlinear time series. In the realm of nonlinear dynamic systems, there exist two contrasting scenarios—complete order and maximum randomness—each characterized by a straightforward structure, resulting in a statistical complexity of 0. However, within the spectrum of these extremes, diverse complex structures are concealed across different types of systems. To address this challenge, SCM was devised. When considering probability distribution $P = \{p_i, i = 1, 2, \ldots, N\}$, Shannon information entropy $S[P] = -\sum_{i=1}^{N} p_i \ln(p_i)$ serves as a tool to measure the uncertainty and irregularity for nonlinear time series. Notably, when $P$ takes the form of a uniform distribution, Shannon entropy attains its maximum value. However, uniform distribution has the simplest structure for its randomness. Nevertheless, a uniform distribution is structurally simplistic due to its inherent randomness. Relying solely on entropy values for measuring the complexity of a system is, therefore, inaccurate and lacks rigor. For reaching the goal of detecting the complexity comprehensively, an additional complexity measure termed “disequilibrium” was introduced. This measure, combining with Shannon entropy, forms the two components of the SCM, facilitating a better understanding of nonlinear time series complexity.

Disequilibrium, often denoted by $Q$, depicts the distance between the probability distribution of system $P$ and uniform distribution $P_e$, where $P_e = [1/N, 1/N, \ldots, 1/N]$. In this passage, Jensen–Shannon (JS) divergence is used to measure the distance between two distributions. For distribution $P$ and $P_e$, JS divergence $\mathcal{J}[P, P_e]$ is described as
\[
\mathcal{J}[P, P_e] = \frac{1}{2} K(P \mid (P + P_e)/2) + \frac{1}{2} K(P_e \mid (P + P_e)/2)
\]
\[
= S[(P + P_e)/2] - \frac{S[P]}{2} - \frac{S[P_e]}{2}, \tag{5}
\]
where $K(P \mid Q) = -\sum_{i} p_i \ln(q_i/p_i)$ is the Kullback–Leibler divergence of the distribution $P$ and $Q$. By this definition, it is easy to see that $\mathcal{J}[P, P_e]$ reaches the maximum value when the probability of one state equals 1, which means the system is completely ordered. The maximum value $Q_b$ can be calculated by
\[
Q_b = -2 \left\{(N + 1)/N \ln(N + 1) - 2\ln(2N) + \ln(N)\right\}^{-1}. \tag{6}
\]
where $N$ is the number of possible states of the system. Then, the disequilibrium $Q$ based on JS divergence can be described as
\[
Q_{JS}[P, P_e] = Q_b \mathcal{J}[P, P_e]. \tag{7}
\]
Finally, the statistical complexity measure based on JS divergence is as follows:
\[
C_{JS}[P] = Q_b \mathcal{J}[P, P_e]/H[P], \tag{8}
\]
where $H[P]$ denotes the normalized Shannon entropy calculated by
\[
H[P] = S[P]/S_{\text{max}}. \tag{9}
\]

In this paper, $H[P]$ is replaced by normalized dispersion entropy (NDE) as we discussed in Sec. II A.
FIG. 2. (a) Bifurcation diagram of the Logistic map. (b) Lyapunov exponent of the Logistic map.

FIG. 3. PCEC and DCEC of the Logistic map with different selected parameters $a$. (a) $H_r - C_r$, PCEC; (b) $H_r - C_r$, DCEC. The markers • and ▲ stand for starting points and ending points, respectively. The marker ■ stands for the points when $r$ or $q$ equals to 1.
C. Rényi entropy and Tsallis entropy

While Shannon entropy is a valuable tool for delineating system irregularities, it has limitations due to its boundedness, making it unsuitable for all cases. To address this constraint, two extensions of traditional Shannon entropy, namely, Rényi entropy\cite{18} and Tsallis entropy\cite{19}, have been introduced.

Rényi entropy builds upon the concept of substituting a quasi-linear mean for the linear one found in Shannon entropy. This approach has proven effective in solving various problems, including the measurement of quantum keys and serving as an index for side information. The \( r \)-th-order Rényi entropy of distribution \( P \) is defined as

\[
S_r[P] = \frac{1}{1-r} \ln \sum_{i=1}^{n} p_i^r, \quad (10)
\]

where \( r \) represents the entropic index. Notably, as \( r \to 1 \), Rényi entropy degenerates into Shannon entropy. The maximum value \( \ln N \) is attained when \( P = P_e \).

