Dispersion network-transition entropy: A metric for characterizing the complexity of nonlinear signals

Bo Geng,[1,2,*] Haiyan Wang,[1,3,†] Xiaohong Shen,[1,2,‡] Hongwei Zhang,[1,2,§] and Yongsheng Yan[1,2,‖]
[1]*School of Marine Science and Technology, Northwestern Polytechnical University, Xi'an, Shaanxi 710072, China*
[2]*Key Laboratory of Ocean Acoustics and Sensing, Northwestern Polytechnical University, Ministry of Industry and Information Technology, Xi'an, Shaanxi 710072, China*
[3]*School of Electronic Information and Artificial Intelligence, Shaanxi University of Science and Technology, Xi'an, Shaanxi 710021, China*

(Received 13 March 2024; accepted 10 July 2024; published 12 August 2024)

Extracting meaningful information from signals has always been a challenge. Due to the influence of environmental noise, collected signals often exhibit nonlinear characteristics, rendering traditional metrics inadequate in capturing the dynamic properties and complex structures of signals. To address this challenge, this study proposes an innovative metric for quantifying signal complexity—dispersion network-transition entropy (DNTE), which integrates the concepts of complex networks and information entropy. Specifically, we assign single cumulative distribution function values to network nodes and utilize Markov chains to represent links, transforming nonlinear signals into weighted directed complex networks. Subsequently, we assess the importance of network nodes and links, and employ the mathematical expression of information entropy to calculate the DNTE value, quantifying the complexity of the original signal. Next, through extensive experiments on simulated chaotic models and real underwater acoustic signals, we confirm the outstanding performance of DNTE. The results indicate that, compared to Lempel-Ziv complexity, permutation entropy, and dispersion entropy, DNTE not only more accurately reflects changes in signal complexity but also exhibits higher computational efficiency. Importantly, DNTE demonstrates optimal performance in distinguishing different categories of chaotic models, ships, and modulation signals, showcasing its significant potential in extracting effective information from signals.

DOI: 10.1103/PhysRevE.110.024205

I. INTRODUCTION

In this age of information technology, the effective acquisition of information plays a pivotal role in various fields. By processing the collected signals, we can extract valuable information from complex data and address various practical problems. The development of signal processing not only drives the advancement of science and technology but also profoundly influences our daily lives, spanning multiple fields such as industrial engineering [1], economics [2], and hydroacoustics [3]. The core of signal processing lies in feature extraction [4]. However, influenced by various factors such as complex environments and equipment self-noise, collected signals often exhibit nonlinear characteristics [5], making traditional feature extraction methods less advantageous in processing real measured signals. Therefore, it is imperative to apply new features that can effectively characterize nonlinear signals.

Nonlinear dynamical features are used to describe the characteristics of nonlinear behavior within dynamical systems, and many scholars have applied them in the application of real measured signals [6–9]. Commonly employed nonlinear dynamical features include Lempel-Ziv complexity (LZC) [10], the Lyapunov exponent [11], and entropy [12]. Specifically, LZC can characterize the rate at which new patterns emerge in a signal, whereas entropy quantifies the uncertainty of the signal. In contrast to the high computational complexity associated with Lyapunov exponent calculation, LZC and entropy not only have lower computational complexity but also vary with the complexity of the signal. Specifically, as the signal becomes more complex, both entropy and LZC values tend to increase [13].

Complex networks are another commonly used method of extracting information from signals. They define nodes and links in a specific way, where nodes represent individuals or elements in the system, and links in the network represent connections or relationships between nodes [14–16]. Unlike metrics such as LZC and entropy, which directly act on signals, complex networks calculate various relationships between nodes and links, and the obtained features can be used to describe the complex structures and dynamic behaviors of various systems. Therefore, complex networks are widely applied in fields such as mechanical engineering, biomedicine, and hydroacoustics [17–19].

*Contact author: bougen@mail.nwpu.edu.cn
†Contact author: hywang@sust.edu.cn
‡Contact author: xhshen@nwpu.edu.cn
§Contact author: zhanghongwei@mail.nwpu.edu.cn
‖Contact author: ysyan@nwpu.edu.cn

Currently, commonly used complex networks include the visibility graph (VG) [20], recurrence network (RN) [21], and transition network (TN) [22]. In VG, each sampling point in the signal data is regarded as a node, and they are connected based on visibility, offering the advantage of not requiring parameter selection [23]. On the other hand, both RN and TN use the embedding dimension to segment signal data as different nodes. However, RN establishes links through similarity calculations, whereas TN considers adjacent nodes as connected links, thereby preserving temporal causality [24].

However, the aforementioned complex networks also face certain issues. For instance, VG encounters potential data loss and exhibits limited applicability to large datasets [24]. Incorrect threshold selection in RN can significantly impact the network, and during network transformation [25], TN may lead to the loss of effective information in the original signal due to the settings of embedding dimension and time delay. To illustrate, consider the pattern π_{123}; based on the temporal sequence, the next pattern can only be π_{23x}. While setting a larger time delay can mitigate this phenomenon, it may also affect subsequent patterns. In addition, existing complex network metrics, such as the global clustering coefficient [26], network transitivity [27], and average path length [28], although capable of providing a profound understanding of the overall structure, properties, and functionality of networks [29], still have certain limitations in quantifying network irregularities.

In response to the existing issues, we propose a unique solution that combines the concepts of complex networks and information entropy, leveraging the advantages of both to characterize the complexity of nonlinear signals from the perspective of network information distribution. Specifically, inspired by the dispersion pattern [30], we initially use the values of the cumulative distribution function after scaling as network nodes. Following the principles of Markov chains, we treat the relationships between connected nodes as links, thus introducing the concept of the dispersion network (DN). Subsequently, we calculate the distribution probabilities of each node and link in the DN and use the Shannon entropy formula to obtain the DN-transition entropy (DNTE). While constructing DN, we do not employ embedding dimensions for pattern segmentation of nodes to reduce information redundancy. In the computation of TE, we also simultaneously consider the importance of both network nodes and links. Therefore, compared with other metrics, DNTE exhibits more stable and accurate characteristics in characterizing signal complexity. Subsequently, experimental validations are conducted on simulated chaotic models and two types of real measured hydroacoustic signal datasets, further confirming the outstanding performance of DNTE in nonlinear signal processing.

The main contribution of this paper is the introduction of an alternative complexity metric, DNTE, which demonstrates excellent performance in nonlinear signal complexity characterization. Furthermore, the structure of this paper is as follows: Section II progressively introduces the theoretical steps of DNTE and discusses its parameters. Section III uses simulation experiments to compare DNTE's ability to detect dynamic changes and differentiate chaotic models, while also evaluating its computational cost. In Sec. IV, the practical application capabilities of DNTE are validated using two types of real measured hydroacoustic datasets. Section V summarizes the entire paper.

II. THEORY

A. Dispersion network

Similar to other complex networks, the proposed DN typically involves two fundamental steps: the initial determination of nodes and the subsequent establishment of links among these nodes. The distinctive feature of DN lies in the requirement for the preprocessing of temporal information before node determination, which is a crucial step in the construction process. Specifically, for a time series X of length L, where $X = \{x_1, x_2, \ldots, x_L\}$, the construction process of the DN is as follows:

Step 1: Through mapping using a normal cumulative distribution, the original sequence is transformed into a new sequence composed of cumulative distribution function values, reducing the impact of the original values' scale and range. The mapping using normal cumulative distribution can be represented as follows [30]:

$$y_k = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x_k} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt, \tag{1}$$

where σ and μ are the standard deviation and mean of the initial sequence X, respectively. We can obtain a new sequence $Y = \{y_1, y_2, \ldots, y_L\}$, which consists of cumulative distribution function values.

