Frontiers in **Network Physiology**

Testing dynamic correlations and nonlinearity in bivariate time series through information measures and surrogate data analysis

Helder Pinto[1,2]*[†], Ivan Lazic[3†], Yuri Antonacci[4], Riccardo Pernice[4], Danlei Gu[5], Chiara Barà[4], Luca Faes[4§] and Ana Paula Rocha[1,2§]

[1]Departamento de Matemática, Faculdade de Ciências, Universidade do Porto, Porto, Portugal, [2]Centro de Matemática da Universidade do Porto (CMUP), Porto, Portugal, [3]Faculty of Technical Sciences, University of Novi Sad, Novi Sad, Serbia, [4]Department of Engineering, University of Palermo, Palermo, Italy, [5]Beijing Jiaotong University, Beijing, China

The increasing availability of time series data depicting the evolution of physical system properties has prompted the development of methods focused on extracting insights into the system behavior over time, discerning whether it stems from deterministic or stochastic dynamical systems. Surrogate data testing plays a crucial role in this process by facilitating robust statistical assessments. This ensures that the observed results are not mere occurrences by chance, but genuinely reflect the inherent characteristics of the underlying system. The initial process involves formulating a null hypothesis, which is tested using surrogate data in cases where assumptions about the underlying distributions are absent. A discriminating statistic is then computed for both the original data and each surrogate data set. Significantly deviating values between the original data and the surrogate data ensemble lead to the rejection of the null hypothesis. In this work, we present various surrogate methods designed to assess specific statistical properties in random processes. Specifically, we introduce methods for evaluating the presence of autodependencies and nonlinear dynamics within individual processes, using Information Storage as a discriminating statistic. Additionally, methods are introduced for detecting coupling and nonlinearities in bivariate processes, employing the Mutual Information Rate for this purpose. The surrogate methods introduced are first tested through simulations involving univariate and bivariate processes exhibiting both linear and nonlinear dynamics. Then, they are applied to physiological time series of Heart Period (RR intervals) and respiratory flow (RESP) variability measured during spontaneous and paced breathing. Simulations demonstrated that the proposed methods effectively identify essential dynamical features of stochastic systems. The real data application showed that paced breathing, at low breathing rate, increases the predictability of the individual dynamics of RR and RESP and dampens nonlinearity in their coupled dynamics.

KEYWORDS

complex systems, coupling, information storage, information theory, mutual information rate, surrogate analysis

1 Introduction

The increasing accessibility of detailed and extensive signal recordings from various sources is opening the way to the representation of complex systems through network structures. For example, in physiology, the traditional reductionist approach, which involves studying the function of an organ system in isolation, is now complemented by a holistic investigation of collective interactions among diverse organ systems in the field of Network Physiology (Bashan et al., 2012; Ivanov et al., 2017; Ivanov, 2021). Network Physiology falls within the broader domain of Network Science, an expansive interdisciplinary field dedicated to advancing theoretical and practical methodologies to improve the comprehension of natural and artificial networks characterized by hierarchical structures (Barabási, 2013).

Within the Network Physiology approach, the human body is commonly represented as a graph, encoding the observed dynamical system with distinct nodes (e.g., organ systems) connected by edges that map functional dependencies (Lehnertz et al., 2020). The primary methods employed to analyze these dependencies include state-space interdependence (Arnhold et al., 1999; Faes et al., 2008a), correlation analyses (Li et al., 2009), and the application of Granger causality (GC) in both the time and frequency domains (Bressler and Seth, 2011). Information dynamics (Lizier, 2012) is a versatile framework that includes most of these approaches and has been widely used to characterize the interdependence of coupled systems in various fields, especially in physiology (Widjaja et al., 2015; Javorka et al., 2018). Entropy-based measures have been demonstrated to be effective in evaluating physiological interactions in cardiorespiratory (Widjaja et al., 2015), cardiovascular (Faes et al., 2019a), and cerebrovascular systems (Bari et al., 2023). These measures assess complexity (Pincus, 1991; Faes et al., 2017) and information stored within a process (Faes et al., 2019b), as well as directed information transfer between processes (Porta and Faes, 2013). Another key characteristic to assess in physiological interactions is the dynamic coupling between two processes which can be evaluated through the computation of the Mutual Information Rate (MIR) (Barà et al., 2023; Pinto et al., 2023). This information-theoretic measure quantifies the degree of information shared between two systems, providing a measure of their dynamic coupling strength and revealing the interdependencies within their dynamic behavior.

Nevertheless, the inherent variability in dynamic systems, particularly physiological ones, requires proper statistical tests to evaluate the significance of estimated information measures. Without such testing, the reliability of these measures remains uncertain. Typically, this assessment involves the use of surrogate analysis methods (Lucio et al., 2012; Lancaster et al., 2018). Surrogate data testing provides a versatile framework applicable to signals that arise from any physical system, allowing the exploration of fundamental questions about the system. Moreover, incorporating information dynamic measures, such as IS and MIR, into these methods enables the testing of specific hypotheses related to the observed data (Jamšek et al., 2010; Cliff et al., 2021). In typical surrogate analysis methods, a null hypothesis is formulated, assuming the absence of a specific characteristic to be tested in the observed data. Subsequently, surrogate data consistent with the null hypothesis is generated, and a discriminating statistic is computed for both the original and surrogate data. If the discriminating statistic of the original data significantly deviates from the surrogate distribution, the null hypothesis is rejected (Schreiber and Schmitz, 2000; Lancaster et al., 2018).

In time series analysis, several methods have emerged to assess the temporal statistical structure and the interaction between interconnected dynamic processes (Theiler et al., 1992; Palus, 1996; Popivanov and Mineva, 1999; Faes et al., 2008b; Lancaster et al., 2018; Calderón-Juárez et al., 2023). In this study, surrogate-based tests were implemented, using IS as a discriminating statistic, along with shuffling procedures and the Iterative Amplitude Adjusted Fourier Transform (IAAFT) algorithm (Schreiber and Schmitz, 1996) to assess the presence of self-dependencies and nonlinear dynamics in univariate processes. Additionally, these tests were extended to bivariate systems, incorporating MIR as a discriminating statistic, to investigate coupling and nonlinearity in coupled systems. All the proposed methods were validated through simulations involving both univariate processes and bivariate systems with linear and nonlinear dynamics. Finally, the framework is applied to a simple exemplary application in cardiovascular and cardiorespiratory time series, collected from healthy subjects monitored during different breathing conditions, to assess changes in the strength of the cardiorespiratory coupling and the extent of nonlinearities.

2 Framework for the surrogate data analysis of dynamic correlations and nonlinearity

Within the surrogate family, there are two distinct approaches: typical realizations and constrained realizations (Lancaster et al., 2018). In typical realizations, a model is fitted to the data, such as estimating the coefficients of an autoregressive (AR) model. Subsequently, Monte Carlo realizations of this model are generated for comparison with the data. However, a drawback of this method is that the user needs to have prior knowledge of the model and its parameters. On the other hand, constrained realizations are computed directly from the data, producing surrogates that replicate all characteristics of the original data except for the specific property being tested (Theiler and Prichard, 1996). Although the constrained realization approach, commonly known as surrogate data, is more widely adopted, it is essential to acknowledge that this method is specifically tailored for hypothesis testing. In contrast to typical realizations, it cannot be employed for the estimation of confidence intervals (Theiler and Prichard, 1996).

In this constrained realization approach, two fundamental steps need to be addressed: the selection of a suitable discriminating statistic and the proper formulation of the null hypothesis to be tested. This ensures that when rejection occurs, it is possible to conclude that the null hypothesis is not suitable to describe the data. Conversely, not rejecting the null hypothesis does not inherently mean its acceptance. Surrogates exclusively test the null hypothesis specified for them, without confirming its accuracy. Since surrogates do not always have a Gaussian distribution, it is more robust to use a non-parametric test where the significance level is determined, for

example, by the 95th percentile of the surrogate distribution. The required number of surrogates depends on whether a one-sided or two-sided test is being conducted. In a one-sided test, the null hypothesis is rejected only if the distribution of surrogate values of the discriminating statistic deviates from those calculated in the original data in one specified direction—either lower or higher. On the contrary, a two-sided test allows for testing the possibility of the discriminating statistic being either higher or lower in the surrogates than in the original data (Theiler et al., 1992; Schreiber and Schmitz, 2000).

The following sections provide a detailed overview of surrogate methods for testing the presence of specific statistical structures in both univariate and bivariate processes. Firstly, in Section 2.1, the discriminating statistics employed in this work, Information Storage (IS) and Mutual Information Rate (MIR), are described in detail, along with the nearest neighbor estimation of these information measures. Then, in Sections 2.2 and 2.3, respectively, methods for assessing the presence of self-dependencies and nonlinearities in univariate processes are introduced. Lastly, Sections 2.4 and 2.5 discuss, respectively, surrogates for testing the existence of dynamic coupling and the nonlinear dynamics in bivariate systems.

2.1 Discriminating statistics

2.1.1 Information-theoretic preliminaries

As information-theoretic analysis of random processes relies on information measures applied to random variables, we will begin by reviewing the entropy measures adopted for describing general static random variables. Subsequently, these measures are extended to stochastic processes representing networks of interacting dynamical systems, enabling the formulation of Information Storage (IS) and Mutual Information Rate (MIR), which are later used as discriminating statistics.

One of the most well-known and widely used information measures is entropy (Shannon, 1948; Cover, 1999). Entropy serves as a measure of uncertainty or disorder associated with a random variable. For a discrete random variable X with a probability mass function $p(x)$, the entropy $H(X)$ is defined by

$$H(X) = \mathbb{E}\left[\ln\frac{1}{p(x)}\right], \tag{1}$$

where $\mathbb{E}[.]$ represents the expectation operator. Since it is formulated in terms of the natural logarithm, the entropy is measured in natural units (*nats*). High entropy values indicate high uncertainty, while low entropy suggests more predictability.

Moving beyond single variables, conditional entropy (CE), denoted as $H(Y|X)$, measures the remaining uncertainty of a random variable Y given the knowledge of X and is defined as

$$H(Y|X) = -\mathbb{E}\left[\ln\left(\frac{p(x,y)}{p(x)}\right)\right], \tag{2}$$

where $p(x,y)$ is the joint probability mass function of X and Y.

