

Interval multiscale sample entropy: A novel tool for interval-valued time series complexity analysis

Ping Tang(唐萍), Bao-Gen Li(李宝根), and Yang Wang(王阳)
Citation: Chin. Phys. B, 2025, 34 (**12**): 120508. DOI: 10.1088/1674-1056/adea56
http://cpb.iphy.ac.cn; http://iopscience.iop.org/cpb

What follows is a list of articles you may be interested in

A novel baseline perspective visibility graph for time series analysis

Huang-Jing Ni(倪黄晶), Zi-Jie Song(宋紫婕), Jiao-Long Qin(秦姣龙), Ye Wu(吴烨), Shi-Le Qi(戚世乐), and Ming Song(宋明)
Chin. Phys. B, 2025, 34 (**8**): 080504. DOI: 10.1088/1674-1056/adce9a

Model-free prediction of chaotic dynamics with parameter-aware reservoir computing

Jianmin Guo(郭建敏), Yao Du(杜瑶), Haibo Luo(罗海波), Xuan Wang(王晅), Yizhen Yu(于一真), and Xingang Wang(王新刚)
Chin. Phys. B, 2025, 34 (**4**): 040505. DOI: 10.1088/1674-1056/adb733

Associated network family of the unified piecewise linear chaotic family and their relevance

Haoying Niu(牛浩瀛) and Jie Liu(刘杰)
Chin. Phys. B, 2025, 34 (**4**): 040503. DOI: 10.1088/1674-1056/adb26a

Impact of the chaotic semiconductor laser output power on time-delay-signature of chaos and random bit generation

Chenpeng Xue(薛琛鹏), Xu Wang(王绪), Likai Zheng(郑利凯), Haoyu Zhang(张昊宇), Yanhua Hong, and Zuxing Zhang(张祖兴)
Chin. Phys. B, 2025, 34 (**3**): 030504. DOI: 10.1088/1674-1056/ada7dd

A novel variable-order fractional chaotic map and its dynamics

Zhouqing Tang(唐周青), Shaobo He(贺少波), Huihai Wang(王会海), Kehui Sun(孙克辉), Zhao Yao(姚昭), and Xianming Wu(吴先明)
Chin. Phys. B, 2024, 33 (**3**): 030503. DOI: 10.1088/1674-1056/ad1a93

Interval multiscale sample entropy: A novel tool for interval-valued time series complexity analysis

Ping Tang(唐萍), Bao-Gen Li(李宝根)†, and Yang Wang(王阳)

Huangshi Key Laboratory of Metaverse and Virtual Simulation, School of Mathematics and Statistics, Hubei Normal University, Huangshi 435002, *China*

(Received 17 February 2025; revised manuscript received 18 May 2025; accepted manuscript online 1 July 2025)

To analyze the complexity of interval-valued time series (ITSs), a novel interval multiscale sample entropy (IMSE) methodology is proposed in this paper. To validate the effectiveness and feasibility of IMSE in characterizing ITS complexity, the method is initially implemented on simulated time series. The experimental results demonstrate that IMSE not only successfully identifies series complexity and long-range autocorrelation patterns but also effectively captures the intrinsic relationships between interval boundaries. Furthermore, the test results show that IMSE can also be applied to measure the complexity of multivariate time series of equal length. Subsequently, IMSE is applied to investigate interval temperature series (2000–2023) from four Chinese cities: Shanghai, Kunming, Chongqing, and Nagqu. The results show that IMSE not only distinctly differentiates temperature patterns across cities but also effectively quantifies complexity and long-term autocorrelation in ITSs. All the results indicate that IMSE is an alternative and effective method for studying the complexity of ITSs.

Keywords: interval multiscale sample entropy, interval-valued time series

PACS: 05.45.Tp, 02.70.Rr, 02.50.Sk, 05.45.–a

DOI: 10.1088/1674-1056/adea56 **CSTR:** 32038.14.CPB.adea56

1. Introduction

Entropy, as a measure of disorder, irregularity, uncertainty, or randomness, was first introduced in the field of thermodynamics by the German physicist R. Clausius. In the mid-20th century, Shannon proposed information entropy, also known as Shannon entropy, based on the concept of thermodynamic entropy.[1] Motivated by this, a multitude of entropy-based methodologies have been developed and implemented across diverse domains in recent decades.[2–9]

S. M. Pincus proposed approximate entropy (ApEn) in 1991,[10] which examines the discrepancy between the matching situation of m-dimensional subfragments and the matching situation of $m+1$-dimensional subfragments within a sequence through phase-space reconstruction. This method can compute the extent of the time series' irregularity and the self-similarity of a sequence. A smaller ApEn value indicates greater self-similarity in the time series.[11] However, J. S. Richman *et al*. found that ApEn is sensitive to the length of the data and lacks consistency. Therefore, based on this, the concept of sample entropy (SE) was proposed.[12] Compared to ApEn, this method eliminates the influence of vector self-matching, improves calculation speed, and ensures consistency.[13]

However, ApEn and SE are limited to a single scale and cannot accurately obtain dynamic information from a sequence. To overcome their shortcomings, multiscale entropy (MSE) was proposed by M. Costa *et al*., which revealed the changes in structural complexity shown by SE over multiple timescales and effectively distinguished the physiological sequence dynamics of healthy and diseased individuals.[14,15] In contrast to conventional entropic approaches, the MSE value resembles an "inverted U" shape as system uncertainty increases, indicating that both entirely ordered and entirely disordered sequences yield lower MSE values. This behavior aligns with the definition of the second type of complexity.[16] Moreover, it has been established that the MSE value of $1/f$ noise exceeds that of Gaussian white noise, signifying that the sequence structure of $1/f$ noise is more intricate than that of Gaussian white noise.[17] This approach has been extensively employed across multiple domains.[6,18,19]

MSE is restricted to the analysis of scalar time series, but multivariate time series are often observed in experimental and biological systems. M. U. Ahmed *et al*. introduced multivariate multiscale sample entropy (MMSE)[20,21] and systematically discussed the effects of embedding the dimensions, time delay, and amplitude magnitude of each signal channel on the results. W. J. Li *et al*. found that the MMSE method had the problems of missing relevant cross-channel data and unwanted deviation of estimation results and introduced improved multivariable multiscale sample entropy (IMMSE).[22] Since then, a variety of entropic methods for the complexity analysis of multichannel signals have been proposed, such as multivariate distance dispersion entropy[23] and cross-channel multiscale permutation entropy.[24] These methods provide a richer toolset for analyzing the complexity of multivariate time se-

†Corresponding author. E-mail: baogenli@hbnu.edu.cn

© 2025 Chinese Physical Society and IOP Publishing Ltd. All rights, including for text and data mining, AI training, and similar technologies, are reserved.

ries.