Step 2: Introduce the class c and perform round function mapping on the normalized sequence Y. The cumulative distribution function values in Y are further converted into a new sequence Z, which consists of integers from 1 to c.

$$z_k = |y_k \times c + 0.5|. \tag{2}$$

Here, $|\cdots|$ denotes rounding the elements within the brackets to the nearest integer. The resulting z_k can be considered as a node in the network, denoted as $\pi_i (i = z_k)$, with a maximum of c classes of nodes.

Step 3: Count the links between adjacent nodes in chronological order and denote them as $\pi_i \to \pi_j$, where i and j are both within the range of $1 - c$, with a total of c^2 classes of possible links. Note that self-loops are allowed, meaning situations where the preceding and succeeding nodes are the same, as self-loops also contain valuable information. By counting the occurrences of π_i and $\pi_i \to \pi_j$, i, $j \in [1, c]$, we can obtain the constructed DN, represented as $[G^c, W^c]$.

$$G^c = [N_{\pi_1}, N_{\pi_2}, \ldots, N_{\pi_c}], \tag{3}$$

$$W^c = \begin{bmatrix} N_{(\pi_1\to\pi_1)} & N_{(\pi_1\to\pi_2)} & \cdots & N_{(\pi_1\to\pi_c)} \\ \cdots & \cdots & \cdots & \cdots \\ N_{(\pi_i\to\pi_1)} & N_{(\pi_i\to\pi_2)} & \cdots & N_{(\pi_i\to\pi_c)} \\ \cdots & \cdots & \cdots & \cdots \\ N_{(\pi_c\to\pi_1)} & N_{(\pi_c\to\pi_2)} & \cdots & N_{(\pi_c\to\pi_c)} \end{bmatrix}, \tag{4}$$

where N_{π_i} and $(\pi_i \to \pi_j)$ are the number of corresponding nodes and connected links, respectively. It is important to note that the constructed DN is a directed network, and $N_{(\pi_i\to\pi_j)}$ is usually not equivalent to $N_{(\pi_j\to\pi_i)} (i \neq j)$.

Figure 1 illustrates the time series and the corresponding DNs (with c set to 10) constructed by the logistic map under

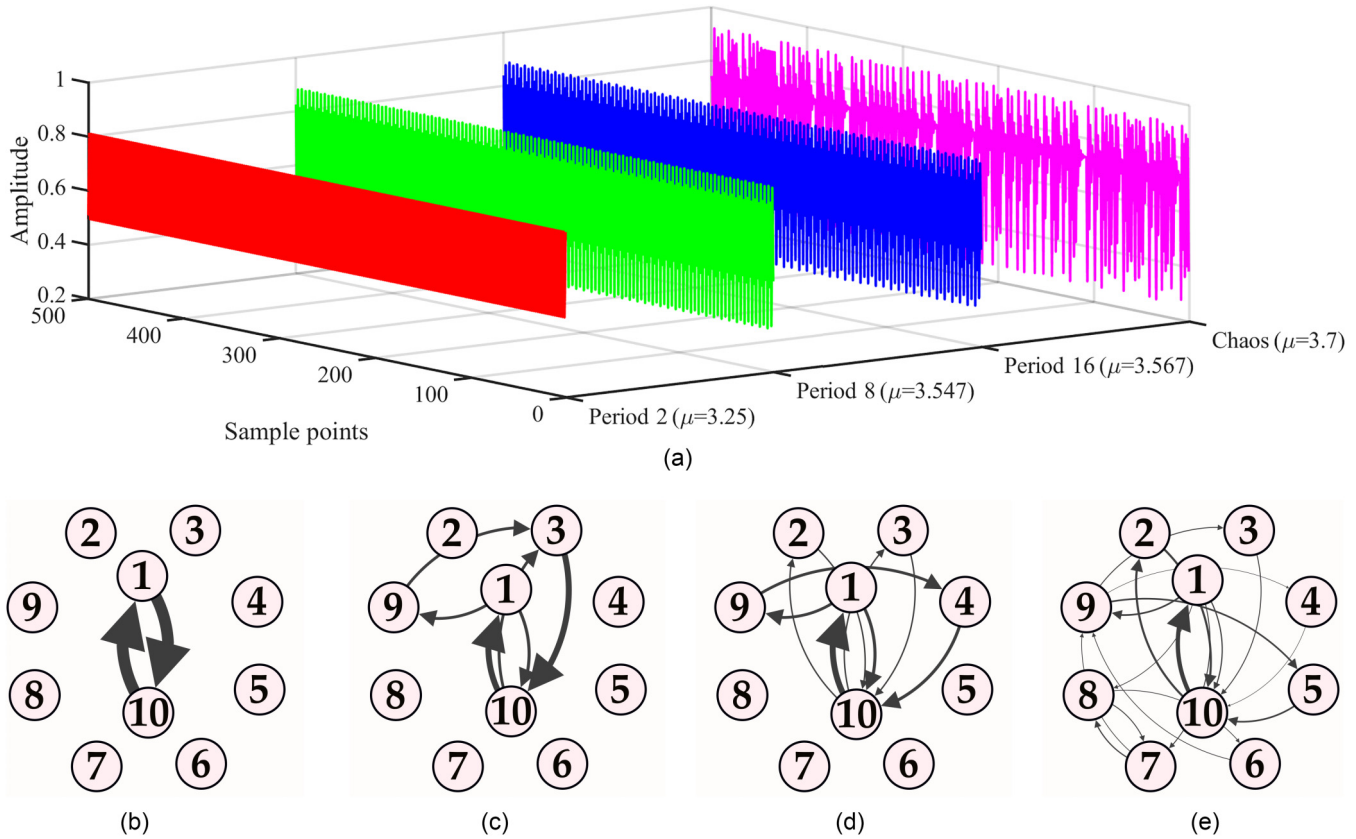


FIG. 1. (a) Time series of logistic map under different bifurcation parameters μ, (b) DN constructed from period 2 ($\mu = 3.25$), (c) DN constructed from period 8 ($\mu = 3.547$), (d) DN constructed from period 16 ($\mu = 3.567$), and (e) DN constructed from the chaotic state ($\mu = 3.7$).

different bifurcation parameters μ, including $\mu = 3.25, 3.547, 3.567$, and 3.7, which correspond to period 2, period 8, period 16, and chaotic states, respectively, and more details about the logistic map can be seen in Sec. III A. From Fig. 1, it can be observed that under the period-2 solution, only nodes 1 and 10 have links, accompanied by the same thickness. However, with an increase in the irregularity of dynamical regimes, both the number of node classes and links increase, which is also reflected in the thickness of each link, demonstrating that networks generated by different dynamical regimes can exhibit distinct characteristics.

B. Transition entropy

Entropy is a fundamental concept in information theory and is frequently employed to measure the intricacy and unpredictability of time series [30,31]. Building upon the DN derived from Eqs. (3) and (4), we introduce the calculation procedure for transition entropy (TE) to obtain the DNTE value, which helps in quantifying the complexity of the obtained DN.

DNTE simultaneously considers both node and link information in the network, providing an effective measure of the overall complexity of the network. It can be used as a nonlinear metric to apply to complex signal applications. Specifically, for the constructed network $[G^c, W^c]$, DNTE involves two steps:

Step 1: Calculate the probabilities P_{π_i} and $P_{\pi_i \to \pi_j}$ for each class of nodes and each class of links, respectively:

$$P_{\pi_i} = \frac{N_{\pi_i}}{L}, \tag{5}$$

$$P_{\pi_i \to \pi_j} = \frac{N_{(\pi_i \to \pi_j)}}{N_{\pi_i}}. \tag{6}$$

Step 2: Calculate the DNTE value E_c by taking P_{π_i} and $P_{\pi_i \to \pi_j}$ as weights and probabilities, respectively, and combining them with the definition of Shannon entropy as follows:

$$E_c = \sum_{i=1}^{c} P_{\pi_i} \sum_{j=1}^{c} P_{\pi_i \to \pi_j} \log(P_{\pi_i \to \pi_j}). \tag{7}$$

By applying Shannon entropy, we combine the information from nodes and links to quantify the network's complexity. In Eq. (7), it is evident that the formula simultaneously captures effective information from both nodes and links in the DN, and the computed value of DNTE is more comprehensive in representing various aspects of the initial time series information.