The amount of information shared between two random variables can be measured by Mutual Information (MI), which is crucial for understanding the relationships and dependencies between variables (Cover, 1999). MI, denoted as $I(X;Y)$, between

X and Y is defined as the expected value of the pointwise mutual information

$$I(X;Y) = \mathbb{E}\left[\ln\left(\frac{p(x,y)}{p(x)p(y)}\right)\right]. \tag{3}$$

The logarithm term inside the expectation measures the ratio of the joint probability to the product of the marginals, reflecting the departure from independence. Positive values of mutual information indicate dependence, while zero suggests independence between X and Y.

2.1.2 Information measures in dynamic processes

This Section describes the use of the information measures defined in Section 2.1.1, applied by taking as arguments proper combinations of the present and past states of the stochastic processes representative of a network of interacting dynamical systems, to formulate a framework quantifying auto-dependencies in univariate processes through IS and dynamic coupling in bivariate systems via the MIR.

Consider two stationary, ergodic, potentially interacting stochastic processes, X and Y. Let us assume that the realization of system states is suitably described as a multivariate stationary stochastic process $S = [X, Y]$. In a temporal reference frame where n denotes the present time, Y_n represents the current state of Y, and $Y_n^- = [Y_{n-1}, Y_{n-2}, \ldots]$ describes its past. The same notation applies to the process X. This simple operation of separating the present from the past allows us to consider the flow of time and study causal interactions within and between processes by examining the statistical dependencies among these variables (Faes and Porta, 2014). Furthermore, by adopting the Markov assumption, which posits a finite-length memory for the investigated system, the past of each process is approximated by vector variables with dimension $1 \times q$, i.e., $X_n^q \approx [X_{n-1}, X_{n-2}, \ldots, X_{n-q}]$ and $Y_n^q \approx [Y_{n-1}, Y_{n-2}, \ldots, Y_{n-q}]$.

2.1.2.1 Information storage

IS can be interpreted as a measure of the process regularity and for the process X can be defined as (Faes et al., 2019b)

$$S_X = I\left(X_n; X_n^q\right) = \mathbb{E}\left[\ln\frac{p\left(x_n, x_{n-1}, \ldots, x_{n-q}\right)}{p(x_n)p\left(x_{n-1}, x_{n-2}, \ldots, x_{n-q}\right)}\right]. \tag{4}$$

Here, the expectation is calculated across various instances $(x_n, x_{n-1}, \ldots, x_{n-q})$ of the random variables $(X_n, X_{n-1}, \ldots X_{n-q})$. Considering the dynamic evolution of a time-evolving system, the concept of IS complements a widely recognized measure of system complexity, which is quantified as the entropy rate. This rate, for ergodic processes, is defined as the conditional entropy of the current state given its past states, denoted as $H_X = H(X_n \mid X_n^q)$ (Cover, 1999; Martins et al., 2020). Thus, IS can be rewritten as

$$S_X = H(X_n) - H_X. \tag{5}$$

Moreover, given that $H(X_n \mid X_n^q) = H(X_n, X_n^q) - H(X_n^q)$, the IS can be reformulated exclusively in terms of entropy, i.e.,

$$S_X = H(X_n) + H(X_n^q) - H(X_n, X_n^q). \tag{6}$$

It is worth noting that the IS for the Y process is defined in exactly the same way, simply by replacing X with Y in the previous

expressions. As the entropy measures previously defined, the IS is also expressed in *nats*.

2.1.2.2 Mutual information rate

The MIR of the bivariate process $\mathbf{S} = \{X, Y\}$ is defined as the limit, if it exists, of the rate at which the MI between dynamic sequences taken from X and Y increases over time (Komaee, 2020; Miao et al., 2020)

$$I_{X;Y} = \lim_{q \to \infty} \frac{1}{q} I\left(X_n^q; Y_n^q\right). \tag{7}$$

The MIR serves as a dynamic measure for the information shared per unit of time between two dynamical systems and was adopted in various forms to quantify dynamic interactions among physiological processes (Baptista and Kurths, 2008; Mijatovic et al., 2021; Faes et al., 2022). This measure of dynamic coupling can be decomposed in terms of other information-theoretic measures, offering valuable insights into the dynamics of each process and the coupling relationships within the bivariate system (Barà et al., 2023). A possible decomposition of MIR is

$$I_{X;Y} = H_X + H_Y - H_{X,Y}, \tag{8}$$

where H_X and H_Y denote the entropy rate of X and Y, respectively, and $H_{X,Y}$ their joint entropy rate. Exploiting the fact that the conditional entropy terms can be written as the difference between two entropies and substituting the above-defined terms in Eq. 8, the MIR can be reformulated as follows

$$\begin{aligned} I_{X;Y} = & H\left(X_n, X_n^q\right) - H\left(X_n^q\right) + H\left(Y_n, Y_n^q\right) - H\left(Y_n^q\right) \\ & - H\left(X_n, Y_n, X_n^q, Y_n^q\right) + H\left(X_n^q, Y_n^q\right), \end{aligned} \tag{9}$$

favoring the estimation of the MIR as the sum of entropy terms, measured also in *nats*.

2.1.3 Estimation of information measures

In this work, the most widely used non-parametric estimator for continuous random variables was applied, the k-nearest neighbor (KNN). Due to its ability to adjust resolution dynamically by altering the distance scale based on the underlying probability distribution (Victor, 2002), and its potential for bias compensation through distance projection (Kraskov et al., 2004), the nearest-neighbor technique has become increasingly popular in recent years for estimating entropy measures in the analysis of time series.

2.1.3.1 Nearest-neighbor estimation

The KNN estimation approach derives an approximation of the probability distribution by analyzing the statistical properties of the distances among neighboring points within the multidimensional spaces defined by the observed variables. This approach is grounded in the results presented in (Kozachenko and Leonenko, 1987), which assert that the mean Shannon information content of a general d-dimensional random variable V can be estimated from the set of N observations $\{v_1, v_2, \ldots, v_N\}$ of the variable as

$$-\mathbb{E}\left[\ln p\left(v_n\right)\right] = \psi(N) - \psi(k) + d\langle\ln \varepsilon_n\rangle, \tag{10}$$

where ψ denotes the digamma function, ε_n represents twice the distance between v_n and its kth nearest neighbor, determined using

the maximum norm and $\langle.\rangle$ denotes the average over all the N realizations of V. It is straightforward, from Eq. 10, to obtain the formula for the KNN estimate of the entropy for X_n, calculated based on the time series $\{x_1, x_2, \ldots, x_N\}$

$$H\left(X_n\right) = \psi(N) - \psi(k) + \langle\ln \varepsilon_n\rangle. \tag{11}$$

According to Eqs 6, 9, the IS and MIR can be computed as a combination of entropy terms. Nevertheless, these entropy terms are calculated in spaces with distinct dimensions, and applying the same neighbor search procedure uniformly across all spaces would yield distinct distance lengths when approximating the probability density in different dimensions. This divergence in distance lengths could introduce estimation biases that cannot be rectified by taking the entropy differences. Hence, in order to keep the same distance length in all explored spaces, the approach discussed in (Kraskov et al., 2004) was used. This method conducts a neighbor search only in the highest-dimensional space and then projects the distances identified in this space to the lower-dimensional spaces. This ensures that these distances serve as the range for counting neighbors in each respective space. Specifically, for the IS estimation, the KNN estimate of $H\left(X_n, X_n^q\right)$ is computed through the neighbor search

$$H\left(X_n, X_n^q\right) = \psi(N) - \psi(k) + (q+1)\langle\ln \varepsilon_n\rangle, \tag{12}$$

with ε_n representing twice the distance from $[x_n, x_n^q]$ to its kth nearest neighbor. Subsequently, based on the calculated distances ε_n, the entropies in the lower-dimensional spaces are assessed using a range search.

$$H\left(X_n^q\right) = \psi(N) - \langle\psi\left(N_{X_n^q} + 1\right)\rangle + q\langle\ln \varepsilon_n\rangle, \tag{13a}$$

$$H\left(X_n\right) = \psi(N) - \langle\psi\left(N_{X_n} + 1\right)\rangle + \langle\ln \varepsilon_n\rangle, \tag{13b}$$

where N_{X_n} and $N_{X_n^q}$ are the number of points whose distance from x_n and x_n^q, respectively, is smaller than $\varepsilon_n/2$. Therefore, IS is obtained subtracting Eq. 12 from the sum of Eq. 13

$$S_X = \psi(N) + \psi(k) - \langle\psi\left(N_{X_n^q} + 1\right)\rangle - \langle\psi\left(N_{X_n} + 1\right)\rangle. \tag{14}$$

Analogously, for the MIR estimation, considering the bivariate system $\mathbf{S} = \{X, Y\}$, the KNN estimate of the entropy in the higher space, i.e., $[X_n, Y_n, X_n^q, Y_n^q]$, is computed through the neighbor search as follows

$$H\left(X_n, Y_n, X_n^q, Y_n^q\right) = -\psi(k) + \psi(N) + 2(q+1)\langle\log\varepsilon_n\rangle \tag{15}$$

where, in this case, ε_n denotes twice the distance from $[x_n, y_n, x_n^q, y_n^q]$ to its kth nearest neighbor. Following this, the entropies in the lower-dimensional spaces are determined through a range search using ε_n. Particularly, in this framework, $H\left(X_n, X_n^q\right)$ and $H\left(X_n^q\right)$ are computed respectively as.