In previous studies, most of the modeling has been done for point-valued time series, but in recent years, researchers have found that the results of studies using point-valued time series are sometimes insufficiently precise, and it is difficult to improve the accuracy of the forecasts.[25] On the other hand, in some cases, scholars are interested in interval-type data across various domains. For example, in economics, the maximum and minimum prices of stocks help determine optimal times for buying and selling.[26] In meteorology, daily maximum and minimum temperatures serve as indicators to guide people's work and daily life.[27] Similarly, in medicine, daily variations in diastolic and systolic blood pressure provide doctors with important information for assessing patients' health conditions.[28]

In contrast to point-valued time series, interval-valued time series (ITSs) not only encompass the precise values collected at each time point but also capture the range of data changes. In addition, ITSs have advantages such as containing more accurate and complete information,[29,30] including more volatile data, being less influenced by outliers, and being robust in prediction.[31] These features have led ITSs to receive widespread attention from scholars. At present, research into ITSs focuses on the fields of prediction[32–36] and clustering.[37–39] However, research into the complexity of ITS systems is relatively scarce. As a key tool to measure system complexity, entropic methods have attracted great interest in recent years.

C. Y. Hu *et al.* introduced the concept of entropy in interval-valued datasets into the field,[40] and they found that utilizing interval-valued data significantly improves the efficiency and effectiveness of information processing and uncertainty management. Furthermore, they applied this entropic concept to stock market data,[41] and the results showed that the information entropy of interval-valued data was significantly higher than that of traditional point-valued data. The greater the information entropy, the more information can be obtained from it, which proves the superiority of ITSs in terms of information theory.[41] However, this method mainly relies on the hypothesis of interval distribution and is not directly applicable to multiscale analysis.

Inspired and motivated by MSE and IMMSE, the interval multiscale sample entropy (IMSE) method is proposed in this paper to quantify the complexity of ITS. Through simulation experiments, interval sample entropy (ISE) is evaluated on different timescales, revealing that the IMSE method can not only identify long-term autocorrelations in ITSs but also capture relevant information between the upper and lower bounds of intervals. The IMSE method is also suitable for multivariate time series of the same length. In an analysis of interval temperature data from different cities, IMSE demonstrates its ability to comprehensively capture the complexity of ITSs and distinctly differentiate their characteristics. In summary, the contributions of this paper are as follows: (i) A novel IMSE method is proposed to measure the complexity of ITSs. (ii) The interval is reconstructed as a whole, the Frobenius norm is introduced to measure the distance of the embedding matrix, and the difference between the two matrices is measured from the perspective of the overall size of the matrix. The test results show that IMSE can also be applied to measure the complexity of multivariate time series of equal length.

IMSE takes ITSs as research objects and explores their dynamic complexity across multiple scales. IMSE considers both time and numerical information in ITSs and does not need to assume an interval distribution. Overall, IMSE is more convenient to use. This paper aims to provide a new perspective on the complexity analysis of ITSs and to introduce a new analytical tool for researchers and practitioners in related fields.

The paper is organized as follows. Section 2 introduces the MSE and IMMSE. Section 3 introduces the computational steps of the IMSE method. Section 4 analyzes the complexity of ITS through simulation experiments and real-world data and demonstrates the performance of the IMSE method. Section 5 presents the conclusions.

2. Established entropy methods

The IMSE method proposed in this paper is derived from MSE and IMMSE. Therefore, in this section, the calculation steps of the MSE and IMMSE methods are first introduced.

2.1. Multiscale entropy

As a meaningful physiological complexity measure, MSE is used to evaluate the relative complexities of normalized time series on multiple scales.[14] The MSE process comprises two primary steps: (1) coarse granulation of the time series; (2) computation of the SE for the coarse-grained sequences. The calculation of SE and the coarse-granulation process are two largely independent methods. The process for calculating the SE is outlined below.

For a given one-dimensional discrete time series $X = \{x_1, x_2, \ldots, x_N\}$ of length N, given the embedding dimension m and the time delay τ, the delay vector $X_m^\tau(i)$ is generated by phase space reconstruction,[42] denoted by

$$X_m^\tau(i) = \left\{x_i, x_{i+\tau}, \ldots, x_{i+(m-1)\tau}\right\}. \tag{1}$$

In Eq. (1), $1 \le i \le N-(m-1)\tau$. In general, the time delay is fixed at $\tau = 1$, and the embedding dimension is set to $m = 2$.[10–12,15] Next, define the distance between any two delay vectors as

$$\mathrm{dis}(i,j) = \|X_m^\tau(i) - X_m^\tau(j)\|_\infty, \tag{2}$$

where $\|\cdot\|_\infty$ stands for the infinite norm, $1 \le i, j \le N-(m-1)\tau$ and $i \ne j$.

Given a tolerance threshold of r, for a fixed $X_m^\tau(i)$, let the number of other vectors whose distance from $X_m^\tau(i)$ is less than r be B_i; then, the probability that the sequence is within the threshold is

$$B_i^m(r) = \frac{B_i}{N-(m-1)\tau-1}. \tag{3}$$

Further, for all $X_m^\tau(i)$, $1 \le i \le N-(m-1)\tau$, the mean value of $B_i^m(r)$ is

$$B_m(r) = \frac{1}{N-(m-1)\tau} \sum_{i=1}^{N-(m-1)\tau} B_i^m(r). \tag{4}$$

In general, the tolerance threshold $r \in [0.1 \times \mathrm{std}, 0.25 \times \mathrm{std}]$, std is the standard deviation of the sequence.

Next, the embedding dimension is increased by one. This ensures that two vectors that exhibit similarity at m points retain the conditional probability of similarity at subsequent points. The previous process is then repeated to compute $B_{m+1}(r)$. The SE of the sequence X can now be defined as

$$\mathrm{SE}(X,m,\tau,r) = -\ln \frac{B_{m+1}(r)}{B_m(r)}. \tag{5}$$

MSE is the SE of the coarse-grained sequence. The following describes the coarse-grained processing process. For a given one-dimensional discrete time series $X = \{x_1, x_2, \ldots, x_N\}$ of length N, given a scale factor ε, the coarse-grained series is defined as

$$\left\{y_j^\varepsilon\right\} = \frac{1}{\varepsilon} \sum_{i=(j-1)\varepsilon+1}^{j\varepsilon} x_i, \tag{6}$$

where $n = \lfloor \frac{N}{\varepsilon} \rfloor$ and $1 \le j \le n$. Clearly, the length of the coarse-grained sequence is $1/\varepsilon$ times the length of the original sequence. We now calculate the SE of the coarse-granulated sequence, $\left\{y_j^\varepsilon\right\}$, and obtain the MSE

$$\mathrm{MSE}(X,\varepsilon,m,\tau,r) = \mathrm{SE}\left(y_j^\varepsilon, m, \tau, r\right). \tag{7}$$