Tsallis entropy possesses the capability to identify nonlinear and multifractal structures when analyzing nonlinear time series. Tsallis \( q \)-entropy of distribution \( P \) is defined by

\[
S_q[P] = \sum_{i=1}^{n} p_i \log_q \frac{1}{p_i}, \quad (11)
\]

The entropic index \( q \) is commonly referred to as the bias parameter. A value of \( q < 1 \) expands the representation of rare events or states in distribution \( P \), while \( q > 1 \) manifests notable events or states in \( P \). This characteristic elucidates why Tsallis \( q \)-entropy can detect more information in nonlinear time series. An another representation of Tsallis \( q \)-entropy is given by

\[
S_q[P] = \frac{1 - \sum_{i=1}^{n} p_i^q}{q - 1}. \quad (12)
\]

where \( \log_q x = \frac{x^{1/q} - 1}{1/q} \) is the \( q \)-logarithm function. The maximum value \( \log_q N \) is attained when \( P = P_e \).

D. Dispersion complexity–entropy curves in two dimension and three dimension

In Secs. II A–II C, we explored dispersion entropy, statistical complexity measures, Rényi entropy\cite{18} and Tsallis entropy\cite{19}, focusing on their capacity to capture the irregularity and characteristics of nonlinear time series from various perspectives. To further uncover information and dynamic structures within time series, we integrate these four methods to formulate modified CEC across different dimensions. By following Ribeiro et al.\cite{24} and Jauregui et al.,\cite{25}

FIG. 4. PCEC and DCEC of the Logistic map with different selected parameters, (a) \( H_q - C_q \) PCEC; (b) \( H_q - C_q \) DCEC. The markers \( \bullet \) and \( \triangle \) stand for starting points and ending points, respectively. The marker \( ■ \) stands for the points when \( r \) or \( q \) equals to 1.
FIG. 5. The variation trend of $H_\ast^r$, $C_\ast^r$, $H_q^\text{min}$, and $C_q^\text{max}$ based on CEC as the logistic parameter $a$ goes up from 3 to 4. (a) $H_\ast^r$ and $C_\ast^r$ for PCEC; (b) $H_q^\text{min}$ and $C_q^\text{max}$ for PCEC; (c) $H_\ast^r$ and $C_\ast^r$ for DCEC; (d) $H_q^\text{min}$ and $C_q^\text{max}$ for DCEC.

The algorithms of Rényi DCEC and Tsallis DCEC are presented as follows.

Assuming that $P = \{p_1, p_2, \ldots, p_N\}$ is the patterns distribution of dispersion entropy, the definition of $H_r^r[P], H_q^q[P], C_r^r[P], C_q^q[P]$ based on dispersion entropy and different entropic index $r$ or $q$ are described as

$$H_r^r[P] = S_r[P]/S_r[P],$$

$$H_q^q[P] = S_q[P]/S_q[P],$$

$$C_r^r[P] = Q_r[P, P_e] H_r^r[P],$$

$$C_q^q[P] = Q_q[P, P_e] H_q^q[P],$$

where $H_r^r[P]$ and $H_q^q[P]$ represent the normalized dispersion entropy (NDE) of distribution $P$ based on Rényi entropy and Tsallis entropy, respectively. These values vary with the entropic indices $r$ and $q$, generating a list of entropy values. Similarly, we can compute a list of $C_r^r[P]$ and $C_q^q[P]$ values. In the equation, $Q_r[P, P_e]$ denotes the Rényi formulation of disequilibrium, expressed as $Q_r[P, P_e] = Q_{\text{ni}} J_r[P, P_e]$, where

$$J_r[P, P_e] = \frac{1}{2} K_r(P \| P_e) + \frac{1}{2} K_r(P_e \| P).$$

In this equation, $K_r(P \| Q)$ is the Rényi formulation of Kullback–Leibler divergence. And $Q_{\text{ni}}$ is the normalized constant calculated by

$$Q_{\text{ni}} = \left\{ \frac{1}{2(r-1)} \ln \left[ (N+1)^{1-r} + N-1 \frac{(N+1)^{1-r}}{4N} \right] \right\}.$$

Published under an exclusive license by AIP Publishing
Likewise, \( Q_q[P,P_e] \) is the Tsallis format of disequilibrium, described as
\[
Q_q[P,P_e] = Q_{q0} \mathcal{J}_q[P,P_e],
\]
where
\[
\mathcal{J}_q[P,P_e] = \frac{1}{2} K_q(P | (P + P_e)/2) + \frac{1}{2} K_q(P_e | (P + P_e)/2)
\]
\[
= \frac{1}{2} \sum_{i=1}^{N} p_i \log_i \left( \frac{P_i + 1/N}{2} \right)
\]
\[
= \frac{1}{2} \sum_{i=1}^{N} \frac{1}{N} \log_i \left( \frac{P_i + 1/N}{2} \right). \quad (19)
\]
In the equation, \( K_q(P | Q) = -\sum p_i \log_i(q_i/p_i) \) is the Tsallis formulation of Kullback-Leibler divergence. And \( Q_{q0} \) is the normalized constant calculated by
\[
Q_{q0} = \frac{(1-q)2^{2-q}N - (1+N)1^{1-q} - N(1+1/N)1^{-q} - N+1}{2^{2-q}N - (1+N)1^{1-q} - N(1+1/N)1^{-q} - N+1}. \quad (20)
\]

In summary, four variables \( H_r[P], H_q[P], C_r[P], \) and \( C_q[P] \) can be computed using the provided formula to illustrate the characteristics of nonlinear time series. Once these values are calculated for a given time series, we can select two or three of them to trace characteristic curves, observing their trajectories with varying values of \( r \) and \( q \). Drawing on the relationships between the variables, distinct nonlinear time series exhibit different curves, enabling the discrimination of stochastic processes, chaotic systems, and periodic systems, as elucidated in Sec. III.