$$H\left(X_n, X_n^q\right) = \psi(N) - \langle\psi\left(N_{X_n X_n^q} + 1\right)\rangle + (q+1)\langle\ln \varepsilon_n\rangle, \tag{16a}$$

$$H\left(X_n^q\right) = \psi(N) - \langle\psi\left(N_{X_n^q} + 1\right)\rangle + q\langle\ln \varepsilon_n\rangle, \tag{16b}$$

where $N_{X_n X_n^q}$ and $N_{X_n^q}$ are the number of points whose distance from $[x_n, x_n^q]$ and x_n^q is smaller than $\varepsilon_n/2$, respectively. Similarly, $H\left(Y_n, Y_n^q\right)$ and $H\left(Y_n^q\right)$ can be derived by replacing X with Y in the corresponding above equations, thus obtaining.
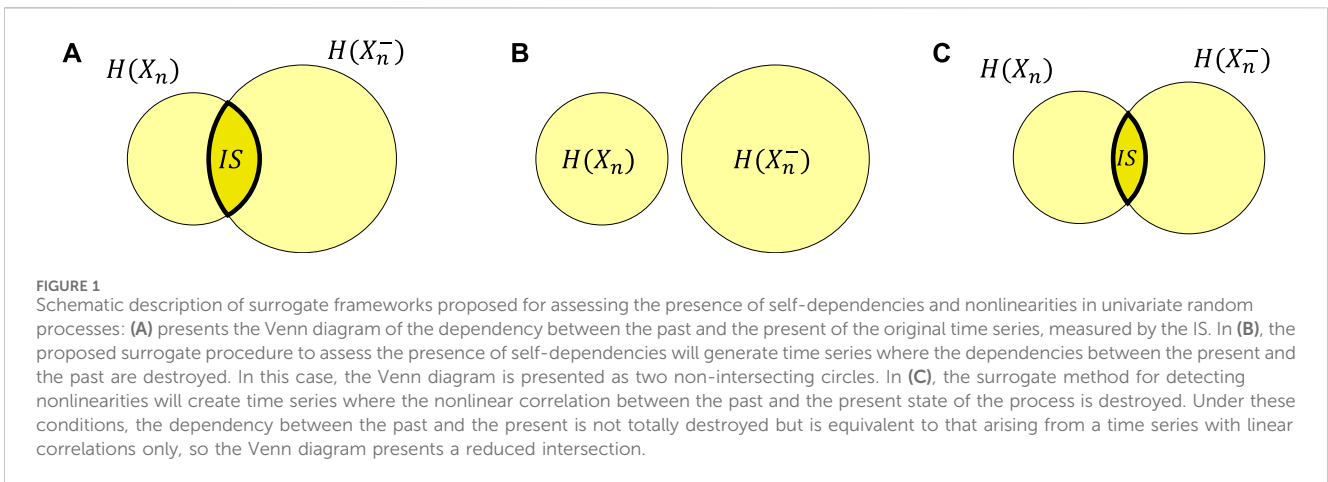
FIGURE 1
Schematic description of surrogate frameworks proposed for assessing the presence of self-dependencies and nonlinearities in univariate random processes: **(A)** presents the Venn diagram of the dependency between the past and the present of the original time series, measured by the IS. In **(B)**, the proposed surrogate procedure to assess the presence of self-dependencies will generate time series where the dependencies between the present and the past are destroyed. In this case, the Venn diagram is presented as two non-intersecting circles. In **(C)**, the surrogate method for detecting nonlinearities will create time series where the nonlinear correlation between the past and the present state of the process is destroyed. Under these conditions, the dependency between the past and the present is not totally destroyed but is equivalent to that arising from a time series with linear correlations only, so the Venn diagram presents a reduced intersection.

$$H(Y_n, Y_n^q) = \psi(N) - \langle\psi(N_{Y_n Y_n^q} + 1)\rangle + (q+1)\langle\ln\varepsilon_n\rangle, \quad (17a)$$

$$H(Y_n^q) = \psi(N) - \langle\psi(N_{Y_n^q} + 1)\rangle + q\langle\ln\varepsilon_n\rangle, \quad (17b)$$

$N_{Y_n Y_n^q}$ and $N_{Y_n^q}$ are the number of points whose distance from $[y_n, y_n^q]$ and y_n^q is smaller than $\varepsilon_n/2$. Then, the joint entropy of the past state of both processes can be computed as:

$$H(X_n^q, Y_n^q) = \psi(N) - \langle\psi(N_{X_n^q Y_n^q} + 1)\rangle + 2q\langle\log(\varepsilon_n)\rangle, \quad (18)$$

where $N_{X_n^q Y_n^q}$ denote the number of points whose distance from $[x_n^q, y_n^q]$ is smaller than $\varepsilon_n/2$. Finally, the KNN estimate of MIR is derived plugging Eqs 15–18 in Eq. 9, thus obtaining

$$I_{X;Y} = \psi(k) + \langle\psi(N_{X_n^q} + 1) + \psi(N_{Y_n^q} + 1) - \psi(N_{X_n X_n^q} + 1)$$
$$- \psi(N_{Y_n Y_n^q} + 1) - \psi(N_{X_n^q Y_n^q} + 1)\rangle. \quad (19)$$

2.2 Surrogates for auto-dependencies in univariate processes

Many physical and biological systems exhibit complex dynamic behaviors resulting from the presence of self-sustained oscillators, interacting subsystems, and feedback loops responding to both internal and external stimuli (Pincus, 1991; Donges et al., 2009; Chialvo, 2010). Therefore, assessing the existence of self-dependencies in these types of systems is crucial. In this context, for detecting auto-dependencies, in this work a surrogate method that employs the IS as a discriminating statistic is proposed. Let us consider the time series $\mathbf{x} = \{x_1, x_2, x_3, \ldots, x_N\}$ as a realization of length N of the univariate stochastic process X. Moreover, assuming an embedding dimension of q, the estimation of IS, as described in Section 2.1.2.1, relies on the $(N - q) \times (q + 1)$ observation matrix, defined as

$$\begin{bmatrix} x_{q+1} & x_q & x_{q-1} & \ldots & x_1 \\ x_{q+2} & x_{q+1} & x_q & \ldots & x_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_N & x_{N-1} & x_{N-2} & \ldots & x_{N-q} \end{bmatrix}. \quad (20)$$

This matrix plays a central role in the proposed surrogate method.

Null Hypothesis. The time series $\mathbf{x}$ comes from a process that does not exhibit self-dependencies; equivalently, $S_X = 0$.

Algorithm. Perform a random shuffle on the column corresponding to the present state of the time series $\mathbf{x}$, which is the first column of the observation matrix $[x_{q+1}, x_{q+2}, \ldots x_N]^T$ (T represents matrix transpose), thereby eliminating any dependency between the present samples and their past (see Figure 1B). After this shuffling process, the IS is estimated. This procedure is repeated M times to obtain the distribution of IS for the surrogates. Finally, a percentile-based test is performed, where the null hypothesis is rejected if, with a significance level α, the IS estimated for the original process exceeds the $(1 - \alpha)^{\text{th}}$ percentile of the surrogate distribution, indicating that the process does have self-dependencies.
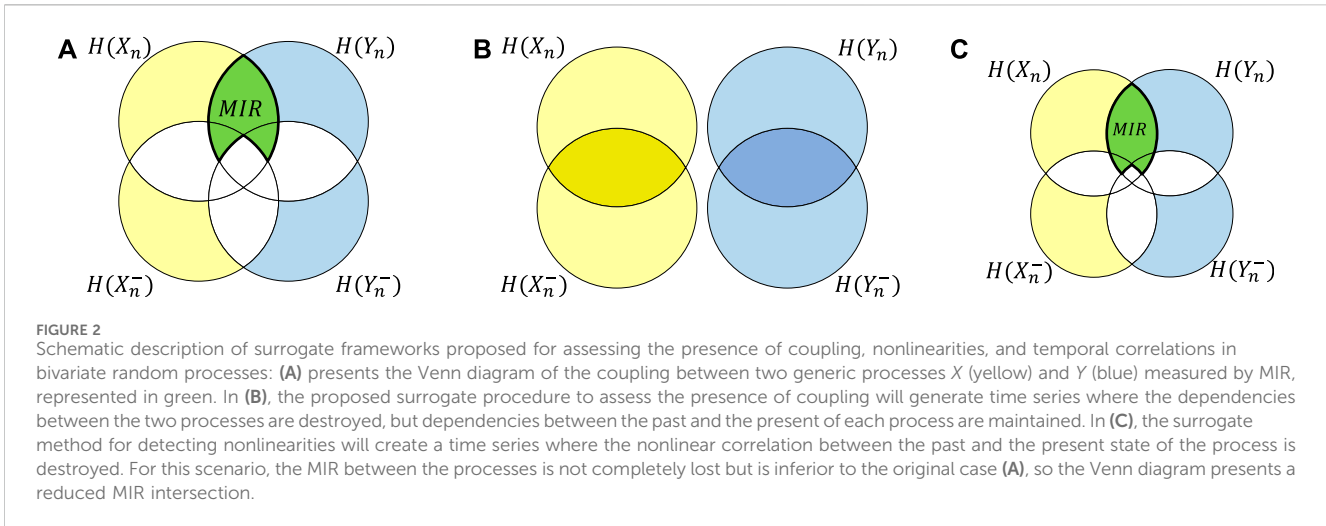
2.3 Surrogates for nonlinear dynamics in univariate processes

The presence of nonlinear dynamics was assessed using the Iterative Amplitude Adjusted Fourier Transform (IAAFT) surrogates introduced by (Schreiber and Schmitz, 1996). This method involves the iterative replacement of Fourier amplitudes with the correct values and rescaling the distribution to achieve a closer match between the distribution and the power spectrum in the original data and the surrogates (Lancaster et al., 2018).

Null Hypothesis. The time series $\mathbf{x}$ comes from a process that exclusively exhibit linear dynamics.

Algorithm. The algorithm involves an iterative process:

1. Store a sorted list of the values of the time series $\mathbf{x}$ and the squared amplitudes of its Fourier transform $S_k^2 = |\sum_{n=0}^{N-1} x_n e^{i2\pi kn/N}|^2$, with $k = 1, \ldots, N$.
2. Randomly shuffle (without replacement) $\mathbf{x}^{(0)}$, where the superscript $^{(0)}$ denotes the iteration number.
3. $\mathbf{x}^{(i)}$ is adjusted to match the desired sample power spectrum. This involves performing a Fourier transform on $\mathbf{x}^{(i)}$, replacing the squared amplitudes $S_k^{2,(i)}$ with S_k^2, and subsequently applying the inverse Fourier Transform. The phases of the complex Fourier components are retained. Therefore, this step ensures the correct spectrum, but it typically results in a modified distribution.

FIGURE 2
Schematic description of surrogate frameworks proposed for assessing the presence of coupling, nonlinearities, and temporal correlations in bivariate random processes: **(A)** presents the Venn diagram of the coupling between two generic processes X (yellow) and Y (blue) measured by MIR, represented in green. In **(B)**, the proposed surrogate procedure to assess the presence of coupling will generate time series where the dependencies between the two processes are destroyed, but dependencies between the past and the present of each process are maintained. In **(C)**, the surrogate method for detecting nonlinearities will create a time series where the nonlinear correlation between the past and the present state of the process is destroyed. For this scenario, the MIR between the processes is not completely lost but is inferior to the original case **(A)**, so the Venn diagram presents a reduced MIR intersection.

4. Arrange the derived series in a ranked order to precisely match the values of $\mathbf{x}$. Unfortunately, this results in a further modification of the spectrum in $\mathbf{x}^{(i+1)}$.