Additionally, in the calculation process of DNTE, it is necessary to normalize the content within the second summation formula. Normalization ensures that the computed entropy value is scaled appropriately, making it easier to interpret and compare across different networks. The specific

ALGORITHM 1. Dispersion network-transition entropy (DNTE).

1: Input time series $X = \{x_1, x_2, \ldots, x_L\}$.
2: Obtain the cumulative distribution function sequence $Y = \{y_1, y_2, \ldots, y_L\}$ through Eq. (1).
3: Obtain the integer sequence $Z = \{z_1, z_2, \ldots, z_L\}$ by applying the round function in Eq. (2).
4: Construct $[G^c, W^c]$ based on the nodes and links in sequence Z.
5: Count the quantity and probabilities of the nodes and links in $[G^c, W^c]$.
6: Calculate DNTE values E_c based on Eq. (7).
7: Normalize the E_c.

procedure of normalization involves dividing the value obtained from Eq. (7) by $\log(c)$, since for each node i there are c possible linking results. Through normalization, the resulting E_c are scaled within the range of 0 to 1, offering a more intuitive representation of the complexity of the network, as a value between 0 and 1 clearly indicates the degree of complexity, with higher values representing greater complexity. Additionally, Algorithm 1 shows the pseudocode of the DNTE.

C. Parameter selection

As elucidated earlier, DNTE solely entails the determination of the parameter c, necessitating less *a priori* knowledge for parameter selection than alternative metrics. c determines the potential number of nodes, while c^2 represents the potential number of links. Hence, a judicious choice of c can effectively capture subtle pattern changes in the time series while ensuring that DNTE remains insensitive to noise, thereby retaining the essential information of the time series.

Figure 2 illustrates the mean and standard deviation of the three types of noise under different classes c. White noise exhibits constant power across all frequencies, resembling a random mixture, while pink noise accentuates lower frequencies, and blue noise accentuates higher frequencies in their respective spectral distributions [32]. The length of each noise class is 6000, with 200 randomly generated instances for each class. From the figure, it can be observed that with smaller c, the distinction among different noise types is more apparent. Larger c, particularly in the case of pink and blue noise, exhibit smaller variances, indicating greater stability in representing complexity. Therefore, we recommend a range for the parameter c between 10 and 20.

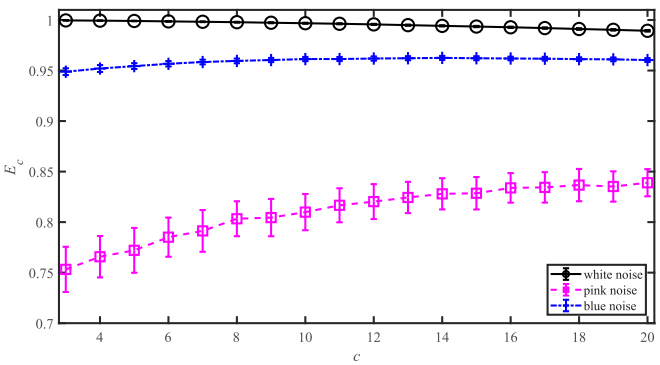


FIG. 2. Mean and standard deviation of the three types of noise under different classes c.

III. TESTS ON THE SIMULATED SIGNALS

In this section, we conduct comparative experiments on the performance of DNTE using simulated signals, including its ability to capture dynamic changes and differentiate between various chaotic models. Quantitative metrics of complexity for comparison include LZC [10], permutation entropy (PE) [31], and dispersion entropy (DE) [30], and their metric values are denoted as ε_1, ε_2, and ε_3, respectively. Additionally, we also conduct a statistical analysis of the computational consumption of these metrics.

A. Experiment (1): Capture of dynamic changes

The logistic map is a classic one-dimensional chaotic map widely used in the study of chaotic phenomena and nonlinear dynamical systems [33,34]:

$$x_{i+1} = \mu x_i(1 - x_i). \tag{8}$$

Here, μ acts as the bifurcation parameter controlling the level of chaos in the entire map, while x_0 denotes the initial point of the map. Typically, in chaotic systems, the specific value of the initial point has minimal impact on the results, and we set it as 0.1. In our experiments, we systematically varied μ within the range [3.5, 4] with an interval of 0.001. For each μ value, we extract the final 5000 iterations out of 10 000 to form the sample sequence. Figure 3(a) illustrates the bifurcation diagram of the logistic map, highlighting the transition from initial period 4 to subsequent period 8, period 16, and a progressively increasing chaotic behavior as the bifurcation parameter undergoes changes. Figures 3(b)–3(e) depict the variation curves of different metrics with the bifurcation parameter μ. Among these, LZC involves no parameter selection; for DNTE, the number of classes c is set to 5, 10, 15, and 20, respectively; the embedding dimension m for PE ranges from 3 to 6; DE involves the selection of both the embedding dimension m and the number of classes c. In addition to the suggested combinations (3,4), (4,3), and (4,4) [30], we also set (2,15) for comparison. This is because an embedding dimension of 2 is equivalent to the statistics of adjacent nodes; hence for completely normal signals, DNTE and DE are the same (same c). However, completely normal signals are almost nonexistent in the real world, and these differences can be reflected through Eqs. (6) and (7).

Figure 3 shows that each metric broadly reflects the overall chaotic variations of the logistic map, yet distinctions persist among them. For instance, within the interval $\mu \in [3.5, 3.545]$, the map resides in a period-4 state. However, both PE ($m = 5$ and 6) and DE ($m = 3$, $c = 4$, and $m = 4$, $c = 4$) curves exhibit frequent jumps in this interval. Furthermore, although the DE curve with parameters $m = 2$ and $c = 15$ is more reflective of the dynamics of the logistic map, the normalized ε_1 in the periodic state is greater than 0.5, which illustrates that when c is taken to be 15, due to the overly fine division of the categories, it may lead to a very similar distribution for some of the pattern categories, which