In general, the more unpredictable or irregular a time series, the higher its entropy value, indicating that it is more complex. A consistently high entropy value across several scale factors signifies that the time series exhibits substantial complexity. If the entropy value consistently declines with variations in scale factors, it signifies that the information within the time series diminishes as the scale grows.[14] Real-world time series frequently encompass information across several scales, and in many instances, MSE is proficient at extracting important features.[6,18,19]

2.2. Improved multivariate multiscale sample entropy

MSE is confined to examining scalar time series, whereas multivariate time series are more common in experimental and biological contexts. Each variable can be treated as an independent time series only when all variables are statistically independent or at least uncorrelated. Ahmed *et al*. extended MSE to multivariate cases by proposing the MMSE method.[20,21] Subsequently, W. J. Li *et al*. proposed the IMMSE method, which addresses the deficiencies of MMSE regarding the absence of cross-channel correlation information and biased estimate outcomes.[22] Similar to the MSE algorithm, this algorithm can be executed by applying the following steps:

1. For a p-element time series $\left\{x_{i,k}\right\}_{i=1}^N$ of length N, a coarse-grained sequence is defined with respect to a scaling factor ϵ, i.e.,

$$\left\{y_{i,k}^\epsilon\right\} = \frac{1}{\epsilon} \sum_{i=(j-1)\epsilon+1}^{j\epsilon} x_{i,k}, \tag{8}$$

where $k = 1, 2, \ldots, p$, $1 \le j \le \frac{N}{\epsilon}$.

2. Let $n = \lfloor \frac{N}{\epsilon} \rfloor$; then, the coarse-grained sequence $\left\{y_{i,k}^\epsilon\right\}$ is a p-element time series of length n, denoted by $\left\{y_{i,k}\right\}_{i=1}^n$, and the multivariate time series is reconstructed in phase space to generate q composite delay vectors, i.e.,

$$\begin{aligned} Y_M(i) = [&y_{i,1}, y_{i+\tau_1,1}, \ldots, y_{i+(m_1-1)\tau_1,1}, \ldots, \\ &y_{i,p}, y_{i+\tau_p,p}, \ldots, y_{i+(m_p-1)\tau_p,p}], \end{aligned} \tag{9}$$

where $1 \le i \le q$, $M = [m_1, m_2, \ldots, m_p] \in \mathbb{R}^p$ denotes the embedding vector, while $\tau = [\tau_1, \tau_2, \ldots, \tau_p] \in \mathbb{R}^p$ represents the time delay vector, and $Y_M(i) \in \mathbb{R}^m \left(m = \sum_{k=1}^p m_k\right)$.

3. For a given $Y_M(i)$ and threshold r, the distance between any two composite delay vectors $Y_M(i)$ and $Y_M(j)$ is defined as the infinite norm. Let A_i represent the number of vectors in the reconstruction space that are situated within a threshold distance from $Y_M(i)$; thus, the probability that the sequence remains inside this threshold is

$$A_i^M(r) = \frac{A_i}{q-1}, \tag{10}$$

where $1 \le j \le q$ and $j \ne i$, the threshold $r \in [0.1 \times \mathrm{tr}(S), 0.25 \times \mathrm{tr}(S)]$, and $\mathrm{tr}(S)$ denotes the sum of the diagonal elements of the covariance matrix for the multivariate time series.

4. For all $Y_m(i)$, the mean value of $A_i^M(r)$ is

$$A_M(r) = \frac{1}{q} \sum_{i=1}^q A_i^M(r). \tag{11}$$

5. In contrast to the univariate embedding dimension m, which increases by one, the embedding dimension M in the multivariate SE algorithm is a vector. Consequently, we increase M to $M+p$, specifically by increasing each m_i in M by one, where $i = 1, 2, \ldots, p$. Repeat steps 3 and 4 to obtain the probability that the distance between any two composite delay

vectors $Y_{M+p}(i)$ and $Y_{M+p}(j)$ is less than a threshold value in the $M+p$ dimension, as follows:

$$A_{M+p}(r) = \frac{1}{q'}\sum_{i=1}^{q'} A_i^{M+p}(r), \tag{12}$$

where $1 \le i, j \le q'$ and $i \ne j$.

6. IMMSE is defined as

$$\mathrm{IMMSE}(M,\tau,r,N) = -\ln\frac{A_{M+p}(r)}{A_M(r)}. \tag{13}$$

3. Interval multiscale sample entropy

Compared with point-valued time series, ITSs contain richer information. Consequently, scholars have conducted extensive research on them.[43] Moore was the first to propose research on interval data;[44] thereafter, the modeling of interval data has developed rapidly. The IMSE method proposed in this paper takes ITS data as the entry point to explore the complexity of ITSs. In this section, the definition of ITS is introduced, and the computational steps of the IMSE method are detailed.

3.1. Interval-valued time series

An interval number can be expressed as a fuzzy number as $\tilde{x} = [x_{\mathrm{L}}, x_{\mathrm{R}}]$, where $x_{\mathrm{L}}, x_{\mathrm{R}} \in R$ and $x_{\mathrm{L}} \le x_{\mathrm{R}}$. Here, x_{L} and x_{R} represent the bottom and upper limits of the interval, respectively. $\tilde{x}$ can also be expressed in center, radius form as $\tilde{x} = (c, r)$, where $r = (x_{\mathrm{R}} - x_{\mathrm{L}})/2$ is the radius of the interval and $c = (x_{\mathrm{L}} + x_{\mathrm{R}})/2$ is the center of the interval.

On this basis, an ITS is a series composed of a set of interval numbers arranged in chronological order, where each interval represents the numerical range of a specific time point. An ITS with length n is defined as follows:[45]

$$\tilde{x}_t = \left\{ \left[x_t^{\mathrm{L}}, x_t^{\mathrm{R}}\right],\ t = 1,2,\ldots,n \right\}. \tag{14}$$

In Eq. (14), n represents the number of time points in the ITS $\{\tilde{x}_t\}$, referred to as the length; t is a unit of time, which may be seconds, minutes, or days; and the interval $\left[x_t^{\mathrm{L}}, x_t^{\mathrm{R}}\right]$ denotes the interval value of $\{\tilde{x}_t\}$ at time t, where $x_t^{\mathrm{L}}, x_t^{\mathrm{R}} \in R$ and $x_t^{\mathrm{L}} \le x_t^{\mathrm{R}}$, $t = 1,2,\ldots,n$. ITSs can be partitioned into an interval lower-bound sequence $\left\{\tilde{x}^{\mathrm{L}}\right\}$ and an interval upper-bound sequence $\left\{\tilde{x}^{\mathrm{R}}\right\}$:

$$\tilde{x}^{\mathrm{L}} = \left\{x_t^{\mathrm{L}}, t = 1,2,\ldots,n\right\},\ \tilde{x}^{\mathrm{R}} = \left\{x_t^{\mathrm{R}}, t = 1,2,\ldots,n\right\}. \tag{15}$$

3.2. Interval multiscale sample entropy

The existing MSE method is primarily aimed at univariate scalar time series analysis and has been adapted to multivariate time series analysis. However, in fields such as meteorology and finance, time series often appear in the form of interval data, which are arranged over time to form ITSs. For example, in meteorology, daily temperature values are often represented by intervals, with the lowest temperature serving as the lower bound and the highest temperature as the upper bound of the interval. Although it is possible to reduce ITSs to univariate time series by calculating the mean or to view the interval upper and lower bounds as a bivariate time series, both of these methods may lose some information. Therefore, it is particularly necessary to extend the MSE and IMMSE approaches to the analysis of ITSs. To this end, the concept of ISE is first introduced, which aims to capture and analyze the complexity of ITSs more comprehensively.