In this study, we employ the \( H_r - C_r \) curve, \( H_q - C_q \) curve, \( H_r - H_q - C_q \) curve, \( H_r - H_q - C_r \) curve, and \( C_q - C_r \) curve under various conditions to analyze the time series. The block diagram of DCEC is illustrated in Fig. 1.

III. EXPERIMENTS IN SIMULATED DATA

In this section, we apply our modified DCEC to analyze the logistic map, color noises, and chaotic systems, showcasing their effectiveness in characterizing nonlinear time series features. Before the experiments, appropriate parameters should be selected carefully, including the number of classes \( c \) for DCEC, the embedding dimension \( m \), and the time delay \( t \).

The parameter \( c \) holds significant importance as it determines the configuration of dispersion patterns. A large \( c \) can result in longer computational times and sensitivity to noise due to the division across numerous dispersion patterns. Conversely, an overly small \( c \) might assign values from different amplitudes to the same pattern, leading to inaccuracies in practical applications. Hence, balancing the objectives of accuracy, noise reduction, and computational efficiency, we opt for \( c = 4 \) in the subsequent experiments.
Similar to the parameter \( c \), the embedding dimension \( m \) is also a crucial consideration. A large \( m \) may lead to prolonged computational times, although it may attain more accurate results. Conversely, if \( m \) is too small, it might fail to capture changes in dynamic structures within the time series in certain cases. Striking a balance between efficiency and accuracy, we select \( m = 4 \) for both PCEC and DCEC in the subsequent experiments.

Regarding the time delay \( t \) for PCEC and DCEC, previous research has proposed numerous optimization algorithms. In our study, we choose the widely used value of 1 as the setting for \( t \) in the experiments. When we construct the CEC, both the entropic index \( r \) and \( q \) range from 0 to 50 in a step length of 0.001.

A. The logistic map

The logistic map is a representative dynamical system that displays nonlinear complex behavior. Its formulation is given by

\[
x_n = ax_{n-1}(1 - x_{n-1}),
\]

where \( a \) serves as the control parameter, which can vary within the range of [3.4]. The mapping series exhibit diverse periods or manifest chaotic behavior to varying degrees for different values of \( a \). The bifurcation diagram in Fig. 2(a) visually captures the characteristics of the mapping series concerning the variations in the parameter \( a \).

For instance, when \( a = 3.05 \), the mapping displays a period of 2, while at \( a = 3.5 \), the period transforms into 8. Once the parameter \( a \) attains the value of 3.569 945 6..., the mapping starts exhibiting chaotic characteristics. In Fig. 2(b), the Lyapunov exponent of the logistic map is presented for different values of \( a \). A positive exponent indicates chaotic behavior, while a negative exponent signifies periodicity. This exponent serves as a valuable index for analyzing the characteristics of the logistic map.

Figure 3 displays the \( H_r - C_r - C_q \) PCEC and \( H_r - C_q \) DCEC associated with the logistic map for various selected parameters \( a \) from the set \( \{3.05, 3.5, 3.55, 3.593, 3.6, 3.7, 3.83, 3.84, 3.9, 3.95, 4\} \). To ensure adequate convergence of the sequences, we generate 10,000 points for each logistic parameter \( a \), starting with an initial value of 0.1. We then select the points from 5001 to 10,000 for experimentation. From the pictures, it is evident both PCEC and DCEC can effectively differentiate between periodic (\( a = 3.05, 3.5, 3.55, 3.83, 3.84 \)) and chaotic sequences (\( a = 3.593, 3.6, 3.7, 3.9, 3.95, 4 \)). The CEC for periodic sequences appear nearly vertical, while those for chaotic sequences exhibit a curved shape. Notably, as the logistic mapping transition into chaos for \( a = 3.593 \), both PCEC and DCEC prompt show curved shapes, contrasting with the vertical lines seen at \( a = 3.55 \). However, DCEC offers some distinct advantages. First, all DCECs for periodic sequences are situated in the lower left of the plane, clearly distinguishing them from chaotic
sequences. Second, as the chaotic degree increases, the DCEC for the six chaotic sequences \((a = 3.593, 3.6, 3.7, 3.9, 3.95, 4)\) progressively shifts toward the upper right in Fig. 3(b), with their shapes gradually becoming more curved. When the sequence is entirely chaotic \((a = 4)\), the starting and ending points of the curve reach the same altitude. These observations demonstrate that DCEC effectively discerns the diverse characteristics of sequences across different parameters \(a\) in the logistic mapping. Similar results are observed in Fig. 4. Compared to PCEC in Fig. 4(a), DCEC in Fig. 4(b) more clearly distinguishes periodic and chaotic sequences, with the curves of periodic sequences appearing at the top and the curves of chaotic sequences orderly descending with increasing chaotic degree. In summary, DCEC proves to be a stable and effective method in various aspects, accurately and robustly detecting the characteristics of nonlinear time series.