5. Repeat steps 3 and 4 to ensure that the power spectrum and distribution are as close to the original data as possible. In this work, seven iterations were used to avoid high computational costs, and indeed, this number of iterations is sufficient to observe a low percentage of false rejections (Schreiber and Schmitz, 1996).

This iterative procedure is repeated M times to obtain the distribution of the IS of the 100 surrogates. The *null hypothesis* is rejected, with a significance level of α, when the IS estimated in the original process exceeds the $(1 - \alpha)^{\text{th}}$ percentile of the surrogate distribution, indicating that the data come from a nonlinear process. Figure 1C represents graphically this surrogate procedure in terms of a Venn diagram. The IAFFT algorithm generates a time series in which the nonlinear correlation between the past and present state of the process is disrupted. In this case, the dependency between the past and the present is not entirely destroyed but is reduced compared to the original case presented in Figure 1A, resulting in a reduced intersection in the Venn diagram.

2.4 Surrogates for the presence of dynamic coupling in bivariate processes

Consider that $\mathbf{x} = \{x_1, x_2, \ldots, x_N\}$ and $\mathbf{y} = \{y_1, y_2, \ldots, y_N\}$ are realizations of length N for the stochastic processes X and Y, respectively. Similar to the approach with IS, described in Section 2.2, assuming an embedding dimension of q, the estimation of the MIR relies on an $(N - q) \times 2(q + 1)$ observation matrix given by

$$\begin{bmatrix} x_{q+1} & x_q & \cdots & x_1 & y_{q+1} & y_q & \cdots & y_1 \\ x_{q+2} & x_{q+1} & \cdots & x_2 & y_{q+2} & y_{q+1} & \cdots & y_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_N & x_{N-1} & \cdots & x_{N-q} & y_N & y_{N-1} & \cdots & y_{N-q} \end{bmatrix}, \quad (21)$$

which is exploited for generating surrogate time series.

Null Hypothesis. The time series $\mathbf{x}$ and $\mathbf{y}$ are realizations of independent processes, or, equivalently, $I_{X;Y} = 0$.

Algorithm. Shuffle the rows of the $(N - q) \times (q + 1)$ sub-matrix of the observation matrix defined as

$$\begin{bmatrix} x_{q+1} & x_q & \cdots & x_1 \\ x_{q+2} & x_{q+1} & \cdots & x_2 \\ \vdots & \vdots & \vdots & \vdots \\ x_N & x_{N-1} & \cdots & x_{N-q} \end{bmatrix}, \quad (22)$$

thus destroying the coupling while maintaining the internal dynamics of X and Y, as illustrated in Figure 2B. It is important to note that the matrix shown relates to process X for illustrative purposes only; likewise, choosing the sub-matrix for process Y would yield the same results. After this step, the shuffled observation matrix is used to estimate the MIR. This procedure is iterated M times to derive the surrogate distribution of the MIR, and the *null hypothesis* is rejected, with a significance level α, if the MIR value estimated in the original time series exceeds the $(1 - \alpha)^{\text{th}}$ percentile of the surrogate distribution, indicating that the processes are coupled.

2.5 Surrogates for the assessment of nonlinear interactions in bivariate processes

Fourier Transform (FT) surrogates, along with any other surrogate types wherein the phases are randomized, such as IAFFT introduced in Section 2.3, can be employed to investigate nonlinear dependencies in multivariate data. Let consider $\mathbf{x} = \{x_1, x_2, \ldots, x_N\}$ and $\mathbf{y} = \{y_1, y_2, \ldots, y_N\}$ as realizations of length N from the bivariate stochastic system $S = \{X, Y\}$.

Null Hypothesis. The time series $\mathbf{x}$ and $\mathbf{y}$ are realizations of a bivariate process exhibiting exclusively linear correlations, or, equivalently, the data constitutes a realization of a multivariate Gaussian process.

Algorithm. The algorithm is similar to the previously presented IAAFT in Section 2.3. The only difference in the procedure lies in step 2, where, instead of applying a random permutation to both

TABLE 1 SID toolbox function contents.

Function name	Description
data manipulation	
surr_SetLag	Creates the desired embedding vector
surr_ObsMat	Forms the observation matrix
KNN supporting functions	
nn_prepare	Construct a tree structure for nearest neighbor search
nn_search	Apply nearest neighbor search for a given query
range_search	Find points that are within a range for a given query
measure estimation	
surr_ISknn	K-nearest neighbor estimation of Information Storage
surr_MIRknn	K-nearest neighbor estimation of Mutual Information Rate
surrogate generation	
surr_ShuffColumn	Create surrogate observation matrices by applying joint random permutation of selected columns
surr_Iaaft	Create surrogate data for a given single process using the IAAFT algorithm
surr_IaaftBivariate	Create surrogate data for a given bivariate process using the IAAFT algorithm

series, the same random number in the range of $[0, 2\pi]$ is added to the phases of both processes (Lancaster et al., 2018). This approach preserves not only the power spectrum but also the cross-spectrum between signals, as highlighted by Prichard and Theiler (1994). In this case, the MIR between the processes is not completely lost but is inferior to the original, illustrated in Figure 2A, so the Venn diagram presents a reduced MIR intersection on Figure 2C.

As before, this procedure is repeated M times to obtain the distribution of the MIR of the M surrogates. The *null hypothesis* is rejected, with a significance level α, when the MIR estimated in the original process exceeds the 95th percentile of the surrogate distribution, indicating the presence of nonlinear dependencies between the processes constituting the bivariate system.

3 Toolbox description

In this section, we present the Surrogates for Information Dynamics measures toolbox (SID) with the utilized measures and surrogate techniques previously outlined. The toolbox is designed within MATLAB and is equipped with all essential functions required for the described investigation within this work, with some demo scripts in order to demonstrate its functionality.

The used functions are within the */SID/functions* folder, which can be grouped into four distinct categories:

- data manipulation,
- KNN supporting functions,
- measure estimation,
- surrogate generation algorithms.

Details for each individual provided function are covered in Table 1. Here, the supporting functions for efficient KNN algorithm

computations are obtained from (Lindner et al., 2011) and are given as compiled MEX files for Windows and Linux systems.

The parameters for the main estimation functions *surr_ISknn* and *surr_MIRknn* are the following:

- Y - the analyzed multivariate process, organized as a matrix with rows as samples and columns as processes,
- V - the assigned embedding vector,
- jj - index of the column for one investigated process,
- ii - index of the column for the second investigated process (MIR only),
- k - number of neighbors for the KNN estimation,
- *metric* - distance metric for the KNN estimation, and
- *surr* - parameter determining which type of surrogate to be created before estimating the measure.

To enable the creation of the proposed surrogates, the functionality is controlled with the *surr* parameter, where for *surr* = 1 the function applies the shuffling method on defined columns of the observation matrix, for *surr* = 2 applies the IAAFT method, and for *surr* = 0 it does not apply any surrogate method and the measure is estimated from the original data.

In order to demonstrate the functionality of the proposed methods, 2 demo scripts are placed within the root directory of the toolbox, */SID*, along with a single sample data obtained from (Faes et al., 2011) concerning the RR and respiration series.

In both *demo_IS_estimation* and *demo_MIR_estimation*, we demonstrate the steps to evaluate the significance of the measure and the presence of nonlinearities within the RR series and respiration series. The script goes through a basic collection of steps which can be condensed as:

- loading and applying z-score normalization on the data,
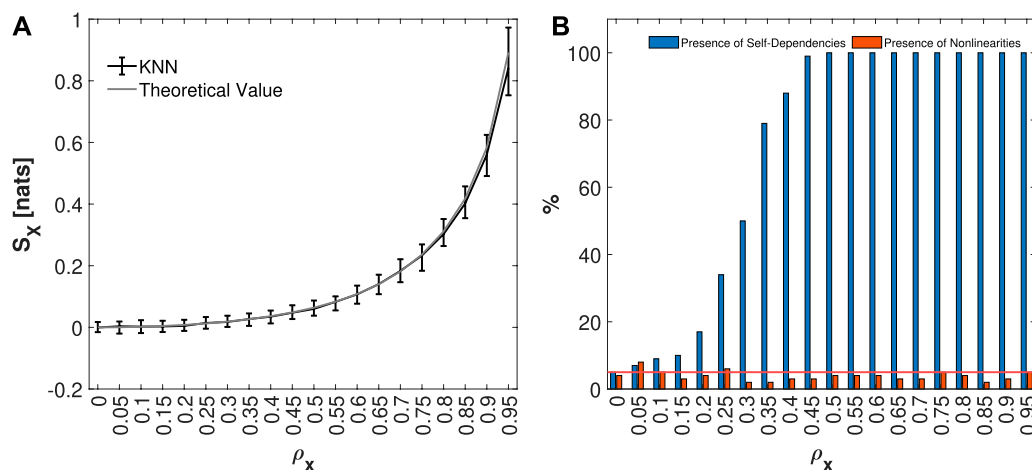
FIGURE 3
Distribution of IS KNN estimates, median and 5%–95% percentiles, computed on 100 simulations of a univariate AR model, varying the pole radius ρ_x from 0 to 0.95, is shown in panel **(A)**. In panel **(B)**, the plots depict the percentage of realizations, out of 100 realizations for each value of the pole radius ρ_x, in which the IS was detected as statistically significant (in blue), and for which nonlinearities (in orange) were found. The red line represents the 5% significance level.

- defining the parameters needed for the measure estimations,
- applying the measure estimation function with the *surr* parameter 0,
- applying the measure estimation function with the *surr* parameter 1 and 2, repeated *num_surrogates* times,
- obtaining the percentile values and comparing them to the originally estimated measure.

The potential users can easily adapt to the scripts with their data, parameters, or other changes. The SID toolbox is freely available at the Biosignals and Information Theory Laboratory (BIT Lab) website and https://github.com/helderpinto97/SID_Toolbox along with the real data application and the codes for the full simulation setup.

4 Simulations studies

In this section, the surrogate methods presented in Section 2 will be tested in simulations of both univariate and bivariate systems, including linear and nonlinear models. Specifically, for univariate processes, simulations were conducted using an AutoRegressive (AR) model and the Hénon-Hénon map to assess the presence of auto-dependencies and nonlinear dynamics. Subsequently, surrogates designed for detecting coupling and nonlinearities in bivariate systems were applied to a Vector AR model and the Coupled Hénon-Hénon Map. For each simulation model, a total of $M = 100$ surrogates were constructed and a significance level of $\alpha = 0.05$ was considered. For KNN estimation, $k = 10$ neighbors and an embedding dimension of $q = 2$ were applied. These are the values typically employed in the literature for KNN estimation. The choice of k involves a trade-off between variance and bias: increasing k reduces the variance but increases the bias. A low embedding was used to mitigate the bias on the estimates associated with the *curse of dimensionality* (Bellman, 1966).