First, the phase space of the ITS is reconstructed, the univariate embedding vectors are generalized to the ITS, and the interval embedding matrix is generated. For a given ITS of length N, we denote the series as $[X] = \{[x_{1,1}, x_{1,2}], [x_{2,1}, x_{2,2}], \ldots, [x_{N,1}, x_{N,2}]\}$. Here, the interval at time t is $[x_{t,1}, x_{t,2}]$, where the interval's lower bound is $x_{t,1}$ and the upper bound is $x_{t,2}$. At this time, each interval can be regarded as a bivariate vector, and the ITS can be represented as

$$[X] = \begin{pmatrix} x_{1,1} & x_{1,2} \\ x_{2,1} & x_{2,2} \\ \vdots & \vdots \\ x_{N,1} & x_{N,2} \end{pmatrix}. \tag{16}$$

The phase space reconstruction of $[X]$ is different from the reconstruction used in the IMMSE method. An ITS, although it can be viewed as a bivariate data treatment, is essentially two dimensions of the same variable. Therefore, it is reconstructed as a whole; then, the interval-embedding matrix $[X_m^{\tau}(i)]$ is denoted by

$$[X_m^{\tau}(i)] = \begin{pmatrix} x_{i,1} & x_{i,2} \\ x_{i+\tau,1} & x_{i+\tau,2} \\ \vdots & \vdots \\ x_{i+(m-1)\tau,1} & x_{i+(m-1)\tau,2} \end{pmatrix}, \tag{17}$$

where m represents the embedding dimension, τ represents the time delay, $[X_m^{\tau}(i)]$ is an $m \times 2$ matrix, i represents the time unit, and $i = 1,2,\ldots,N-(m-1)\tau$.

It is now necessary to calculate the distance between any two interval-embedding matrices in the reconstructed space. Compared with the delay vectors obtained from traditional univariate time series reconstruction, the interval-embedding matrix not only records the values of the time points but also includes the range, i.e., the upper and lower limits, of each time point. Therefore, the interval-embedding matrix contains richer information. Here, the Frobenius norm is introduced to measure the distance between two embedding matrices. The Frobenius norm of a matrix is defined as the square root of the sum of the squares of all its elements.[46] It provides a comprehensive perspective on the overall size of the matrix and

measures the difference or similarity between two matrices. Compared to other norms, the Frobenius norm takes into account the information of all matrix elements and thus more comprehensively reflects the differences between two matrices. Based on these advantages, the Frobenius norm of the difference between two interval embedding matrices is defined as their distance, that is,

$$D(i,j) = \|[X_m^\tau(i)] - [X_m^\tau(j)]\|_{\mathrm{F}}, \tag{18}$$

where $\|\cdot\|_{\mathrm{F}}$ denotes the Frobenius norm, $[X_m^\tau(i)]$ and $[X_m^\tau(j)]$ represent two different interval embedding matrices, $i,j = 1,2,\ldots,N-(m-1)\tau$, and $i \neq j$.

Given a tolerance threshold r^*, for a given $[X_\tau^m(i)]$, calculate the probability that the distance between $[X_m^\tau(i)]$ and $[X_m^\tau(j)]$ is within the tolerance as follows:

$$C_i^m(r^*) = \frac{C_i}{N-(m-1)\tau-1}, \tag{19}$$

where C_i denotes the number of all embedding matrices with a separation of less than r^*, $i = 1,2,\ldots,N-(m-1)\tau$ and $i \neq j$. In the ISE method, we invoke the threshold range in the IMMSE method and set $r^* \in [0.1 \times \mathrm{tr}(S), 0.25 \times \mathrm{tr}(S)]$, where $\mathrm{tr}(S)$ denotes the trace of S, and S is the covariance matrix of $[X]$.

For all i, the probability that the sequence is within the threshold range is

$$C_m(r^*) = \frac{1}{N-(m-1)\tau} \sum_{i=1}^{N-(m-1)\tau} C_i^m(r^*). \tag{20}$$

Finally, increasing the embedding dimension m to $m+1$ and repeating the steps of Eqs. (18)–(20), we obtain

$$C_{m+1}(r^*) = \frac{1}{N-m\tau} \sum_{i=1}^{N-m\tau} C_i^{m+1}(r^*). \tag{21}$$

Therefore, we define the ISE of $[X]$ as

$$\mathrm{ISE}([X], m, \tau, r^*) = -\ln \frac{C_{m+1}(r^*)}{C_m(r^*)}. \tag{22}$$

Taking the ITS as a bivariate time series, the coarse-granulated processing method is the same as the IMMSE method (see formula (8)); we therefore set $k = 1,2$, $1 \leq i \leq \frac{N}{\epsilon}$. Then, the ISE of the coarse-granulated sequence $\{y_{i,k}^\epsilon\}$ is calculated as follows:

$$\mathrm{IMSE}([X], \epsilon, m, \tau, r^*) = \mathrm{ISE}\left(\{y_{i,k}^\epsilon\}, m, \tau, r^*\right), \tag{23}$$

where IMSE denotes the interval multiscale sample entropy.

4. Interval complexity analysis

To investigate the relative complexity of ITS, the ISE is calculated across various scales. Consistent with MSE theory,[14] it is assumed that the relative complexity of ITS data can be assessed based on the following criteria: (1) if the ISE value of one ITS is higher than that of another at most scales, it is considered more complex; (2) if the ISE value decreases monotonically with the scale factor, it indicates that the main information of the sequence resides at the smallest scale; (3) if the ISE monotonically increases or remains constant with scale, the sequence is considered to exhibit long-term autocorrelation.