To be more precise, four feature values including \(H^* r\), \(C^* r\), \(H q_{\text{min}}\), and \(C q_{\text{max}}\) are selected to assess our method. \(H^* r\) and \(C^* r\) are the normalized entropy and statistic complexity, respectively, when \(r = 50\). Additionally, \(H q_{\text{min}}\) is the minimum value of normalized entropy, and \(C q_{\text{max}}\) is the maximum value of statistical complexity within the curves. Figure 5 shows the variation trend of these four values based on PCEC and DCEC, respectively, as the logistic parameter \(a\) goes up from 3 to 4. Notably, in Figs. 5(a) and 5(b), it is obvious that when \(a < 3.5699456\), which means the series is periodic, all four features based on PCEC exhibit pronounced fluctuations with the changing parameter \(a\). In contrast, in Figs. 5(c) and 5(d), for DCEC, \(H^* r\) and \(C^* r\) remain at low values with minimal change. For \(H q_{\text{min}}\) and \(C q_{\text{max}}\), slight variations are observed when \(a\) goes up from 3 to 3.5699456, followed by regular changes when \(a > 3.5699456\), coinciding with the series transformed into chaos. To provide further detail, Table I displays the Kendall rank correlation between these four values \((H^* r, C^* r, H q_{\text{min}}, \text{and} C q_{\text{max}})\) and Lyapunov exponent depicted in Fig. 2(b) for both PCEC and DCEC. The numerical results indicate that all four feature values in DCEC exhibit a stronger correlation with the Lyapunov exponent. In conclusion, when examining the logistic map, the features derived from DCEC not only exhibit greater robustness but also better capture the dynamic characteristics inherent in the original nonlinear time series compared to PCEC.

**Table I.** The Kendall rank correlation between four feature values \((H^* r, C^* r, H q_{\text{min}}, \text{and} C q_{\text{max}})\) and Lyapunov exponent. Bold represents the one that has a stronger correlation with the Lyapunov exponent.

<table>
<thead>
<tr>
<th></th>
<th>(H^* r)</th>
<th>(C^* r)</th>
<th>(H q_{\text{min}})</th>
<th>(C q_{\text{max}})</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCEC</td>
<td>0.4776</td>
<td>0.4518</td>
<td>0.7130</td>
<td>-0.7105</td>
</tr>
<tr>
<td>DCEC</td>
<td><strong>0.7251</strong></td>
<td><strong>0.7229</strong></td>
<td><strong>0.7671</strong></td>
<td><strong>-0.7621</strong></td>
</tr>
</tbody>
</table>
B. The color noises

Noise, a form of stochastic fluctuation, is a central focus in nonlinear dynamics research. Among various types, white noise stands out as the most familiar, given its ubiquitous generation. Beyond white noise, there exist different color noises characterized by the power spectrum of the original signal. The definitions of white noise, blue noise, pink noise, and red noise based on power at a constant bandwidth are provided in Table II. White noise, with its maximum randomness, exhibits a uniform power spectrum density. Blue noise finds extensive application in computer sampling techniques, featuring a power spectral density proportional to frequency, concentrating energy at high frequencies. On the other hand, red noise and pink noise share similar characteristics in the frequency domain, with their power spectrum density inversely proportional to frequency. The majority of their energy is concentrated at low frequencies, with red noise being particularly pronounced in this regard.

We utilize both PCEC and DCEC methodologies to scrutinize the distinct structures of four noise types and draw comparisons. In this experiment, all noise sequence lengths are set to 5000. In Fig. 6, \( H_r - C_r \), PCEC, \( H_q - C_q \) DCEC, \( H_r - C_q \) PCEC, and \( H_q - C_q \) DCEC for these four noise types are depicted. In Figs. 6(a) and 6(c), both the \( H_r - C_r \), PCEC and \( H_q - C_q \) PCEC for the four noises exhibit similar variations. The \( H_r - C_r \) curves extend from the bottom right to the top left, while the \( H_q - C_q \) curves form loops. In contrast, the DCEC plots for the four noises in Figs. 6(b) and 6(d) display noticeably different trends. In Fig. 6(b), the DCEC \( H_q - C_q \) curves for white noise and blue noise resemble those in PCEC [Fig. 6(a)]. However, the curves for pink noise and red noise exhibit distinct shapes and trends compared to white noise and blue noise. Regarding the DCEC \( H_q - C_q \) curves in Fig. 6(b), the curves for white noise form nearly loop-like patterns, with both start and end points at (1,0), while pink noise and red noise show a turn followed by a surge in value, and red noise initially follows a straight path. To provide clearer results, three-dimensional representations of PCEC and DCEC in spaces indexed by \( H_r - H_q - C_q \) and \( H_r - H_q - C_r \) are presented in Fig. 7. All three axes range from [0,1], constructing a three-dimensional complexity–entropy space. The results mirror those in Fig. 6, as PCEC for the four noises in Figs. 7(a) and 7(c) exhibit similar trends, especially for blue noise and pink noise. However, DCEC in Figs. 7(b) and 7(d) effectively differentiate between the four types of noises based on their distinct trends and shapes. Overall, the DCEC method proves to be a valuable tool for distinguishing noises with varying power spectral density, highlighting its capability for detecting nonlinear time series structures.