4.1 Detection of self-dependencies and nonlinear dynamics in univariate processes

4.1.1 AutoRegressive model

The identification of auto dependencies and the presence of nonlinear dynamics is initially tested through the use of IS in conjunction with the surrogate analysis techniques introduced earlier in the AR model (Brockwell and Davis, 2009)

$$X_n = 2\rho_x \cos\left(2\pi f_x\right)X_{n-1} - \rho_x^2 X_{n-2} + \varepsilon_n, \tag{23}$$

where $f_x = 0.3$, the pole radius ρ_x was varied from 0 to 0.95 in steps of 0.05, to ensure the stability of the generated models (Barnett and Seth, 2014), and ε_n denotes white Gaussian noise with zero mean and unit variance. For each value of ρ_x, 100 realizations of length $N = 1000$ were generated.
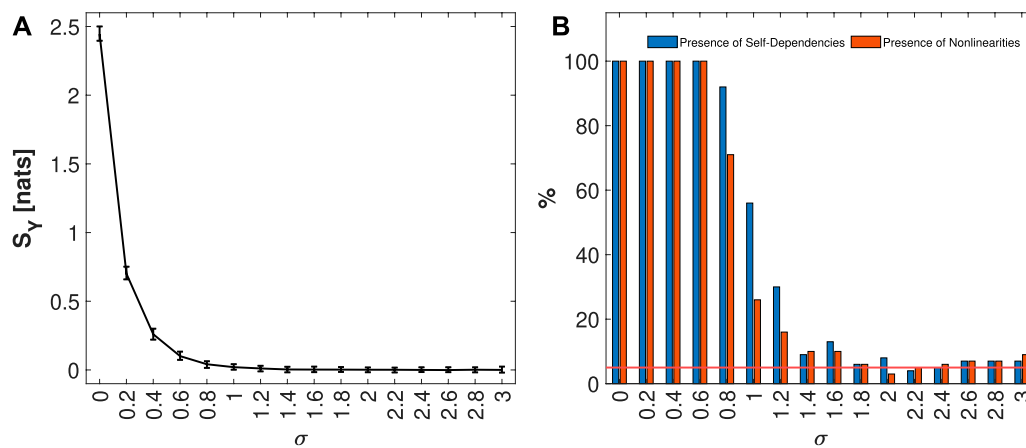
The median and 5%–95% percentile range of the IS values computed for the AR model, for each ρ_x considered, are presented in Figure 3A. Since the AR model is a Gaussian process it is possible to obtain the theoretical value of IS, which is represented, in the same panel, in gray color. In Figure 3B, the blue bars indicate the percentage of the 100 generated realizations in which IS was detected as statistically significant for each ρ_x value, while the orange bars represent the fraction of realizations in which nonlinearities were identified by the IAFFT algorithm. As ρ_x increases, the regularity of the process increases, leading to an increase in IS, as illustrated in Figure 3A. Correspondingly, Figure 3B shows an increase in the significance of the measure, represented by the blue bars. On the other hand, since the AR model is linear, only a small percentage of realizations exhibit nonlinearities, as indicated by the orange bars, showing that the rate of false detection oscillates around the nominal 5% level.

4.1.2 Hénon map

The univariate Hénon Map, as defined by (Hénon, 1976), is described by the following equation

$$X_n = 1 - 1.4X_{n-1}^2 + 0.3X_{n-2}. \tag{24}$$

FIGURE 4
Distribution of IS KNN estimates, median and 5%–95% percentile range, computed on 100 simulations of Y process defined in Eq. 25, varying the variance noise σ from 0 to 3, is shown in panel **(A)**. In panel **(B)**, the plots depict the percentage of realizations, out of 100 simulation runs for each value of σ, in which the IS was detected as statistically significant (in blue), and for which nonlinearities (in orange) were found. The red line represents the 5% significance level.

For this simulation, the process Y, defined as

$$Y = X + \xi, \qquad (25)$$

was used, where ξ denotes Gaussian noise with zero mean and variance σ. By incrementally increasing the noise variance from 0 to 3 in steps of 0.2, the past dependencies, as well as the nonlinear dynamics of the Hénon Map defined in Equation 24 (considering the classic parameters for chaotic behavior), are masked. This approach enables the assessment of surrogate methods proposed to test the statistical significance of IS and the presence of nonlinear dynamics. For each value of σ considered, 100 simulations of length $N = 1000$, using a burn-in period of $N_{trans} = 1000$ to avoid transient behavior, were computed using random initial conditions.

Figure 4A presents the median and 5%–95% percentile range of the IS values computed for the process Y formulated in Eq. 25, for each value of the noise variance σ considered. In Figure 4B, the blue bars indicate the percentage of the 100 generated realizations in which IS was detected as statistically significant, while the orange bars represent the fraction for which the IAFFT algorithm detected nonlinearities. The increase in noise variance conceals the dependence between the present and the past of the process, leading to a reduction of the information storage, as illustrated in Figure 4A. Consequently, Figure 4B shows a decrease in the presence of statistically significant self-dependencies. Additionally, the higher noise variance masks nonlinearities, contributing to the observed downward trend in the presence of nonlinear dynamics. In summary, these results, together with those discussed in Section 4.1.1, highlight the effectiveness of the surrogate methods proposed in this work for detecting self-dependencies and the presence of nonlinearities in univariate processes.

4.2 Detection of dynamic coupling and nonlinear interactions in bivariate processes
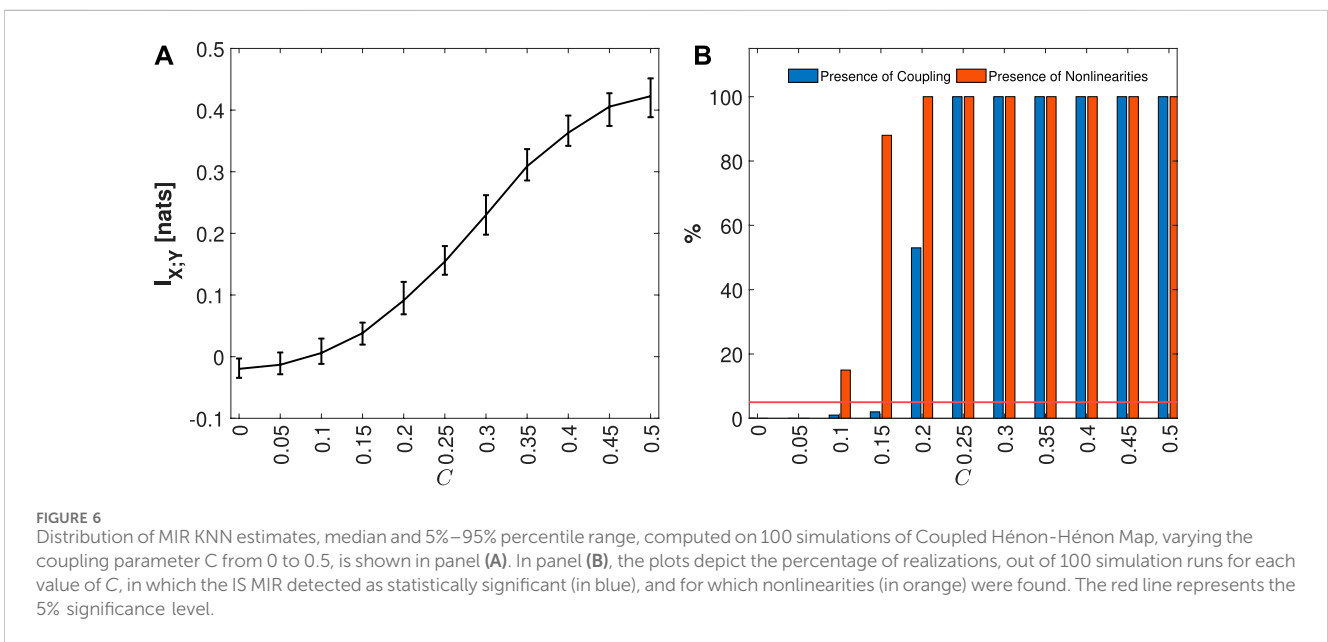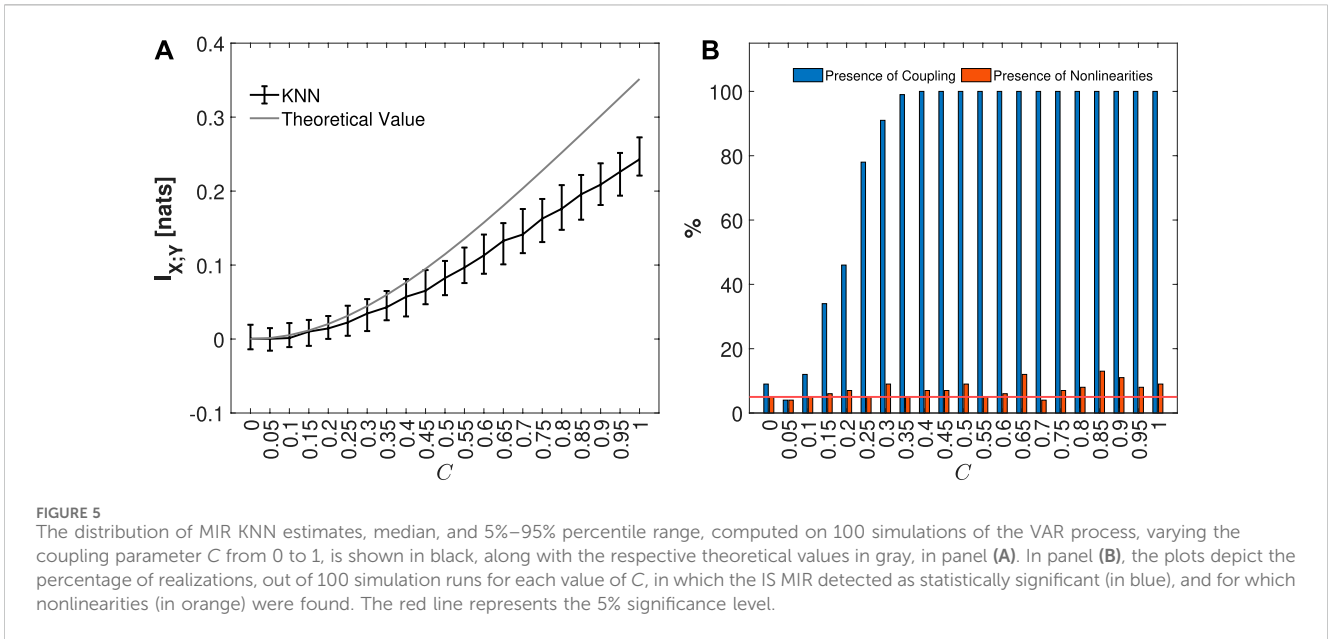
4.2.1 Vector AutoRegressive model