In this part, three sets of simulation experiments are designed and conducted. The effectiveness of IMSE is illustrated, and the above assumptions are verified by using the IMSE method on various synthetic signals. We then use an interval temperature time series to evaluate the performance of IMSE. All the following experimental data have been normalized and standardized. To ensure that the upper bound of the interval is always greater than the lower bound, we add a quantity Δ to the upper bound sequence of the interval, where $\Delta = \max\left\{\tilde{x}^{\mathrm{L}} - \tilde{x}^{\mathrm{R}}\right\}$. Here, $\tilde{x}^{\mathrm{L}}$ and $\tilde{x}^{\mathrm{R}}$ denote the normalized lower and upper bound sequences, respectively.

4.1. Simulation analysis

First, univariate MSE analyses show that the sample entropy value for uncorrelated random white noise decreases consistently with increasing scale. However, the sample entropy for remotely correlated $1/f$ noise remains stable across all scales. This suggests that $1/f$ noise is structurally more complex than random white noise.[14]

To demonstrate that this conclusion also holds in the case of interval data, we generate four sets of interval data: (a) the two interval boundary sequences are mutually independent random white noise signals; (b) the two interval boundary sequences are $1/f$ noise signals that are independent of each other; (c) the upper interval boundary sequence is a white noise signal, and the lower interval boundary sequence is a $1/f$ noise signal; (d) the upper interval boundary sequence is a $1/f$ noise signal, and the lower interval boundary sequence is a white noise signal. Additionally, to evaluate the effectiveness of IMSE in complex systems more comprehensively, four groups of independent interval sequences are simulated and generated: (a) interval white noise; (b) interval $1/f$ noise; (c) interval Brown noise; (d) interval blue noise. The embedding dimension $m = 2$, the time delay $\tau = 1$, the threshold $r^* = 0.25 \times \mathrm{tr}(S)$, and the data length $N = 5000$ are configured. The ISE values of the interval data at different scales are then calculated, with the results displayed in Fig. 1.

Figure 1(a) illustrates that the four sets of interval data exhibit varying trends as the scale factor increases. The interval $1/f$ noise series exhibits the smallest entropy value when the scale is one but maintains a stable value as the scale factor increases. For interval mixed noise, the pink line (upper

bound of the interval is white noise) and the green line (upper bound of the interval is $1/f$ noise) essentially coincide. Their entropy is intermediate at a scale factor of one and decreases monotonically with increasing scale factor. In contrast, the interval white noise sequence exhibits maximum entropy at a scale of one, which diminishes swiftly as the scale factor escalates. This shows that white noise has the highest degree of randomness and irregularity at the smallest scale; as the scale factor increases, the random details are gradually smoothed out, and its complexity decreases. As the white noise sequence decreases, the $1/f$ noise sequence increases, and its complexity increases at larger scales. This is because the power spectral density of $1/f$ noise is inversely proportional to frequency, exhibiting long-range correlation and a self-similar structure. These properties enable $1/f$ noise to maintain a certain level of complexity at larger scales.

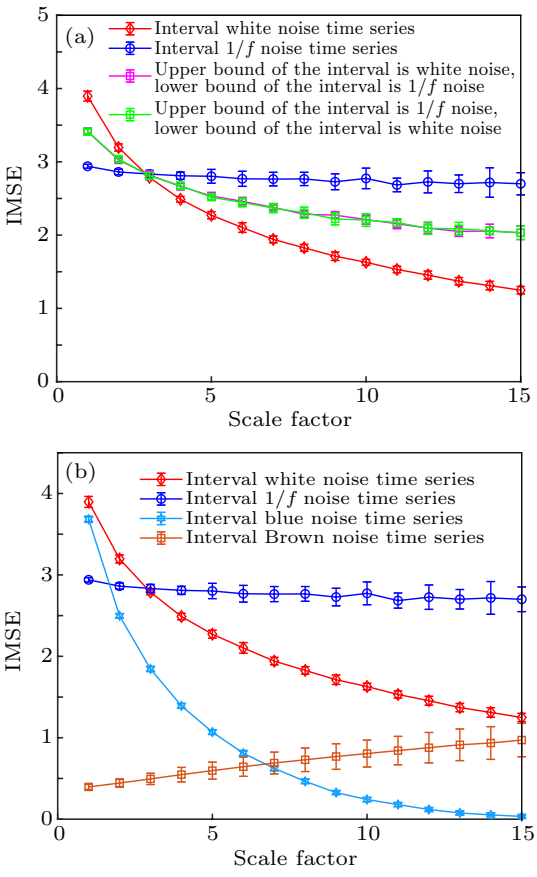


Fig. 1. (a) IMSE results for interval white noise, interval $1/f$ noise, and interval mixed noise. (b) IMSE results for interval white noise, interval $1/f$ noise, interval Brown noise, and interval blue noise. The curves represent the mean of 20 independent realizations and the error bars represent the standard deviation.

In Fig. 1(b), it can be observed that Brown noise exhibits the lowest entropy, whereas white noise shows the highest entropy at a scale factor of one. As the scale increases, the entropy of both white and blue noise declines rapidly — with blue noise decreasing more steeply. In contrast, $1/f$ noise maintains stable entropy across all scales, and Brown noise's entropy increases slightly with scale. This is attributed to the spectral density function and fractal characteristics of Brown noise and $1/f$ noise, which resemble those of complex signals.[47] These experimental results demonstrate that IMSE can effectively distinguish between different types of noise.

Second, the MSE algorithm can only consider the long-term autocorrelation of a single piece of data, whereas ISE considers the interval as a whole. Therefore, it is also necessary to consider the correlation between the upper and lower bound sequences. To validate the efficacy of IMSE in determining the correlation between the upper and lower bound signals of the interval, four groups of bivariate interval time series are simulated using the method described in Ref. [21]: (a) the upper and lower bound sequences of the interval are correlated white noise sequences; (b) the upper and lower bound sequences of the interval are uncorrelated white noise sequences; (c) the upper and lower bound sequences of the interval are correlated $1/f$ noise sequences; (d) the upper and lower bound sequences of the interval are unrelated $1/f$ noise sequences. The embedding dimension $m = 2$, the time delay $\tau = 1$, the threshold $r^* = 0.25 \times \mathrm{tr}(S)$, and a data length of $N = 5000$ are configured. The computation results are presented in Fig. 2.

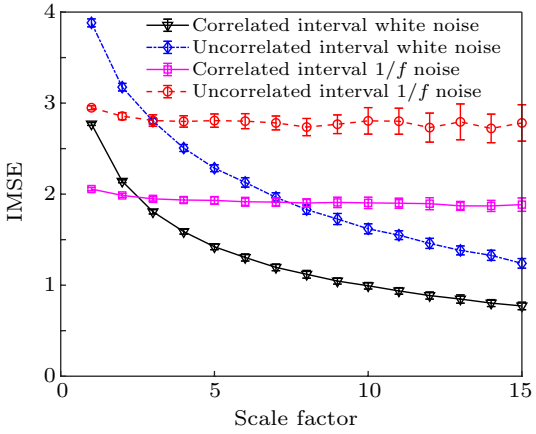


Fig. 2. IMSE results for correlated interval white noise, uncorrelated interval white noise, correlated $1/f$ noise, and uncorrelated $1/f$ noise. The curves represent the mean of 20 independent realizations, and the error bars represent the standard deviation.