C. Chaotic systems

Analyzing consecutive complex systems through discretization and employing nonlinear time series analysis constitutes a common area of investigation within the dynamic systems field. In this
section, we utilize DCEC to examine various chaotic systems, specifically, the logistic map, Hénon map, Chen system, Duffing system, Lorenz system, and Rossler system. Our aim is to observe similarities and differences in the curves of these systems, shedding light on the underlying structural relationships. To further enrich our analysis, we introduce 1/f noise to draw comparisons between chaotic systems and stochastic systems. The length of all series is set at 50,001, and for the four consecutive chaotic systems (Chen, Duffing, Lorenz, and Rossler), we discretize time from 0 to 500 with a step size of 0.01. The corresponding equations of these systems are provided below:

1. Logistic map:
   \[
   x_n = ax_{n-1}(1 - x_{n-1}).
   \]
   Parameter value: \(a = 4\); initial condition: \(x_0 = 0.1\).

2. Hénon map:
   \[
   \begin{align*}
   x_n &= 1 + y_{n-1} - ax_{n-1}, \\
   y_n &= bx_{n-1}.
   \end{align*}
   \] (23)
   Parameter values: \(a = 1.4\), \(b = 0.3\); initial conditions: \(x_0 = 0.1\), \(y_0 = 0.1\).

3. Chen system:
   \[
   \begin{align*}
   \dot{x} &= a(y - x), \\
   \dot{y} &= (c - a)x - xz + cy, \\
   \dot{z} &= xy - bz.
   \end{align*}
   \] (24)
   Parameter values: \(a = 35\), \(b = 3\), \(c = 28\); initial conditions: \(x_0 = 0\), \(y_0 = 1.001\), \(z_0 = 0\).

4. Duffing system:
   \[
   \begin{align*}
   \dot{x} &= y, \\
   \dot{y} &= x - x^3 - ky + z, \\
   \dot{z} &= -fsin(ft).
   \end{align*}
   \] (25)
   Parameter values: \(k = 1.15\), \(f = 1\); initial conditions: \(x_0 = 0\), \(y_0 = 0\), \(z_0 = 1\).

5. Lorenz system:
   \[
   \begin{align*}
   \dot{x} &= p(y - x), \\
   \dot{y} &=rx - y - xz, \\
   \dot{z} &= xy - bz.
   \end{align*}
   \] (26)
   Parameter values: \(p = 0.2\), \(r = 10\), \(b = 8/3\); initial conditions: \(x_0 = 0\), \(y_0 = 0\), \(z_0 = 0\).

The corresponding equations of these systems are provided below:
Parameter values: \( p = 10, \ r = 28, \ b = 8/3 \); initial conditions: \( x_0 = 10, \ y_0 = 0, \ z_0 = 1 \).

(6) Rossler system:

\[
\begin{align*}
\dot{x} &= -y - z \\
\dot{y} &= x + ay \\
\dot{z} &= b + z(x - c).
\end{align*}
\]  

Parameter values: \( a = 0.2, \ b = 0.4, \ c = 5.7 \); initial conditions: \( x_0 = 1, \ y_0 = 0, \ z_0 = 0 \).