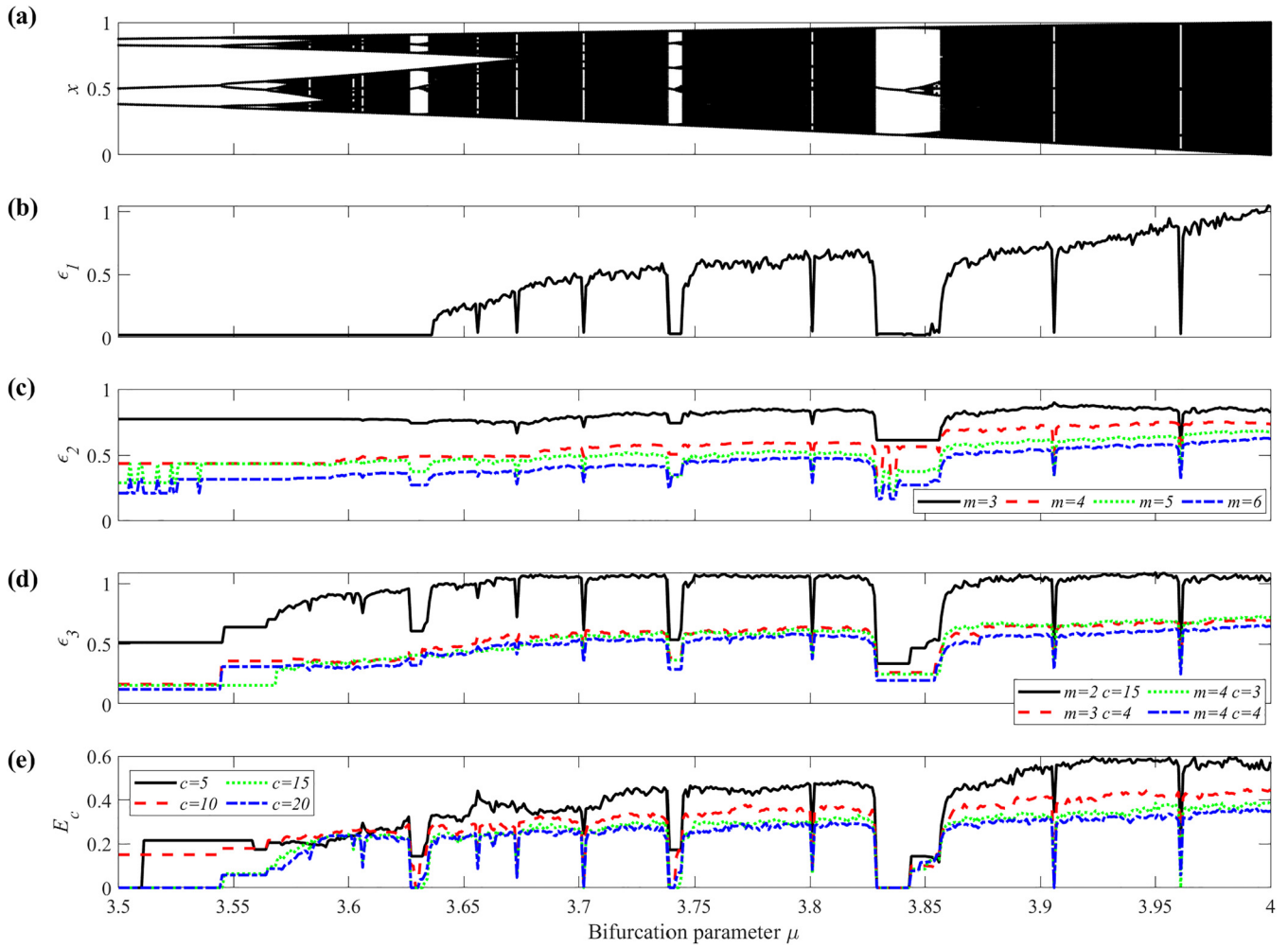


FIG. 3. Variation curves of different metrics with the bifurcation parameter μ of the logistic map. (a) Bifurcation diagram of the logistic map, (b) LZC, (c) PE, (d) DE, and (e) DNTE.

in turn makes the amount of useful information of the time series decrease. In contrast, it is worth noting that the DNTE curves depicted in Fig. 3(e) (except when c is set to 5) not only accurately describe the transition between the periodic and chaotic states, but also accurately reflect the periodic state of the logistic map (with an E_c of 0), and also show sensitivity to subtle bifurcations within the periodic state, particularly when $\mu \in [3.5, 3.635]$. In conclusion, compared with the other three metrics, DNTE exhibits significantly superior performance in detecting dynamic changes in the logistic map.

B. Experiment (2): Differentiation of the different chaotic models

Next, we introduce the Hénon map [18] and Lorenz map [35] to conduct experiments aimed at distinguishing different chaotic models. The formulas for the Hénon map and Lorenz map can be expressed as follows:

$$x_{i+1} = 1 + y_i - ax_i^2 \; y_{i+1} = bx_i, \tag{9}$$

$$\dot{x} = \sigma(y - x) \;\; \dot{y} = \rho x - y - xz \;\; \dot{z} = xy - \beta z. \tag{10}$$

Here, the parameters for the Hénon map are set as $a = 1.3$ and $b = 0.3$, while the parameters for the Lorenz map are set as $\sigma = 10$, $\rho = 28$, and $\beta = 8/3$. For each chaotic model, we select various sequence lengths ranging from 1000 to 10 000, with intervals of 1000. It is worth noting that we also omit the initial 5000 iterations of each map to ensure sequence stability. By calculating the E_c ($c = 15$), ε_2 ($m = 5$), and ε_1 and ε_3 ($m = 3$, $c = 4$; $m = 2$, $c = 15$) values for different sequence lengths of the chaotic models, the mean and standard deviation of the distributions are obtained as shown in Fig. 4, and all metrics are normalized for ease of comparison.

Observation of Fig. 4 reveals that the LZC can effectively identify nearly all three types of chaotic models. However, there is a slight sample confounding observed in the logistic and Lorenz maps, with large standard deviations for all models. Moreover, the metric values of the three chaotic models in DE ($m = 3$, $c = 4$) are very close, making it difficult to distinguish them effectively. Furthermore, the entropy value of DE ($m = 2$, $c = 15$) is markedly elevated, which leads to the logistic map and the Lorenz map being nearly identical, making them challenging to distinguish. Only PE and DNTE are found to be effective in distinguishing the three chaotic models, with their entropy values exhibiting minimal standard

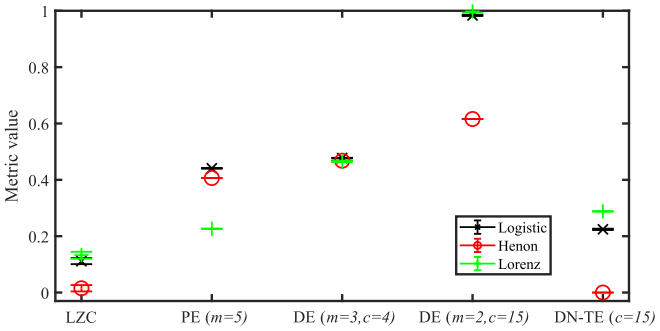


FIG. 4. Mean and standard deviation of various metrics for three chaotic models.

deviation. These findings indicate that PE and DNTE are capable of accurately reflecting the entropy values of input signals, even when the sample series are limited in size.

C. Experiment (3): Comparison of time consumption

The time consumed in the computation of various metrics is also of greater significance in practical application. Among the four metrics considered, both LZC and PE exhibit time complexities of $O(N^2)$, while DE and DNTE boast time complexities of $O(N)$. Figure 5 illustrates the average computational consumption of these metrics across varying time series lengths. Each time series length is randomly generated for 200 records.

It is clear from Fig. 5 that increasing the series length results in a linearly increasing computational cost for each metric. In terms of computational efficiency, PE and LZC emerge as the most time-consuming metrics, followed by DE, where the increase in the number of categories also leads to a longer computation time. It is worth noting that the DNTE algorithm is the most efficient. In particular, LZC and PE exhibit computational consumption three orders of magnitude higher than that of DNTE, and the results underscore the practical advantage of our proposed DNTE in providing superior computational efficiency.

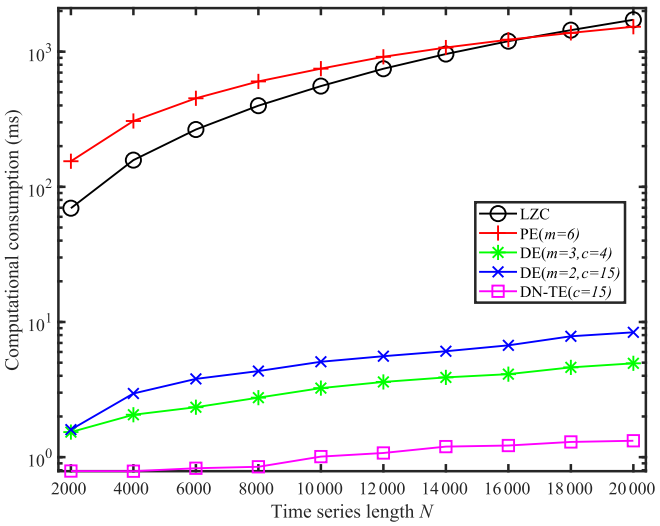


FIG. 5. Computational consumption of LZC, PE ($m = 6$), DE ($m = 3$, $c = 4$), DE ($m = 2$, $c = 15$), and DNTE ($c = 15$) for different time series lengths.