Figure 2 shows that when the upper and lower bound sequences of the interval are $1/f$ noise, the entropy remains stable as the scale factor increases, regardless of correlation, and the entropy of the correlated $1/f$ noise is always lower than that of the uncorrelated $1/f$ noise. When the upper and lower bound sequences are correlated or uncorrelated white noise, their entropy decreases monotonically with increasing scale factor, and the entropy of correlated white noise is always lower than the entropy of uncorrelated white noise. When the scale is sufficiently large, the entropy of $1/f$ noise is higher than the entropy of white noise, regardless of whether the upper and lower sequences are correlated. The above results are consistent with the fact that when the upper and lower bounds are related, the interval value time series is more regular. Thus,

IMSE demonstrates the capability to find long-term autocorrelations within the sequence and to capture pertinent information within the upper and lower boundaries of the interval.

Finally, the utility of the IMSE is further demonstrated by calculating the IMSE results for different orders of bivariate vector autoregressive (AR) processes. The pth-order AR model can be expressed as

$$X_t = \sum_{i=1}^{p} \varphi_i X_{t-i} + \epsilon_t, \tag{24}$$

where ϵ_t denotes Gaussian white noise and φ_i is the autocorrelation coefficient. The AR model delineates the correlation between the present value and its historical counterparts, utilizing earlier temporal data of the variable to forecast its future values. Thus, as the order p increases, the number of historical values associated with the current value increases, the correlation increases, and it becomes easier to predict. In this paper, the process of generating three sets of bivariate AR is simulated (the parameters are the same as those used above). The correlation coefficients used in this simulation of AR are the same as those in Ref. [22]. The results are shown in Fig. 3.

Figure 3 clearly demonstrates that the IMSE approach can differentiate between AR processes of varying orders. The IMSE values of these three sets of bivariate AR models decrease with increasing regression order, and all show a tendency to decrease and then increase with increasing scale factor. This precisely satisfies the properties of AR processes, where as the regression order increases, correlation increases, complexity decreases, and entropy decreases.

The above simulated data can be treated as equal-length bivariate time series. Thus, the results of using IMMSE[22] to calculate their entropy are shown in Fig. 4. The results in Figs. 1–3 are largely consistent with those in Fig. 4, which shows that IMSE can be used for equal-length bivariate time series. According to Eq. (18), the Frobenius norm can also calculate the distance between multivariate time series embedding matrices. From the calculation process of SE,[10] we infer that IMSE can be used for equal-length multivariate time series.

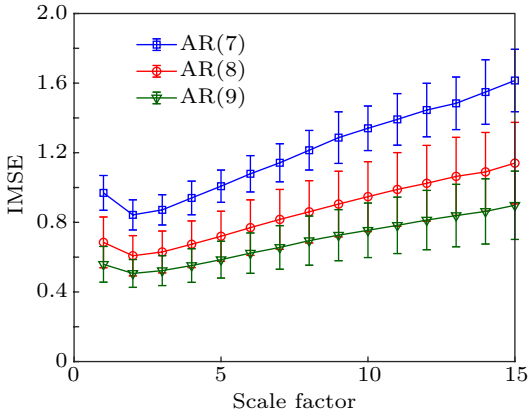


Fig. 3. IMSE results for different bivariate AR processes. The curves represent the mean of 20 independent realizations and the error bars represent the standard deviation.

Next, the steps of the IMSE method can also be extended to the case of equal-length multivariate time series. For a p-element time series $\{x_{i,k}\}_{i=1}^{N}(k = 1, 2, \ldots, p)$ of length N, the values of different elements at the same time are reconstructed by considering them as a whole to generate the embedding matrix, denoted by

$$X_m^{\tau}(i) = \begin{pmatrix} x_{i,1} & x_{i,2} & \cdots & x_{i,p} \\ x_{i+\tau,1} & x_{i+\tau,2} & \cdots & x_{i+\tau,p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{i+(m-1)\tau,1} & x_{i+(m-1)\tau,2} & \cdots & x_{i+(m-1)\tau,p} \end{pmatrix}, \tag{25}$$

where $X_m^{\tau}(i)$ is a matrix of $m \times p$. The following algorithm steps are the same as those of IMSE, as shown in Eqs. (18)–(23).

To verify the applicability of the algorithm to equal-length multivariate time series, two sets of equal-length three-variable time series are generated. The first set initially contains three independent white noise channels; the number of white noise channels is then gradually reduced while the number of $1/f$ noise channels is increased. The second set is composed of three-variable AR processes with different orders. The parameters are the same as those used above, and the extended multivariable IMSE algorithm is tested. The results are shown in Fig. 5.

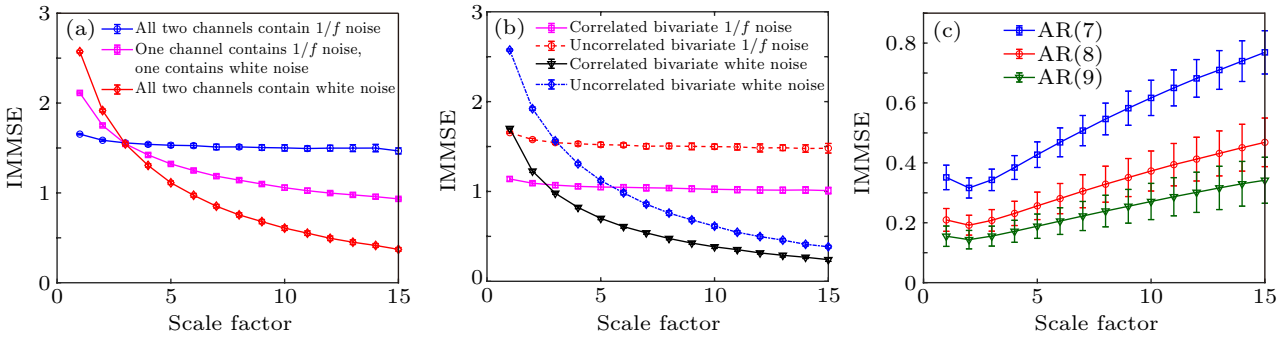


Fig. 4. (a) IMMSE results for bivariate data containing white noise and $1/f$ noise. (b) IMMSE results for correlated and uncorrelated white noise and $1/f$ noise. (c) IMMSE results for different bivariate AR processes. We set the embedding dimension $M = [2, 2]$, the time delay $\tau = [1, 1]$, and the threshold $r = 0.25 \times \mathrm{tr}(S)$. The curves represent the mean of 20 independent realizations and the error bars represent the standard deviation.