Figure 8 illustrates \( H_r - H_q - C_q, \ H_r - H_q - C_r, \) and \( C_q - C_r \) DCEC of six chaotic systems alongside 1/f noise. Noteworthy findings and conclusions have emerged from the analysis. In both Figs. 8(a) and 8(b), the six chaotic systems display similar trajectories in three-dimensional space, varying with the parameters \( r \) and \( q \). However, these trajectories exhibit distinct shapes when compared to the trajectory of 1/f noise. This successful differentiation in the visual representation suggests an effective distinction between chaotic systems and noise. More crucially, Fig. 8 showcases both the commonalities and distinctions among the six chaotic systems. The DCEC of the six chaotic systems exhibit analogous shapes and trends. Nonetheless, their inflection points differ, facilitating a more detailed recognition and clustering. For instance, as depicted in Fig. 8(a), the \( H_r - H_q - C_q \) curves of Duffing, Lorenz, and Rossler systems closely resemble each other, implying similar structures and classification into the same type. Similarly, Logistic and Hénon system cluster together as another type. The curve of the Chen system, positioned between these two types, stands out due to its unique structure when compared with the remaining five systems. This approach enables the differentiation of the six chaotic systems into three distinct types. Furthermore, in comparison with the four consecutive systems (Chen, Duffing, Lorenz, and Rossler), the logistic map and Hénon map have more complex structures for their \( C \) values are higher than others. However, their entropy value \( H \) remains nearly identical. In contrast, 1/f noise displays a high entropy value and low complexity, characteristic of its random and simplistic structure. Figure 8(c) illustrates the complexity plane, revealing the nonlinear relationship between \( C_q \) and \( C_r \). Notably, the curves of Chen, Duffing, Lorenz, and Rossler systems trend downward, while those of the Logistic and Hénon maps trend upward, highlighting their distinct structures. Additionally, it is observed that at higher values of \( r \) and \( q \), the complexity of 1/f noise resembles that of chaotic systems at lower \( r \) and \( q \). This observation suggests potential relationships between noises and chaotic systems.

To comprehensively explore the characteristics, we utilize the multiscale time series method \(^{45}\) for further analysis. In Fig. 9, we present \( H_q \) min and \( C_q \) max values of six chaotic systems and 1/f noise across different scales. Notably, as the time scale increases from 1 to 10, entropy values \( H \) exhibit growth, while complexity values \( C \) decline. However, distinct systems demonstrate varying levels of sensitivity to the coarse-grained scale. It is evident that 1/f noise stands out as the most robust, whereas both the logistic map and
Hénon map display heightened sensitivity. This observation suggests that despite the more intricate structures of logistic map and Hénon map, they lack the stability observed in other chaotic systems.

IV. APPLICATIONS IN REAL-WORLD DATA

A. Rolling bearing fault diagnosis

In the field of engineering, rolling bearings play a pivotal role as a crucial component in rotating machinery. Despite their widespread use, these bearings frequently experience breakdowns in various components due to prolonged exposure to high-pressure and high-speed working conditions. Statistical data indicate that a substantial portion, approximately 40%–50%, of mechanical failures can be attributed to issues with rolling bearings. Given this prevalence, it becomes imperative to develop accurate methods for diagnosing faults in rolling bearings. The primary challenge in rolling bearing fault diagnosis lies in determining the specific fault type based on the received signal. In essence, the goal is to identify which part of the rolling bearing is malfunctioning. This task holds significant importance for ensuring the reliability and performance of rotating machinery.

In this section, we use the detected bearing data from the Case Western Reserve University. The dataset comprises information collected from normal bearings as well as bearings with single-point defects at the drive end and fan end. The sample rate is categorized into two frequencies, namely, 12 and 48 kHz, with a consistent rotating speed of approximately 1750 rpm across all cases. The fault points are characterized by diameters of 0.007 in., 0.014 in., 0.021 in., and 0.028 in. Due to some instances of missing data in the dataset, our analysis focuses specifically on the normal type and five distinct fault types. Specifically, we consider the fault data of 12 kHz frequency at the drive end with a damage diameter of 0.007 in. The relevant details are summarized in Table III. Each signal’s length exceeds 100,000 points, and for experimental purposes, we select the initial 100,000 points for analysis.

In prior studies, methods such as empirical mode decomposition (EMD), ensemble empirical mode decomposition EEMD, variational mode decomposition (VMD), and other signal

<table>
<thead>
<tr>
<th>Fault type</th>
<th>Code</th>
<th>Damage diameter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>NOR</td>
<td>0.007 in.</td>
</tr>
<tr>
<td>Ball fault</td>
<td>BA</td>
<td>0.007 in.</td>
</tr>
<tr>
<td>Inner Race fault</td>
<td>IR</td>
<td>0.007 in.</td>
</tr>
<tr>
<td>Outer Race 3 o’clock fault</td>
<td>OR3</td>
<td>0.007 in.</td>
</tr>
<tr>
<td>Outer Race 6 o’clock fault</td>
<td>OR6</td>
<td>0.007 in.</td>
</tr>
<tr>
<td>Outer Race 12 o’clock fault</td>
<td>OR12</td>
<td>0.007 in.</td>
</tr>
</tbody>
</table>

FIG. 12. The classification result of 180 fault bearing samples based on different training indexes. (a) $H_q \min$, (b) $C_q \max$, (c) $C_{-\text{index}}$, (d) $H_q \min$ and $C_q \max$; (e) $H_q \min$, $C_q \max$, and $C_{-\text{index}}$. 