TABLE I. Specific types and labels of ShipsEar signals.

Class	Type	Label
A	Dredge	80__04_10_12_adricristuy
B	Motorboat	72__23_07_13_H3_lancha2
C	Passengers	9__10_07_13_marDeOnza_Espera
D	Ocean liner	69__23_07_13_H2_costaVoyager
E	Background noise	82__27_09_13_H3_lluvia

IV. APPLICATION TO REAL MEASURED HYDROACOUSTIC SIGNALS

The ability of DNTE to reflect dynamic changes and differentiate between chaotic models has been confirmed through simulated experiments. In this section, we conduct classification experiments on real measured hydroacoustic signals. The data include the publicly available ShipsEar dataset and measured signals of various modulation classes, and the compared metrics include LZC, PE ($m = 6$), and DE ($m = 2$, $c = 6$), and their values are also denoted as ε_1, ε_2, and ε_3, respectively. By comparing the feature distribution and recognition rates of different metrics for ship signals or modulation signals, we aim to validate the superiority of DNTE in the application of real measured signals.

A. Case (1): ShipsEar dataset

The ShipsEar dataset was recorded along the Atlantic coast in northwest Spain and was obtained from [36,37]. Because of the high intensity and diversity of port traffic, the ShipsEar dataset includes recordings of various types of ships. The collected signals are categorized into five major classes based on ship size, including four classes of ships and one class of environmental noise data. Each class of signal has a sampling frequency of 52.734 Hz, and we select a signal segment from each major class for experimentation. Further information on the selected signals can be found in Table I.

For each category of ShipsEar signal, we randomly select 200 nonoverlapping samples, each with a length of 10 000 sampling points, to create the experimental sample set. We then calculate the DNTE values E_c for these samples using parameter settings of $c = 15$. In addition, we calculate ε_1, ε_2, and ε_3 for comparison purposes, and the resulting feature distribution is displayed in the violin plot in Fig. 6, where the black line represents the mean values.

In Fig. 6(a), it appears that only class D can effectively distinguish itself from the other four classes of signals. Some samples from class A overlap with those from class B and class C, and a similar situation is observed between class B and class E. In Fig. 6(b), the entropy values of samples from the other four classes, except for class C, are mostly clustered in one region, indicating a poorer discriminative effect of PE. Additionally, the feature distributions in Figs. 6(c) and 6(d) are highly analogous. This is due to the fact that both DE and DNTE employ normal cumulative distribution mapping. In DE, the distributions for class A and class C are more

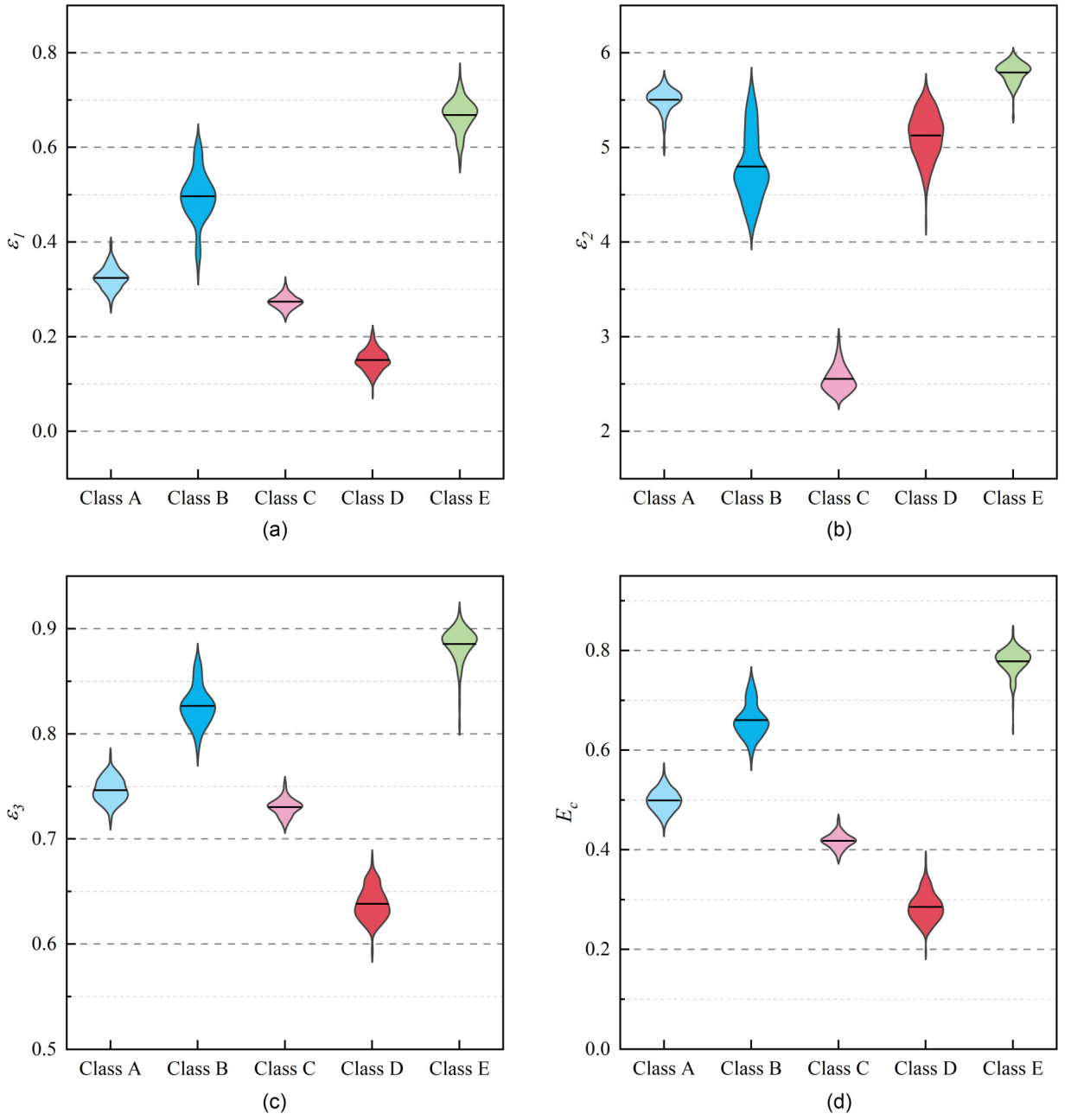


FIG. 6. Feature distributions of different metrics for five classes of the ShipsEar dataset: (a) LZC, (b) PE, (c) DE, and (d) DNTE.

closely aligned, while in DNTE, there is no discernible overlap between the two. In comparison, the effect of the feature distributions is significantly superior to that of LZC and PE.

In order to demonstrate the effectiveness of DNTE more intuitively, we also introduce the K-nearest neighbor (KNN) classifier by virtue of its simplicity and intuition [38], and validate the performance of DNTE through classification accuracy. By calculating the Euclidean distance between the test samples and all samples in the training set, the five nearest training samples are chosen. The class of each test sample is then determined using a majority voting method, achieving signal classification. Figure 7 shows the recognition rates of various metrics from the ShipsEar dataset.