The surrogate procedures for the significance and the presence of nonlinear coupling using the MIR are first tested

in a stable bivariate Vector AutoRegressive (VAR) unidirectionally coupled model defined as (Barnett and Seth, 2014; Faes et al., 2015)

$$
\begin{aligned}
X_n &= 2\rho_x \cos\left(2\pi f_x\right)X_{n-1} - \rho_x^2 X_{n-2} + \varepsilon_n, \\
Y_n &= 2\rho_y \cos\left(2\pi f_y\right)Y_{n-1} - \rho_y^2 Y_{n-2} + C \cdot X_{n-2} + \xi_n,
\end{aligned}
\qquad (26)
$$

where ε_n and ξ_n are independent white Gaussian noises with zero mean and unit variance, and $\rho_x = 0.3, f_x = 0.3$ and $\rho_y = 0.3, f_y = 0.1$. In this simulation, X and Y exhibit second-order autoregressive behavior, characterized by two complex-conjugate poles with a modulus of $\rho_{x,y}$ and phases of $\pm\ 2\pi f_{x,y}$. This configuration establishes autonomous wide-band oscillations at frequencies of 0.3 and 0.1 Hz for X and Y, respectively. Directional connections are set from X to Y, at lag 2, modulated by the parameter C. For each value of the coupling parameter C, which varies between 0 and 1 in steps of 0.05, 100 realizations of length $N = 1000$ were generated.

The median and 5%–95% percentile range of the MIR values computed for the VAR model, considering each coupling parameter C, are presented in Figure 5A. In Figure 5B, the blue bars indicate the percentage of the 100 generated realizations in which MIR was detected as statistically significant, while the orange bars represent the fraction of simulations in which nonlinearities were identified. As expected, the increase in the parameter C leads to an increase in $I_{X;Y}$, as reported in Figure 5A. It is also important to note that for low values of C, negative values of MIR are observed. This can be attributed to the bias of the KNN estimator (Kraskov et al., 2004; Mijatovic et al., 2021), which tends to underestimate the MIR, resulting in values near zero becoming negative but not statistically significant, as illustrated in Figure 6B. Consequently, in Figure 5B, the presence of significant coupling increases along with the coupling parameter. On the other hand, due to the linearity of the VAR model, there is a low fraction of realizations, distributed around the nominal 5% error rate, where nonlinearities were detected for each C value considered.

FIGURE 5
The distribution of MIR KNN estimates, median, and 5%−95% percentile range, computed on 100 simulations of the VAR process, varying the coupling parameter C from 0 to 1, is shown in black, along with the respective theoretical values in gray, in panel **(A)**. In panel **(B)**, the plots depict the percentage of realizations, out of 100 simulation runs for each value of C, in which the IS MIR detected as statistically significant (in blue), and for which nonlinearities (in orange) were found. The red line represents the 5% significance level.

FIGURE 6
Distribution of MIR KNN estimates, median and 5%−95% percentile range, computed on 100 simulations of Coupled Hénon-Hénon Map, varying the coupling parameter C from 0 to 0.5, is shown in panel **(A)**. In panel **(B)**, the plots depict the percentage of realizations, out of 100 simulation runs for each value of C, in which the IS MIR detected as statistically significant (in blue), and for which nonlinearities (in orange) were found. The red line represents the 5% significance level.

4.2.2 Coupled Hénon-Hénon maps
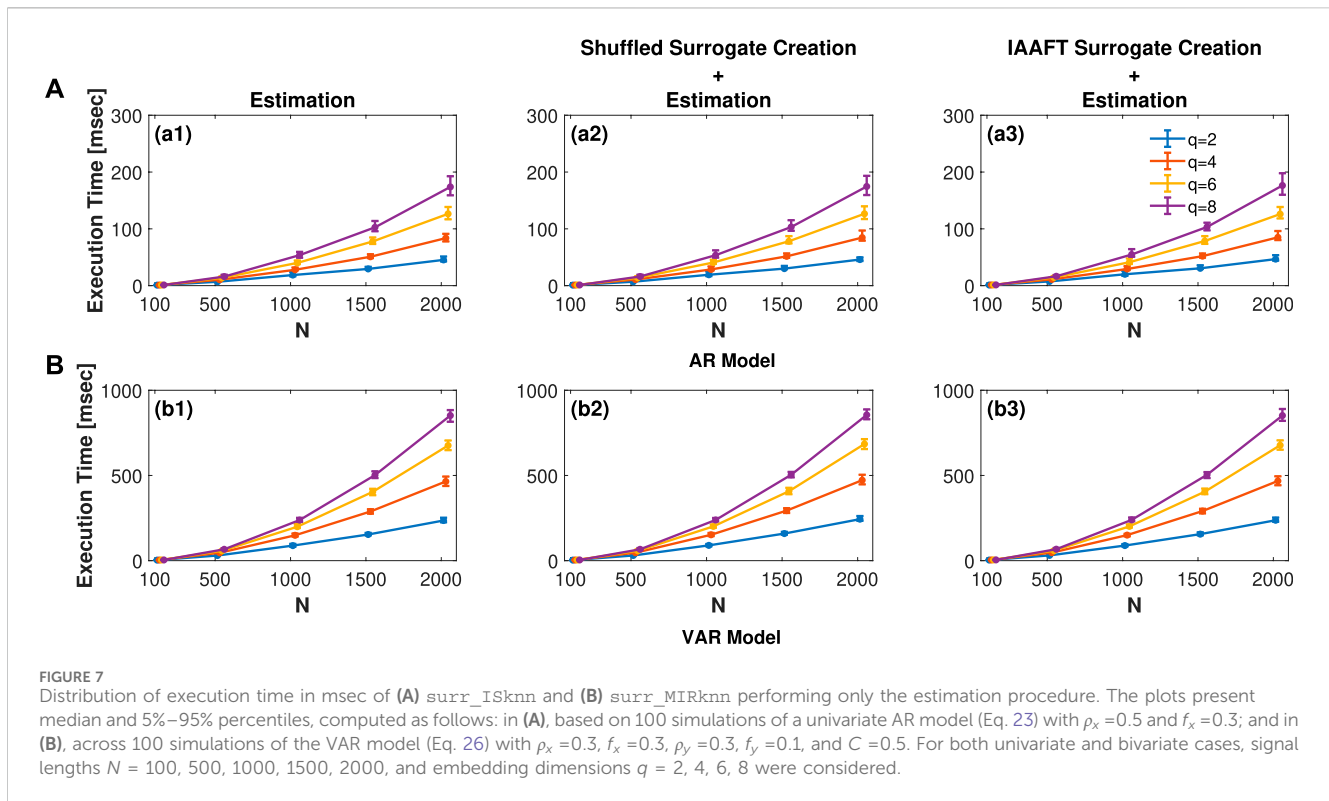
The unidirectionally coupled Hénon-Hénon system is described by the following equations (Sausedo-Solorio and Pisarchik, 2011)

$$X_n = 1.4 - X_{n-1}^2 + 0.3X_{n-2},$$
$$Y_n = 1.4 - \left(CX_{n-1}Y_{n-1} + (1-C)Y_{n-1}^2\right) + 0.3Y_{n-2}. \quad (27)$$

For this simulation, the coupling parameter C was varied from 0 to 0.5 in steps of 0.1. The selection of this interval for C is primarily based on the system manifestation of chaotic behavior within this range, without the occurrence of synchronization, which is observed at higher values of the coupling parameter (Krakovská et al., 2018). For each C, 100 realizations of length $N = 1000$ with a burn-in period of $N_{trans} = 1000$ to avoid the transitory behavior were

computed using random initial conditions. As the parameter C increases, the nonlinear coupling increases, on the other hand, the internal dynamics of Y is reduced.

Figure 6A presents the median and 5%–95% percentile range of the MIR values computed for the Coupled Hénon-Hénon Map, considering each coupling parameter C. In Figure 6B, the blue bars depict the percentage of the 100 generated realizations where MIR was detected as statistically significant, while the orange bars represent the fraction of simulations in which nonlinearities were identified. As expected, with the increase in the coupling parameter C, the $I_{X;Y}$ rises, as observed in Figure 6A. This observation is further highlighted in Figure 6B, illustrating the significance of this measure and the increased detection of nonlinear coupling. As in the case of the VAR model, for low coupling values, some realizations exhibit

FIGURE 7
Distribution of execution time in msec of **(A)** `surr_ISknn` and **(B)** `surr_MIRknn` performing only the estimation procedure. The plots present median and 5%–95% percentiles, computed as follows: in **(A)**, based on 100 simulations of a univariate AR model (Eq. 23) with ρ_x =0.5 and f_x =0.3; and in **(B)**, across 100 simulations of the VAR model (Eq. 26) with ρ_x =0.3, f_x =0.3, ρ_y =0.3, f_y =0.1, and C =0.5. For both univariate and bivariate cases, signal lengths N = 100, 500, 1000, 1500, 2000, and embedding dimensions q = 2, 4, 6, 8 were considered.

negative values due to the bias of the KNN estimator (Kraskov et al., 2004; Mijatovic et al., 2021).