Published under an exclusive license by AIP Publishing
decomposition approaches have commonly been used to extract features from bearing signals for fault diagnosis. However, these methods rely on intricate theories, resulting in high computing times. Additionally, determining the number of intrinsic mode functions (IMF) poses a challenge. Entropy-based methods have also been employed for diagnosis, but they often fall short of achieving the anticipated effectiveness. In our research, we introduce a novel method for diagnosing bearing faults. We extract features from DCEC, as discussed in Secs. I–III. The advantage of this method lies in its consideration not only of entropy value but also of complexity value, which can depict the varied inner structures of the signal. Figure 10 illustrates \(H_q - H_q - C_q\) and \(H_q - H_q - C_q\) DCEC for six bearing types. These figures reveal that normal bearings exhibit the highest complexity degree and the lowest entropy value, indicating a more complex structure compared to damaged circumstances.

To differentiate between the six types, we extract \(H_q\) min and \(C_q\) max from the curves, marked with red and blue stars in Figs. 10(a) and 10(b) respectively. These two indexes serve as feature values for different bearing types. However, a limitation is observed as both indexes struggle to distinguish NOR and OR3 types, potentially leading to diagnostic inaccuracies. To address this issue, we introduce another index based on \(C_q\) – \(C_q\) DCEC, termed the C-index. Figure 11 displays curves for the six bearings, and the C-index is defined by the distance between the lowest \(C_q\) point and the highest \(C_q\) point in the \(C_q\) – \(C_q\) DCEC. This index is marked by dotted lines. Notably, the curve for the normal bearing tends to form a loop when compared to the curves of the other five fault-bearing types.

Using these three indexes, various types of faults can be effectively distinguished. To demonstrate the efficacy of our method, we randomly select 30 samples of length 10,000 for each type from signals, resulting in a total of 180 training data. The values of the training data, based on different combinations of indexes, are presented in Figs. 12(a)–12(c). In Figs. 12(a), 12(b), and 12(d), it is evident that relying solely on the observation of \(H_q\) min and \(C_q\) max can lead to confusion, particularly for NOR and OR3 types. However, the C-index value in Fig. 12(c) proves effective in distinguishing them. Therefore, in Fig. 12(e), we combine these three indexes as features to classify the fault bearings, successfully clustering them into distinct types.

Subsequently, we employ the multivariate support vector machine (MSVM) to train models based on different indexes and make comparisons. MSVM, a valuable tool in the machine-learning domain, performs well in multi-classification tasks. For classification experiments, we randomly select 20 test samples from each type, totaling 120 samples. The classification results and accuracy based on different combinations of indexes are presented in Fig. 13 and summarized in Table IV. The results indicate that the C-index extracted from \(C_q - C_q\) curves exhibits higher accuracy than \(H_q\) min and \(C_q\) max. Furthermore, when we combine the three indexes as features, the accuracy reaches 100%. This strongly attests to the method’s clarity and effectiveness in diagnosing fault bearings.

### B. The stock market indices

Financial time series analysis is a hot research topic that aims at detecting the regularities hidden in the variations of stock indices. Numerous researchers have conducted extensive investigations in this particular field. Through theories in mathematics and methods of signal processing, the relations or distinctions between different stock markets can be discovered. In this section, complexity–entropy curves are used to detect the change of inner structure in financial markets along with the time period. We choose three Chinese indices, three American indices, and three Europe indices. The countries and Bloomberg codes of nine indices are listed in Table V. For each stock indices, three-period (Jan 1, 2000–Dec 31, 2004; Jan 1, 2007–Dec 31, 2011; Jan 1, 2014–Dec 31, 2018) financial time series are selected and the log-return of each day can be calculated by

\[
x_n = \log(S_n) - \log(S_{n-1}).
\]

The \(H_q - H_q - C_q\) DCEC and the \(C_q\) max values under different scales of nine stock market indices are depicted in Figs. 14 and 15, respectively. From these illustrations, interesting conclusions can be drawn. First, the \(C_q\) values of the nine indices exhibit a consistent upward trend over time, as evident in Fig. 14. This suggests a gradual increase in the complexity of financial markets worldwide over the past 20 years. Furthermore, a more detailed analysis of Figs. 14 and 15 reveals that the complexity levels of three American indices (S&P500, DJI, NASDAQ) were higher than those of other regions in the second period (2007–2011), with the S&P500 index showing particularly elevated complexity. This corresponds to the occurrence of the financial crisis in the United States during