Figure 7 shows that LZC and PE achieve 100% recognition rate for only one signal class, which is consistent with the analysis from the violin plots. However, their recognition and discrimination performances are lower for the other four classes in the sample set, resulting in lower recognition rates. In contrast, both DNTE and DE demonstrated high recognition rates across all five classes, although they did not achieve perfect recognition for any single class. Specifically, the DNTE proposed in this paper achieves a recognition rate of 97.4% on the five classes of the ShipsEar dataset, surpassing DE by 1.6%. In summary, these results confirm the effectiveness of DNTE in the classification of hydroacoustic signals.

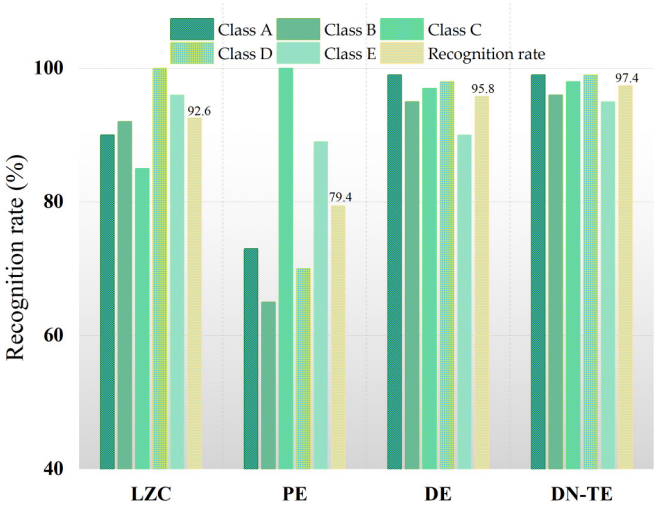


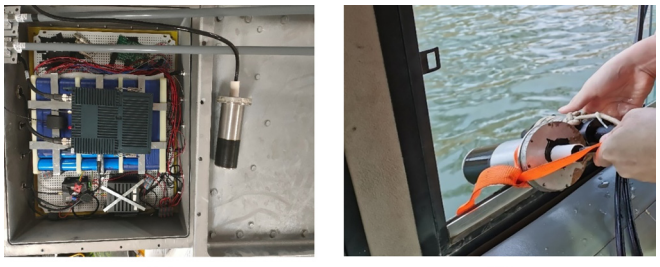
FIG. 7. Recognition rates of various metrics from the ShipsEar dataset.

B. Case (2): Modulation signals

In this subsection, we describe the measured data collection involving five classes of modulation signals. The modulation signals are collected at the Ganhe Reservoir in Xianyang City, Shaanxi Province, China, and the modulation classes include 2FSK, 4FSK, BPSK, DSSS, OFDM, and QPSK.

The geographical coordinates of the Ganhe Reservoir are latitude 34 °532 807′ and longitude 108 °429 436′. During the signal acquisition process, the signal transmitting device is deployed at the shore of the water area, while the sense recognition node device is deployed on a small boat approximately 1000 m from the shoreline. The node device is equipped with accessories such as batteries and waterproof enclosures, as depicted in Fig. 8(a), and the device placement scene is illustrated in Fig. 8(b). Special attention should be paid to the fact that the distance between the small boat and the launching equipment may change because of the water current, and the overall control is between 800 and 1200 m. Other parameter settings used in the experiment are detailed in Table II, including the depth of the transmitter and hydrophone, the center frequency of the carrier, and the sampling frequency of the hydrophone.

The total acquisition duration for each signal is 240 s, with the signal transmitter emitting a 20-s modulation signal at



(a) (b)

FIG. 8. (a) Sense recognition node devices. (b) Device layout diagram.

TABLE II. Specific information of modulation signals.

Parameter	Setting
Transmit depth	3 m
Hydrophone depth	3 m
Signal carrier center frequency	8 KHz
Hydrophone sampling frequency	50 KHz
Modulation class	2FSK, 4FSK, BPSK, DSSS, OFDM, and QPSK

10-s intervals. For each class of modulation signal, we take the initial 20-s signal as the experimental data and divide it into 100 samples at equal intervals, each with 10 000 sampling points. The ε_1, ε_2, ε_3, and E_c are calculated for these samples, and the resulting violin distribution is shown in Fig. 9.

In Fig. 9, in terms of feature stability, both LZC and DNTE exhibit feature distributions within relatively small ranges without significant interval spans. In particular, DNTE demonstrates stable representations for BPSK, DSSS, and OFDM with a small range of entropy value fluctuations. Regarding the distribution of various modulation signals, LZC, DE, and PE are capable of distinguishing a maximum of two signal classes without any overlap with other classes. However, the E_c's for 4FSK, BPSK, DSSS, and OFDM show clear distinctions, with only 2FSK and QPSK exhibiting small regions of overlap. Therefore, it can be observed that DNTE has the best stability in distinguishing among the six modulation signal classes, and the feature distributions make it easy to differentiate between various modulation signals.

Furthermore, the recognition results with the introduction of a classifier are shown in Fig. 10, where 50% of the samples are randomly selected as the training set and the remaining samples are used as the test set. Figure 10 shows that LZC, PE, and DE can achieve a 100% recognition rate for only two modulation signal classes, whereas the proposed DNTE accurately recognizes four of them. In terms of the average recognition rate, DNTE achieves a recognition rate of 97.67%, which is 6% higher than that of the second-ranked DE. These findings demonstrate that the method proposed in this paper offers an alternative approach for the application of nonlinear signals.

C. Supplemental experiments

To further quantify the time complexity of various metrics in the application of real measured signals, we conduct a statistical analysis of their computational time. Table III presents the computation times of various metrics for two datasets. It

TABLE III. Computation times of various metrics for two datasets.

Metric	ShipsEar (s)	Modulation signal (s)
LZC	626.28	367.65
PE	802.4	523.27
DE	65.17	16.82
DNTE	33.82	8.79