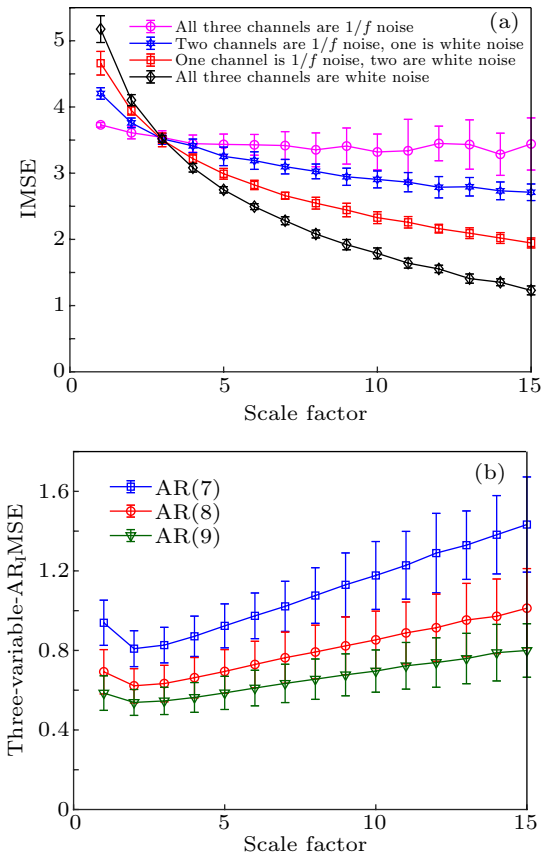


Fig. 5. (a) IMSE results for three-variable white noise, three-variable $1/f$ noise, and three-variable mixed noise. (b) IMSE results for three-variable AR processes. The curves represent the mean of 20 independent realizations, and the error bars represent the standard deviation.

From Fig. 5, it can be observed that applying the IMSE algorithm to three-variable time series basically gives the same results as those of IMMSE.[22] Therefore, we believe that IMSE is also applicable to equal-length multivariate time series.

4.2. Real-world data analysis

In this section, we select an interval temperature series comprising daily maximum and minimum temperatures from Shanghai, Kunming, Chongqing, and Nagqu in China, spanning the period 2000–2023. Missing data are input using linear interpolation. These data are provided by NOAA. First, the data are normalized and standardized. Then, the IMSE of the interval temperature series and the IMMSE of the bivariate time series composed of minimum and maximum temperatures are calculated. The results are presented in Fig. 6.

In Fig. 6, SH, KM, CQ, and NQ stand for Shanghai, Kunming, Chongqing, and Naqu, respectively. From Fig. 6(a), it can be observed that the IMSE values for Shanghai, Kunming, and Chongqing exhibit an overall decreasing trend. This is due to the greater volatility of temperature data on smaller timescales. Kunming's IMSE values are higher than those of the other three cities across all scales, showing the highest complexity. This phenomenon is likely due to Kunming's unique geographical and climatic conditions. Kunming is located in a low-latitude plateau, with complex topography and a humid subtropical climate, featuring four mild seasons, but with distinct wet and dry seasons and a large temperature difference between day and night. These factors collectively contribute to the high complexity observed in Kunming's climate. The IMSE values for the city of Nagqu are lower than those of the other three cities across all scales, showing low complexity. Moreover, the values remain stable across all scales, indicating that the series has long-term autocorrelation. These specifics are related to the geographic location of the city of Nagqu. Nagqu is located in the Qinghai-Tibet Plateau, with an average altitude of more than 4500 m, and the terrain is simple, mainly plateau. Affected by altitude and plateau topography, the climate is always cool and shows strong regularity and stability. Therefore, the interval temperature series in Nagqu shows strong regularity compared to the other three cities. Figure 6(b) shows a trend similar to that in Fig. 6(a). However, when the scale reaches a larger range, IMSE shows better differentiation ability than IMMSE. In summary, IMSE can effectively capture the complexity differences in temperature sequences between these four city intervals.

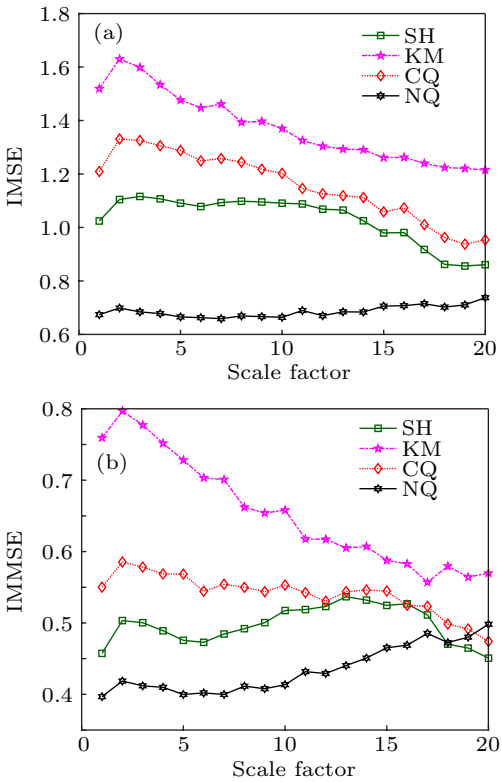


Fig. 6. (a) IMSE results for interval temperature time series in each city. (b) IMMSE results for bivariate temperature time series in each city. The parameter settings are consistent with those in the simulation experiment.

5. Conclusions

This article proposes the IMSE method for analyzing the complexity of ITSs, inspired by MSE and IMMSE. Through

simulation experiments and real-world data analysis, we validate the effectiveness and feasibility of IMSE.

In the simulation experiments, we generated different types of ITS. The results show that the IMSE method is able to accurately capture the complexity changes of these series across different scales. Sequences with long-term autocorrelation maintain high entropy values at all scales, whereas those with high stochasticity exhibit a monotonic decrease in entropy as the scale factor increases. Additionally, the IMSE method can distinguish whether the signals at the upper and lower bounds of the intervals are correlated or not, further validating its advantages in analyzing ITSs. From the results in Figs. 1–5, it can be observed that our IMSE method is also applicable to multivariate time series.

In the analysis of real-world data, interval temperature series comprising daily maximum and minimum temperatures from Shanghai, Kunming, Chongqing, and Naqu in China, spanning the period 2000–2023, are selected as the research object. Our findings indicate that the IMSE method is able to clearly distinguish the temperature series of each city. The IMSE values of Shanghai, Kunming, and Chongqing show an overall decreasing trend, suggesting that these three cities have greater volatility and complexity of temperature data at smaller timescales. The IMSE values for Kunming are higher than those of the other three cities across all scales, showing the highest complexity. The IMSE values for Naqu are lower than those of the other three cities across all scales, showing lower complexity, and remain stable at all scales, suggesting long-term autocorrelation of the series. Comparing the IMSE results for interval temperature sequences with the IMMSE results for bivariate temperature sequences, it is found that IMSE shows better discriminatory ability than IMMSE when the scale reaches a large range.