4.3 Estimators complexity

In this section the execution time of the two main functions surr_ISknn and surr_MIRknn, introduced in Section 3, are analyzed in the AR and VAR model, introduced, respectively, in Sections 4.1.1, 4.2.1. The analysis was carried out on a PC with Windows 11 Home OS and an Intel(R) Core(TM) i7-10750H CPU @ 2.60GHz, with 16 GB of RAM and using MATLAB (R2023b) Update 7.
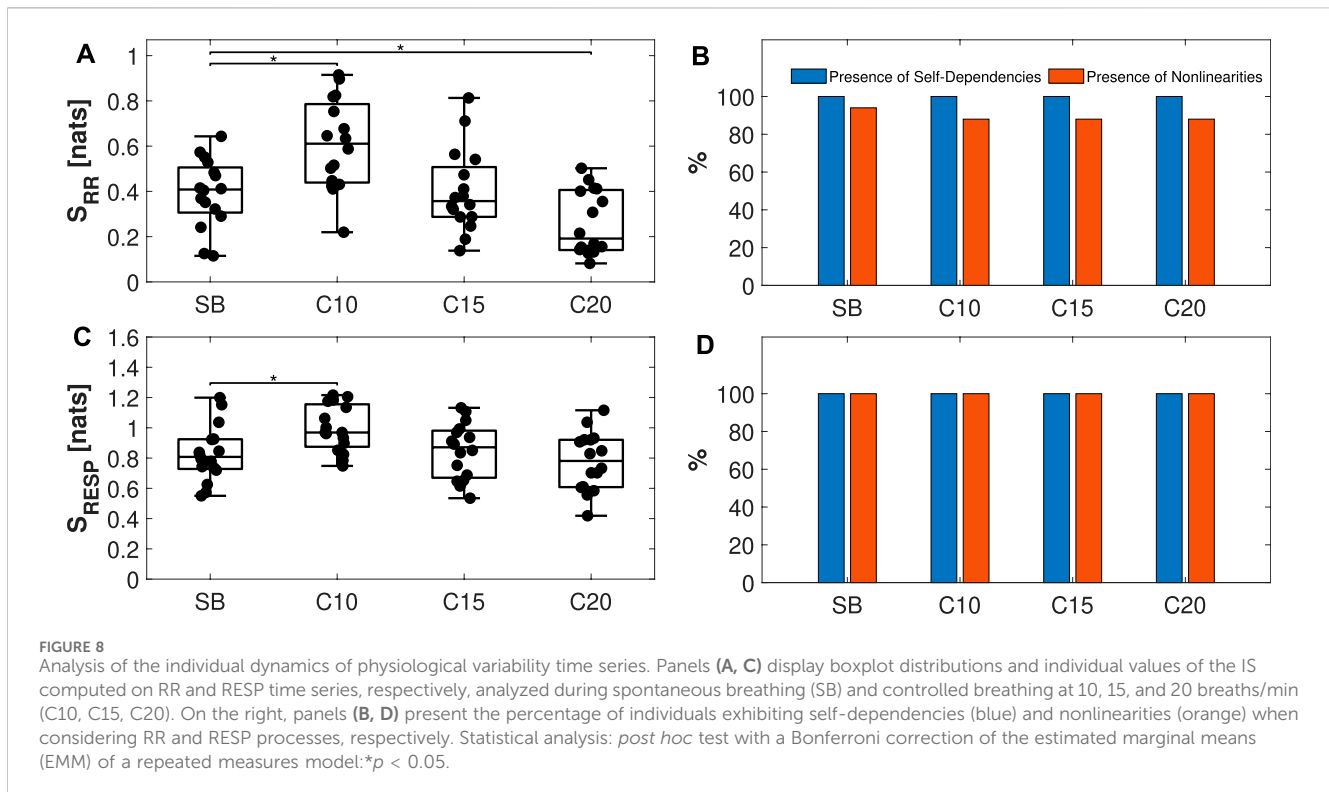
Figure 7 illustrates the distribution of execution times for surr_ISknn in panel **(a)** and surr_MIRknn in panel **(b)**. Each plot displays the median and 5%–95% percentiles, calculated as follows: in **(a)**, based on 100 simulations of a univariate AR model with $\rho_x = 0.5$ and $f_x = 0.3$; and in **(b)**, derived from 100 simulations of the VAR model with $\rho_x = 0.3$, $f_x = 0.3$, $\rho_y = 0.3$, $f_y = 0.1$, and $C = 0.5$. Both cases considered signal lengths N = 100, 500, 1000, 1500, 2000, and embedding dimensions q = 2, 4, 6, 8.

As expected, the execution time of `surr_MIRknn`, illustrated in Figure 7A, is higher than that of `surr_ISknn`, presented in Figure 7B, since the former is employed in bivariate systems where the complexity of calculations is superior. Additionally, for both functions, increasing the signal length results in more time consumed in the range search, leading to an expected increase in execution time, as observed. Similarly, when increasing the embedding dimension q, the dimension of the observation matrix becomes greater, consequently requiring more time for estimation.

One important point to stress is that most of the time consumed is for KNN estimation of the information measures. If some of the surrogate procedures proposed herein are used, a small increment in the results presented in Figure 7 will be observed. The magnitude of this increment is expected to be proportional to the number of surrogates utilized for the test. In particular, for the univariate case, under the worst-case scenario with N = 2000 and q = 8, we observe a median increase in execution time of approximately 1 ms for the shuffled surrogate creation followed by IS estimation, illustrated in Figure 7(a2). Conversely, when using IAAFT surrogates followed by IS estimation, there is an increment in execution time of approximately 3 ms, presented in Figure 7(a3). In the case of the VAR model, with N = 2000 and q = 8, there is a median increase in time of approximately 4 ms for the creation of shuffled surrogates along with the MIR estimation as presented in Figure 7(b2). Conversely, when utilizing IAAFT surrogates with MIR estimation to assess the presence of nonlinear dynamics, a median increase of approximately 2 ms is observed in Figure 7(b3).

5 Analysis of cardiorespiratory interactions

This section reports the application of the surrogate approaches proposed to evaluate the presence of self-dependencies, nonlinear dynamics, and coupling, as presented in Sections 2.2, 2.3, 2.4, 2.5, in a physiological dataset. The aim is to prove the effectiveness of the proposed approaches for assessing the cardiorespiratory interactions during both spontaneous and paced breathing.

FIGURE 8
Analysis of the individual dynamics of physiological variability time series. Panels **(A, C)** display boxplot distributions and individual values of the IS computed on RR and RESP time series, respectively, analyzed during spontaneous breathing (SB) and controlled breathing at 10, 15, and 20 breaths/min (C10, C15, C20). On the right, panels **(B, D)** present the percentage of individuals exhibiting self-dependencies (blue) and nonlinearities (orange) when considering RR and RESP processes, respectively. Statistical analysis: *post hoc* test with a Bonferroni correction of the estimated marginal means (EMM) of a repeated measures model:*$p < 0.05$.

5.1 Experimental protocol

Cardiorespiratory interactions were explored considering a historical database collected for analyzing short-term physiological variability during paced breathing (Faes et al., 2011). The experiments were conducted at the Cardiology Unit of S. Chiara Hospital in Trento, Italy. All participants provided informed consent, and the experimental protocol received approval from the hospital Ethical Committee. The dataset comprises physiological time series obtained from 16 young and healthy individuals observed in the supine position during four distinct experimental conditions: spontaneous breathing (SB), controlled breathing at 10 breaths/min (C10), at 15 breaths/min (C15), and at 20 breaths/min (C20). The acquired signals were the surface electrocardiographic signal (ECG, lead II) and the respiratory nasal flow (by differential pressure transducer). Signals were collected simultaneously and digitized at 1 kHz sampling rate and 12-bit precision. The sequence of heart periods (RR intervals) was extracted from the ECG signal, representing the time intervals between two consecutive R peaks in the ECG. The respiratory amplitude values (RESP) were extracted from the nasal respiration flow signal sampled at the onset of each heart period. The analysis involved synchronous time series consisting of 300 values for each subject and condition.
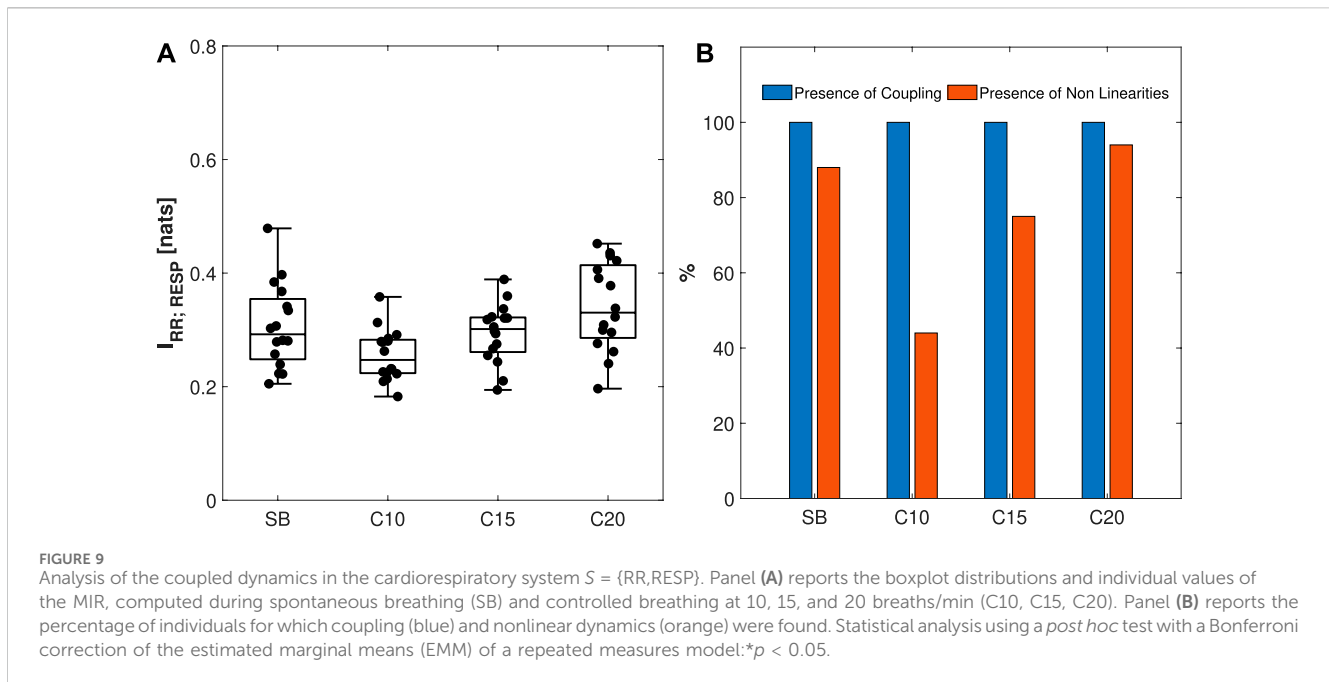
5.2 Statistical analysis

Significant changes in the IS and MIR across the pairs of experimental conditions were assessed using repeated measures models (Davis, 2002; Crowder and Hand, 2017; Davidian, 2017).