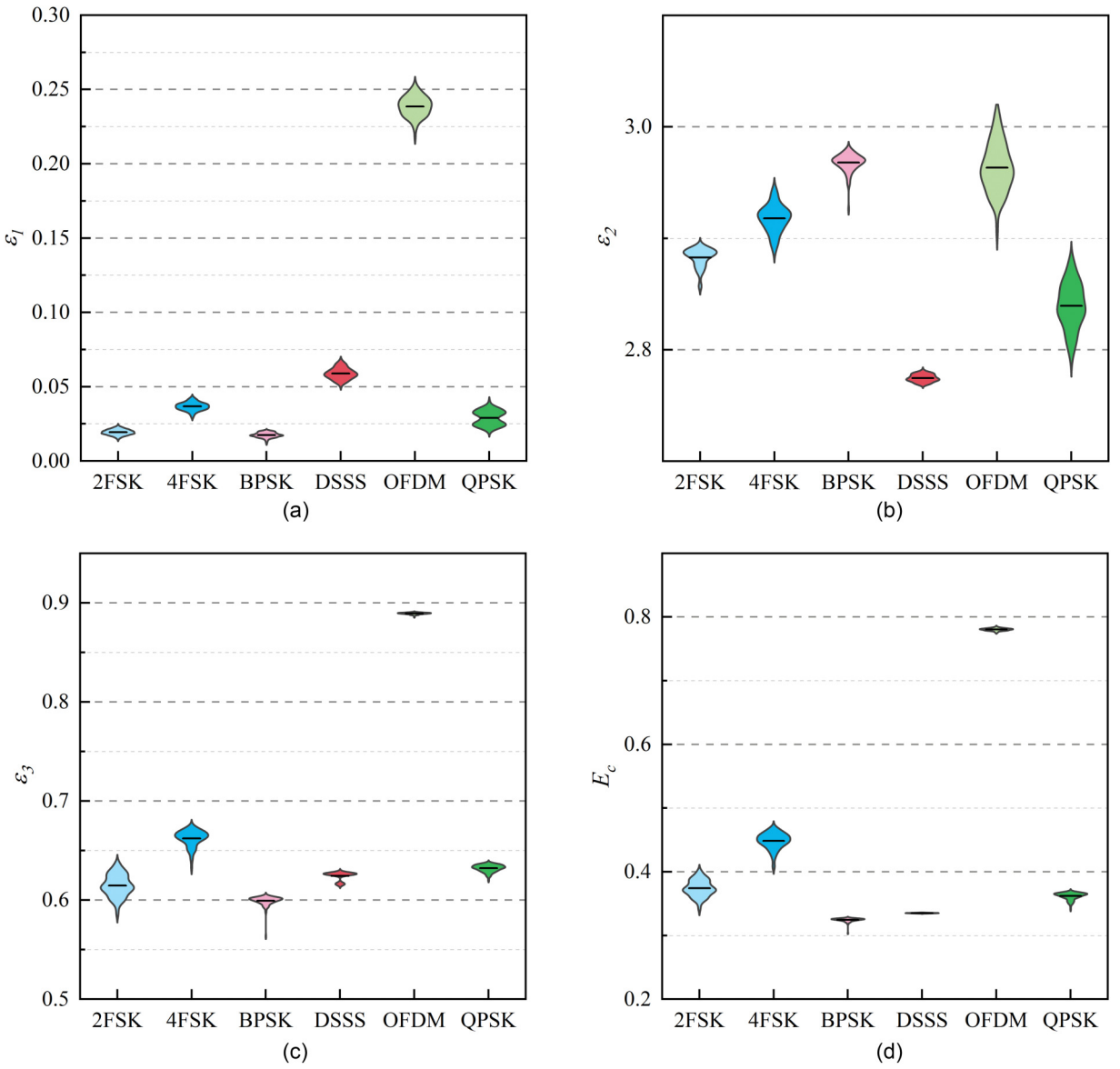


FIG. 9. Feature distributions of different metrics for six classes of modulation signals: (a) LZC, (b) PE, (c) DE, and (d) DNTE.

can be observed that, on any dataset, PE requires the longest computation time, followed by LZC, while DNTE consumes the least time. For a set of 600 samples with a length of 10 000 data points, DNTE requires only 8.79 s of computation time, further demonstrating the efficiency of the proposed DNTE in hydroacoustic applications.

Furthermore, to mitigate the impact of overfitting on the experiments, we adjust the training percentage in the

TABLE IV. Recognition rates and standard deviations for the ShipsEar dataset under different training percentages.

Training percentage	Recognition rates (%)			
	LZC	PE	DE	DNTE
20%	91.14 ± 0.013	73.42 ± 0.020	95.49 ± 0.013	97.17 ± 0.009
30%	91.26 ± 0.010	72.92 ± 0.015	95.92 ± 0.011	97.20 ± 0.007
40%	90.95 ± 0.014	72.56 ± 0.013	96.11 ± 0.009	97.14 ± 0.005
50%	90.82 ± 0.011	72.48 ± 0.016	96.14 ± 0.012	97.08 ± 0.006
60%	90.74 ± 0.015	71.63 ± 0.014	96.25 ± 0.009	98.03 ± 0.007
70%	90.46 ± 0.009	71.49 ± 0.020	95.96 ± 0.016	97.11 ± 0.009
80%	90.32 ± 0.012	71.02 ± 0.024	95.79 ± 0.013	96.89 ± 0.011

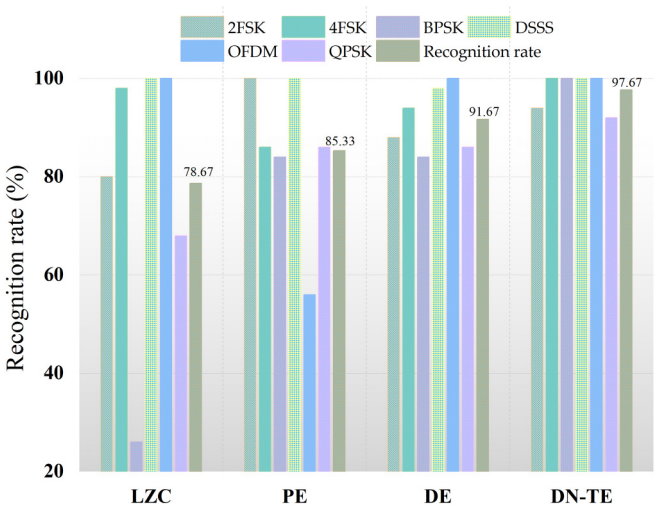


FIG. 10. Recognition rates of various metrics of modulation signals.

comparative experiments to 20%, 30%, 40%, 50%, 60%, 70%, and 80%. Additionally, for each classification experiment, we conduct 200 trials and calculate the mean and standard deviation of recognition rates to ensure the reliability of the experimental data. The results obtained from the two datasets are presented in Tables IV and V, respectively.

Analysis of Tables IV and V shows that DNTE consistently achieves the highest recognition rates and has a smaller standard deviation under any training percentage, whether it is the ShipsEar dataset or the modulation signals we collected, which indicates the robust performance of DNTE. Importantly, the rankings of the various metrics remain stable, unaffected by changes in the training percentage. It is noteworthy that in the modulation signal dataset, variations in the training percentage affect the performance of the metrics. Specifically, the recognition rates of various metrics tend to decrease as the training set decreases. For example, compared with an 80% training percentage, the recognition rate decreases by 4.35% under a 20% training percentage. However, a similar trend is not observed for the ShipsEar dataset, highlighting the importance of having sufficient data in practical experiments. Despite these variations, our proposed DNTE consistently outperforms other metrics under any dataset, demonstrating superior performance in terms of both robustness and efficiency.

V. CONCLUSIONS

This paper proposes a unique complexity measure, DNTE, which combines complex networks and information entropy to address the challenges in effective information extraction for nonlinear signals. The performance of DNTE is validated using both classical chaotic models and measured hydroacoustic signals. The main conclusions of this study are as follows:

1. DN is proposed by using scaled cumulative distribution function values as network nodes and establishing links based on temporal information. The results indicate that the obtained network can effectively reflect different dynamic regimes.

2. DNTE is proposed by simultaneously assessing the importance of each node and various links in the DN. Then quantification is achieved using the calculation formula of Shannon entropy, ensuring a comprehensive capture of effective information within the network.

3. In the comparative experiments involving simulated chaotic models, DNTE not only demonstrates superior efficacy in capturing the dynamic changes of time series, but also exhibits excellent performance in distinguishing between different models compared to LZC, PE, and DE. Furthermore, DNTE incurs the lowest computational cost.

4. In practical comparative experiments, regardless of whether the ShipsEar dataset or measured modulation signals were employed, the DNTE, combined with classifiers, consistently outperforms other metrics, even when small-sample training sets are used. Specifically, in classification experiments across six modulation signal classes, the recognition rate of DNTE reaches 97.67%, surpassing the second-ranking DE by 6%, thus adequately demonstrating its superiority in nonlinear signal complexity characterization.

In the future, we expect to expand DNTE to multivariate as well as multiscale forms, enabling a wider range of applications.

The datasets analyzed during the current study are available from the corresponding author on reasonable request.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (Key) (Grant No. 62031021), and the National Natural Science Foundation of China (Grant No. 62271404).

TABLE V. Recognition rates and standard deviations for modulation signals under different training percentages.