Overall, our IMSE approach has been effectively validated using both simulation data and real data. The IMSE method is able to effectively capture the complexity and long-term autocorrelation of ITSs. It can also clearly distinguish the characteristics of the temperature series in each city. Thus, it provides a novel alternative tool for the study of ITSs.

Acknowledgement

Project supported by Hubei Provincial Department of Education Science and Technology Plan Project (Grant No. B2022165).

References

[1] Shannon C E 1948 *Bell Syst. Tech. J.* **27** 379
[2] Tang J and Liu X Q 2019 *Acta Phys. Sin.* **68** 149801 (in Chinese)
[3] Zhang J C, Ren W K and Jin N D 2020 *Chin. Phys. Lett.* **37** 090501
[4] Wang S Y, Shi C F, Qian G B and Wang W L 2018 *Acta Phys. Sin.* **67** 018401 (in Chinese)
[5] Silva L E V and Murta L O J 2012 *Chaos* **22** 043105
[6] Silva L E V, Cabella B C T, Neves U P C and Murta L O J 2015 *Phys. A* **422** 143
[7] Xu M J, Shang P J and Huang J J 2016 *Commun. Nonlinear Sci. Numer. Simul.* **35** 17
[8] Yao W P 2025 *Acta Phys. Sin.* **74** 040502 (in Chinese)
[9] Ponta L, Murialdo P and Carbone A 2021 *Phys. A* **570** 125777
[10] Pincus S M 1991 *Proc. Natl. Acad. Sci. USA* **88** 2297
[11] Pincus S M 1995 *Quantitative Neuroendocrinology* (New York: Academic Press) p. 336
[12] Richman J S and Moorman J R 2000 *Am. J. Physiol. Heart Circ. Physiol.* **278** H2039
[13] Ramdani S, Bouchara F and Lagarde J 2009 *Chaos* **19** 013123
[14] Costa M, Goldberger A L and Peng C K 2002 *Phys. Rev. Lett.* **89** 068102
[15] Costa M, Goldberger A L and Peng C K 2005 *Phys. Rev. E* **71** 021906
[16] Mitchell M 2009 *Complexity: A Guided Tour* (New York: Oxford University Press) pp. 94–114
[17] Zhang Y C 1991 *J. Phys. I* **1** 971
[18] Wang J, Shang P J, Zhao X J and Xia J N 2013 *Int. J. Mod. Phys. C* **24** 1350006
[19] Wang J, Shang P J, Xia J N and Shi W B 2015 *Phys. A* **421** 583
[20] Ahmed M U, Li L, Cao J T and Mandic D P 2011 *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, August 30–September 03, 2011, Boston, USA, p. 810
[21] Ahmed M U, and Mandic D P 2011 *Phys. Rev. E* **84** 061918
[22] Li W J, Shen X H, Li Y A and Chen Z 2023 *Chaos* **33** 063125
[23] Niu Y, Ding R N, Zhou M N, Sun J, Dou M L, Wen X, Cui X H, Yao R, Wei J and Xiang J 2025 *Nonlinear Dyn.* **113** 8443
[24] Li W J, Shen X H, Li Y A, Chen Z and Shen Y P 2024 *Chaos* **34** 093143
[25] Xiong T, Li C G and Bao Y K 2017 *Econ. Model* **60** 11
[26] Jiang M R, Chen W, Xu H L and Liu Y X 2024 *Pattern Recognit.* **145** 109920
[27] Zhang T, Zhou Y Y, Zhao K G, Zhu Z Y, Chen G, Hu J and Wang L 2022 *Earth Syst. Sci. Data* **14** 5637
[28] Lu Y, Linderman G C, Mahajan S, Liu Y T, Huang C X, Khera R, Mortazavi B J, Spatz E S and Krumholz H M 2023 *Circ. Cardiovasc. Qual. Outcomes* **16** e009258
[29] Billard L and Diday E 2003 *J. Am. Stat. Assoc.* **98** 470
[30] Yang D C, Guo J E, Sun S L, Han J and Wang S Y 2022 *Appl. Energy* **306** 117992
[31] Xiong T, Bao Y K, Hu Z Y and Chiong R 2015 *Inf. Sci.* **305** 77
[32] Zhou Y T, Fan Y, Chen Z Y and Sun J C 2017 *Chin. Phys. Lett.* **34** 050502
[33] Li Q Q, Zheng R Z, Ji A B and Ma H Y 2025 *Spat. Stat.* **65** 100875
[34] Yang K, Cheng Z S, Li M C, Wang S Y and Wei Y J 2024 *Appl. Energy* **353** 122102
[35] Tao Z F, Ni W Q and Wang P 2024 *Eng. Appl. Artif. Intell.* **133** 108007
[36] Liu S H, Xie G, Wang Z Z and Wang S Y 2024 *Appl. Energy* **359** 122613
[37] Maharaj E A, Teles P and Brito P 2019 *Stat. Comput.* **29** 1011
[38] Sun L R, Zhu L J, Li W C, Zhang C H and Balezentis T 2022 *Inf. Sci.* **606** 910
[39] Sun L R, Mao H F, Zheng C, Zhang C H and Balezentis T 2023 *Appl. Math. Model.* **123** 627
[40] Hu C Y and Hu Z H 2020 *Information Processing and Management of Uncertainty in Knowledge-Based Systems*, June 15–19, 2020, Lisbon, Portugal, p. 407
[41] Hu C Y and Hu Z H 2020 *Information Processing and Management of Uncertainty in Knowledge-Based Systems*, June 15–19, 2020, Lisbon, Portugal, p. 422
[42] Takens F 1981 *Dynamical Systems and Turbulence, Warwick 1980*, 1979/80, Coventry, UK, p. 366
[43] Wang P, Gurmani S H, Tao Z F, Liu J P and Chen H Y 2024 *J. Forecast.* **43** 249
[44] Moore R E, Kearfott R B and Cloud M J 2009 *Introduction to Interval Analysis* (Philadelphia: Society for Industrial and Applied Mathematics) pp. 51–83
[45] Wang X, Yu F S, Pedrycz W and Yu L 2019 *Expert Syst. Appl.* **125** 293
[46] Golub G H and Van Loan C F 2013 *Matrix Computations - 4th Edition*, (Baltimore: Johns Hopkins University Press) pp. 71–73
[47] Omidvarnia A, Zalesky A, Mansour L S, Van De Ville D, Jackson G D and Pedersen M 2021 *Neuroimage* **230** 117760