<table>
<thead>
<tr>
<th>Country</th>
<th>Bloomberg code</th>
<th>Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>China</td>
<td>SSE</td>
</tr>
<tr>
<td>2</td>
<td>China</td>
<td>SZI</td>
</tr>
<tr>
<td>3</td>
<td>China</td>
<td>HSI</td>
</tr>
<tr>
<td>4</td>
<td>US</td>
<td>S&amp;P500</td>
</tr>
<tr>
<td>5</td>
<td>US</td>
<td>DJI</td>
</tr>
<tr>
<td>6</td>
<td>US</td>
<td>NASDAQ</td>
</tr>
<tr>
<td>7</td>
<td>France (EU)</td>
<td>CAC</td>
</tr>
<tr>
<td>8</td>
<td>Germany (EU)</td>
<td>DAX</td>
</tr>
<tr>
<td>9</td>
<td>UK (EU)</td>
<td>FTSE</td>
</tr>
</tbody>
</table>
FIG. 13. The predicted result of 120 fault bearing test samples based on different training indexes. (a) $H_q$ min; (b) $C_q$ max; (c) $C$ index; (d) $H_q$ min and $C_q$ max; (e) $H_q$ min, $C_q$ max, and $C$ index.

that period, which significantly impacted stock indices. Three European indices (CAC, DAX, and FTSE) and HSI were also influenced by the crisis for the $C_q$ values increase to a certain extent. Three European indices and HSI were also influenced by the crisis, leading to an increase in $C_q$ values to a certain extent. However, the complexity degree of SSE and SZI remained almost unchanged during this period compared to the first period (2000–2004), indicating that the financial crisis did not exert a substantial influence on these two Chinese indices. Simultaneously, this observation underscores the resemblance between the HSI index and the financial markets in Europe and America. Hence, DCEC serves to highlight and distinguish the impact of the financial crisis on different regions. Finally, focusing on the multiscale results in Fig. 15, it becomes apparent that $C_q$ max values of stock indices in different regions gradually diverge over time, particularly noticeable in the third period (2014–2018).

Additionally, SSE and SZI exhibit a high level of complexity during this period. These findings suggest that global financial markets have become increasingly diverse over time. In conclusion, statistical complexity–entropy curves prove to be a valuable tool for detecting the structures and regularities of financial markets, serving as an effective means for analyzing financial time series.

V. CONCLUSION

In this study, we introduce the concept of DCEC as a novel approach for characterizing nonlinear time series with diverse structures. This methodology stands out from its predecessors, offering a more distinct segmentation of probability states. Our validation of this method, employing both simulated and real-world data, confirms its efficacy.

Our experiments commence with simulated data. Initially, we employed a logistic map to compare the outcomes derived from both DCEC and the previously established PCEC. Analysis through $H_r - C_r$ and $H_q - C_q$ curves reveals effective ability of DCEC to distinguish periodic sequences and chaotic sequences with different degrees. Moreover, the features extracted from DCEC demonstrates robustness to variations in logistic parameters and exhibit a stronger correlation with the Lyapunov exponent. Subsequently, four types of color noises are utilized for experiments. Both two-dimensional $H_r - C_r$, $H_q - C_q$, and $H_r - H_q - C_q$ and $H_r - H_q - C_q$ underscore the ability of DCEC to distinguish noises with varying power spectral density. Furthermore, we employed DCEC for the analysis of diverse chaotic systems and 1/f noise using $H_r - H_q - C_r$, $H_r - H_q - C_q$, and $C_q - C_r$ curves. This approach effectively delineated the relationships and distinctions among various chaotic systems. It uniquely identified 1/f noise through distinct curve shapes, thereby highlighting our method’s capability in discerning between chaotic systems and stochastic processes.

To demonstrate the practical applicability of DCEC, we integrated it into real-world scenarios, specifically in fault-bearing diagnosis and financial time series analysis. In fault bearing diagnosis, we
innovated a method to categorize fault types using selected indices from \( H_r - H_\alpha - \xi_r \) and \( H_\alpha - H_q - \xi_q \). Employing these indices to train the multivariate support vector machine (MSVM) model significantly improved accuracy rates. For financial time series, we analyzed stock indices from various countries over different periods. The examination of DCEC and \( C_p \) max values across multiple scales shed light on hidden patterns and distinct structures in these financial markets.

In summary, DCECs prove to be a highly effective tool for analyzing the structures of various nonlinear time series and distinguishing them. This approach is also applicable to addressing challenges in numerous other research domains. We trust that the methods presented in this paper can provide assistance and serve as references for future research endeavors.

**ACKNOWLEDGMENTS**

The financial support from the funds for the National Natural Science Foundation of China (NNSFC) (No. 62171018) is gratefully acknowledged.

**AUTHOR DECLARATIONS**

**Conflict of Interest**

The authors have no conflicts to disclose.

**Author Contributions**

Runze Jiang: Conceptualization (equal); Data curation (equal); Formal analysis (equal); Methodology (equal); Software (equal); Validation (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal). Pengjian Shang: Formal analysis (equal); Funding acquisition (supporting); Methodology (equal); Visualization (equal); Writing – review & editing (equal).

**DATA AVAILABILITY**

The data that support the findings of this study are openly available in Case Western Reserve University bearing data center at https://engineering.case.edu/bearingdatacenter/apparatus-and-procedures, Ref. 46 and available from the corresponding author upon reasonable request.

**REFERENCES**