Training percentage	Recognition rates (%)			
	LZC	PE	DE	DNTE
20%	78.41 ± 0.025	83.60 ± 0.018	86.03 ± 0.027	93.76 ± 0.017
30%	77.74 ± 0.019	84.43 ± 0.017	89.96 ± 0.020	95.27 ± 0.014
40%	77.22 ± 0.019	84.92 ± 0.016	91.86 ± 0.017	96.62 ± 0.012
50%	76.47 ± 0.019	85.19 ± 0.015	92.11 ± 0.015	97.28 ± 0.008
60%	75.93 ± 0.019	85.39 ± 0.019	93.03 ± 0.015	97.51 ± 0.009
70%	74.89 ± 0.019	85.92 ± 0.022	93.63 ± 0.017	97.76 ± 0.009
80%	74.53 ± 0.022	85.79 ± 0.025	93.05 ± 0.021	97.82 ± 0.012

The author contributions are as follows: B.G.: conceptualization, methodology, software, writing—original draft; H.W.: data curation, project administration, writing—review and editing; X.S.: visualization, investigation, writing—review and editing; H.Z.: software, investigation, visualization; Y.Y.: supervision, data curation, resources.

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

[1] B. Han, Y. Zhou, and G. Yu, Second-order synchro extracting wavelet transform for nonstationary signal analysis of rotating machinery, Signal Process. **186**, 108123 (2021).

[2] A. Bariviera, Z. Luciano, and O. Rosso, An analysis of high-frequency cryptocurrencies prices dynamics using permutation-information-theory quantifiers, Chaos **28**, 075511 (2018).

[3] E. S. Nejevenko and A. A. Sotnikov, Adaptive modeling for hydroacoustic signal processing, Pattern Recognit. Image Anal. **16**, 5 (2006).

[4] C. Zhang, A. Mousavi, S. Masri, G. Gholipour, K. Yan, and X. Li, Vibration feature extraction using signal processing techniques for structural health monitoring: A review, Mech. Syst. Signal Process. **177**, 109175 (2022).

[5] P. Shi, X. Fan, J. Ni, and G. Wang, A detection and classification approach for underwater dam cracks, Struct. Health Monit. **15**, 551 (2016).

[6] S. Jiao, B. Geng, Y. Li, Q. Zhang, and Q. Wang, Fluctuation-based reverse dispersion entropy and its applications to signal classification, Appl. Acoust. **175**, 107857 (2021).

[7] Y. Li, B. Tang, S. Jiao, and Q. Su, Snake optimization-based variable-step multiscale single threshold slope entropy for complexity analysis of signals, IEEE Trans. Instrum. Meas. **72**, 6505313 (2023).

[8] F. Liu, G. Li, and H. Yang, A new feature extraction method of ship radiated noise based on variational mode decomposition, weighted fluctuation-based dispersion entropy and relevance vector machine, Ocean Eng. **266**, 113143 (2022).

[9] Y. Li, Y. Zhou, and S. Jiao, Variable-step multiscale Katz fractal dimension: A new nonlinear dynamic metric for ship-radiated noise analysis, Fractal Fract. **8**, 9 (2024).

[10] A. Lempel and J. Ziv, On the complexity of finite sequences, IEEE Trans. Inf. Theory **22**, 75 (1976).

[11] N. V. Kuznetsov, T. A. Alexeeva, and G. Leonov, Invariance of Lyapunov exponents and Lyapunov dimension for regular and irregular linearizations, Nonlinear Dyn. **85**, 195 (2016).

[12] C. E. Shannon, A mathematical theory of communication, Bell Syst. Tech. J. **27**, 623 (1948).

[13] Y. Li, B. Geng, and S. Jiao, Dispersion entropy-based Lempel-Ziv complexity: A new metric for signal analysis, Chaos Solitons Fractals **161**, 112400 (2022).

[14] J. Zhang and M. Small, Complex network from pseudoperiodic time series: Topology versus dynamics, Phys. Rev. Lett. **96**, 238701 (2006).

[15] H. Mo and Y. Deng, Identifying node importance based on evidence theory in complex networks, Physica A (Amsterdam) **529**, 121538 (2019).

[16] H. Li and Z. Liu, Multivariate time series clustering based on complex network, Pattern Recognit. **115**, 107919 (2021).

[17] Z. Zhang, Y. Qin, L. Jia, and X. Chen, Visibility graph feature model of vibration signals: A novel bearing fault diagnosis approach, Materials **11**, 2262 (2018).

[18] X. Wang, X. Han, Z. Chen, Q. Bi, S. Guan, and Y. Zou, Multi-scale transition network approaches for nonlinear time series analysis, Chaos Solitons Fractals **159**, 112026 (2022).

[19] X. Zheng, C. Feng, T. Li, and B. He, Analysis of autonomous underwater vehicle (AUV) navigational states based on complex networks, Ocean Eng. **187**, 106141 (2019).

[20] L. Lacasa, B. Luque, F. Ballesteros, J. Luque, and J. C. Nuño, From time series to complex networks: The visibility graph, Proc. Natl. Acad. Sci. USA **105**, 4972 (2008).

[21] R. Donner, M. Small, J. Donges, N. Marwan, Y. Zou, R. Xiang, and J. Kurths, Recurrence-based time series analysis by means of complex network methods, Int. J. Bifurcation Chaos **21**, 1019 (2011).

[22] X. Sun, M. Small, Y. Zhao, and X. Xue, Characterizing system dynamics with a weighted and directed network constructed from time series data, Chaos **24**, 024402 (2014).

[23] S. Bai and M. Niu, The visibility graph of n-Bonacci sequence, Chaos Solitons Fractals **163**, 112500 (2022).

[24] Y. Zou, R. Donner, N. Marwan, J. F. Donges, and J. Kurths, Complex network approaches to nonlinear time series analysis, Phys. Rep. **787**, 1 (2018).

[25] N. Marwan, J. Donges, Y. Zou, R. K. Donner, and J. Kurths, Complex network approach for recurrence analysis of time series, Phys. Lett. A **373**, 4246 (2009).

[26] D. Watts and S. Strogatz, Collective dynamics of 'small-world' networks, Nature (London) **393**, 440 (1998).

[27] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, Complex networks: Structure and dynamics, Phys. Rep. **424**, 175 (2006).

[28] D. Wang and Y. Xue, Average path length and degree distribution of networks generated by random sequence, Mod. Phys. Lett. B **35**, 2150347 (2021).

[29] K. Iwayama, Y. Hirata, K. Takahashi, K. Watanabe, K. Aihara, and H. Suzuki, Characterizing global evolutions of complex systems via intermediate network representations, Sci. Rep. **2**, 423 (2012).

[30] M. Rostaghi and H. Azami, Dispersion entropy: A measure for time-series analysis, IEEE Signal Process. Lett. **23**, 610 (2016).

[31] C. Bandt and B. Pompe, Permutation entropy: A natural complexity measure for time series, Phys. Rev. Lett. **88**, 174102 (2002).

[32] Y. Li, B. Tang, B. Geng, and S. Jiao, Fractional order fuzzy dispersion entropy and its application in bearing fault diagnosis, Fractal Fract. **6**, 544 (2022).

[33] G.-C. Wu, D. Baleanu, H.-P. Xie, and F.-L. Chen, Chaos synchronization of fractional chaotic maps based on the stability condition, Physica A (Amsterdam) **460**, 374 (2016).

[34] G. Wu and D. Baleanu, Chaos synchronization of the discrete fractional logistic map, Signal Process. **102**, 96 (2014).

[35] X. Mao, P. Shang, M. Xu, and C.-K. Peng, Measuring time series based on multiscale dispersion Lempel–Ziv complexity and dispersion entropy plane, Chaos Solitons Fractals **137**, 109868 (2020).

[36] ShipsEar: An underwater vessel noise database, available at https://underwaternoise.atlanttic.uvigo.es/.

[37] D. Santos-Domínguez, S. Torres-Guijarro, A. Cardenal-López, and A. Pena-Gimenez, ShipsEar: An underwater vessel noise database, Appl. Acoust. **113**, 64 (2016).

[38] L. Gao, D. Li, and L. Yao, Sensor drift fault diagnosis for chiller system using deep recurrent canonical correlation analysis and k-nearest neighbor classifier, ISA Trans. **122**, 232 (2022).