Estimated marginal means (EMM) were calculated between each paced breathing condition and the spontaneous (SB vs. C10, SB vs. C15, and SB vs. C20) (Searle and Milliken, 1980). A Z-test is then applied to determine the significance of these differences at a significance level of $p < 0.05$, with the Bonferroni correction applied for multiple comparisons ($n = 3$). The model's residuals were checked for whiteness. MATLAB software (MathWorks, Natick, MA, United States) was used to build the models and compute EMM.

5.3 Results of real data analysis

Figures 8A, C present the boxplot distributions with the overlapping individual values of S_{RR} and S_{RESP}, respectively. On the right, Figures 8B, D report the percentage of subjects presenting self-dependencies (in blue) and nonlinearities (in orange) when taking into account the RR and RESP time series, respectively. Results evidence a significant increase in IS from SB to C10, followed by a decrease when going from C10 to C20 with regard to RR, and a statistically significant difference from SB (panel (a)). On the contrary, although the trend is the same, only the increase from SB to C10 is found significant when taking into account RESP time series (panel (c)). The values across all considered phases were statistically significant, indicating the presence of self-dependencies on RR and RESP time series. Additionally, the majority of the tested individuals exhibit nonlinear dynamics, especially for RESP time series, as illustrated in Figure 8D, with a tendency towards decreasing the nonlinearity of the RR series while increasing the frequency of the paced breathing, as presented in Figure 8B.

FIGURE 9
Analysis of the coupled dynamics in the cardiorespiratory system S = {RR,RESP}. Panel **(A)** reports the boxplot distributions and individual values of the MIR, computed during spontaneous breathing (SB) and controlled breathing at 10, 15, and 20 breaths/min (C10, C15, C20). Panel **(B)** reports the percentage of individuals for which coupling (blue) and nonlinear dynamics (orange) were found. Statistical analysis using a *post hoc* test with a Bonferroni correction of the estimated marginal means (EMM) of a repeated measures model:*$p < 0.05$.

In Figure 9A, the distribution and individual values of $I_{RR;RESP}$ for the four experimental conditions are presented. The percentage of subjects exhibiting significant coupling (in blue) and nonlinearities (in orange) is reported in Figure 9B. Figure 9A shows a slight not statistically significant decrease from the SB to the C10 phase followed by the recovery of MIR values in C15 and C20. Nonetheless, all the MIR estimates are statistically significant, as illustrated in Figure 9B. With regard to the presence of nonlinearities (panel (b)), a marked decrease is observed for C10 if compared to SB, followed by a recovery when going to C15 and C20.

5.4 Discussion

Firstly, the presence of self-dependencies and nonlinear dynamics was individually analyzed for each of the processes, RR and RESP, using IS. Physiologically, the higher IS detected during C10, as observed in Figures 8A, C, suggests less complex RR and RESP dynamics (Radhakrishna et al., 2000). This observation may be linked to the mechanism of Respiratory Sinus Arrhythmia (RSA), i.e., modulation of heart rate with respiration. The RSA mechanism is more prominent during forced ventilation at low breathing rates (Saul et al., 1989), and its effects, compared to spontaneous breathing, tend to diminish when increasing respiratory rate (Song and Lehrer, 2003). Furthermore, RSA and other cardiovascular control mechanisms, such as cardio-ventilatory coupling and synchronization of respiratory stroke volume (Elstad et al., 2018), not only produce alterations in the complexity of respiratory dynamics but also similar variations in the complexity of cardiac dynamics. Nonetheless, a difference is found with regard to the nonlinearities between the two systems: controlled breathing produces a slight decrease of nonlinearities in cardiac dynamics with breathing rate, which is instead not observed in respiratory dynamics (Fang et al., 2008; Dimitriev et al., 2019).

The analysis of the coupling of the bivariate system S = {RR, $RESP$} using MIR highlights that the coupling between RR and RESP is not lost during different paced breathing rates (Mary et al., 2018), being the slight decrease observed in Figure 9 from the SB to the C10 phase not significant, and is followed by a complete recovery in C15 and C20. Such results are similar to what was reported in (Barà et al., 2023). A decrease in nonlinearities was observed in the C10 phase when compared with the other experimental conditions. This suggests that lower breathing rates reduce the nonlinearity of cardiorespiratory coupling (Fang et al., 2008), reflecting the possible effect of entrainment between low-frequency and high-frequency (respiration-induced) oscillations of HRV. This entrainment causes the entire HRV pattern to be centered around a single (broad band) stochastic oscillation, resulting in more linear dynamics. The trend differs when comparing $I_{RR;RESP}$ and S_{RR} in terms of nonlinearities: RR nonlinearities appear to be mainly associated with the controlled breathing condition and seem less related to the frequency rate, unlike $I_{RR;RESP}$ nonlinearities, which appear instead to depend only on the breath rate.

6 Conclusion

This study introduces surrogate approaches to examine specific statistical properties in univariate and bivariate dynamic processes. These approaches explore self-dependencies and nonlinearities in univariate processes, along with coupling and nonlinearities in bivariate systems. The proposed methods were validated through simulations and subsequently applied to investigate cardiorespiratory interactions from physiological time series.

The simulation results revealed that the proposed approaches successfully detect key dynamical characteristics of stochastic systems related to the internal properties of individual systems and the coupling structure between two systems, as well as to the nature (linear/nonlinear) of the underlying dynamics. Moreover, the application to real data has allowed to infer interesting different behaviors with regard

to cardiorespiratory interactions during spontaneous and paced breathing. In particular, controlled breathing increases the predictability of both RR and RESP dynamics. A decrease in nonlinearities when increasing the breathing rate is observed in cardiac, but not in respiratory dynamics. Moreover, the different paced breathing rates do not alter the cardiorespiratory coupling, while its nonlinearities are reduced during lower breathing rates.

Future developments will aim to test the surrogate methods discussed herein on various biosignals in the context of network physiology (Ivanov, 2021), particularly focusing on the analysis of brain-heart interactions (Koutlis et al., 2021; Antonacci et al., 2023; Varley et al., 2023). The proposed integrated approach for evaluating dynamic dependencies and nonlinearities in univariate and bivariate time series, implemented with the provided toolbox, will foster the assessment of the statistical temporal structure in coupled processes within Network Physiology and its related fields.

Data availability statement

The data analyzed in this study is subject to the following licenses/restrictions: The physiological raw data can be shared by the corresponding author upon request if data privacy can be guaranteed according to the rules of the European General Data Protection Regulation (EU GDPR). Requests to access these datasets should be directed to LF, luca.faes@unipa.it.

Ethics statement

The studies involving humans were approved by the Cardiology Unit of S. Chiara Hospital in Trento, Italy. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

HP: Software, Visualization, Writing–original draft, Writing–review and editing, Validation. IL: Software, Validation, Visualization, Writing–original draft, Writing–review and editing. YA: Methodology, Writing–review and editing, Validation. RP: Writing–review and editing. DG: Software, Validation, Writing–review and editing. CB: Software, Validation, Writing–review and editing. LF: Conceptualization, Methodology,

Supervision, Writing–review and editing. AR: Conceptualization, Supervision, Writing–review and editing.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) LF, RP, AR and YA declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

References

Antonacci, Y., Barà, C., Zaccaro, A., Ferri, F., Pernice, R., and Faes, L. (2023). Time-varying information measures: an adaptive estimation of information storage with application to brain-heart interactions. *Front. Netw. Physiology* 3, 1242505. doi:10.3389/fnetp.2023.1242505

Arnhold, J., Grassberger, P., Lehnertz, K., and Elger, C. (1999). A robust method for detecting interdependences: application to intracranially recorded eeg. *Phys. D. Nonlinear Phenom.* 134, 419–430. doi:10.1016/S0167-2789(99)00140-2

Baptista, M., and Kurths, J. (2008). Transmission of information in active networks. *Phys. Rev. E* 77, 026205. doi:10.1103/PhysRevE.77.026205

Barà, C., Sparacino, L., Pernice, R., Antonacci, Y., Porta, A., Kugiumtzis, D., et al. (2023). Comparison of discretization strategies for the model-free information-theoretic assessment of short-term physiological interactions. *Chaos An Interdiscip. J. Nonlinear Sci.* 33, 033127. doi:10.1063/5.0140641

Barabási, A. L. (2013). Network science. *Philosophical Trans. R. Soc. A Math. Phys. Eng. Sci.* 371, 20120375. doi:10.1098/RSTA.2012.0375

Bari, V., Gelpi, F., Cairo, B., Anguissola, M., Pugliese, S., Maria, B. D., et al. (2023). Characterization of cardiovascular and cerebrovascular controls via spectral causality analysis in patients undergoing surgical aortic valve replacement during a three-month follow-up. *Physiol. Meas.* 44, 094001. doi:10.1088/1361-6579/acf992

Barnett, L., and Seth, A. K. (2014). The mvgc multivariate granger causality toolbox: a new approach to granger-causal inference. *J. Neurosci. Methods* 223, 50–68. doi:10.1016/j.jneumeth.2013.10.018

Bashan, A., Bartsch, R. P., Kantelhardt, J. W., Havlin, S., and Ivanov, P. C. (2012). Network physiology reveals relations between network topology and physiological function. *Nat. Commun.* 3 (1), 702–709. doi:10.1038/ncomms1705

Bellman, R. (1966). Dynamic programming. *Science* 153, 34–37. doi:10.1126/science.153.3731.34

Bressler, S. L., and Seth, A. K. (2011). Wiener–granger causality: a well established methodology. *NeuroImage* 58, 323–329. doi:10.1016/j.neuroimage.2010.02.059

Brockwell, P. J., and Davis, R. A. (2009). *Time series: theory and methods*. Springer science and business media. doi:10.1007/978-1-4419-0320-4

Calderón-Juárez, M., Cruz-Vega, I. B., González-Gómez, G. H., and Lerma, C. (2023). Nonlinear dynamics of heart rate variability after acutely induced myocardial ischemia by percutaneous transluminal coronary angioplasty. *Entropy* 25, 469. doi:10.3390/e25030469

Chialvo, D. R. (2010). Emergent complex neural dynamics. *Nat. Phys.* 6 (10), 744–750. doi:10.1038/nphys1803

Cliff, O. M., Novelli, L., Fulcher, B. D., Shine, J. M., and Lizier, J. T. (2021). Assessing the significance of directed and multivariate measures of linear dependence between time series. *Phys. Rev. Res.* 3, 013145. doi:10.1103/PHYSREVRESEARCH.3.013145

Cover, T. M. (1999). *Elements of information theory*. John Wiley and Sons. doi:10.1002/047174882X

Crowder, M. J., and Hand, D. J. (2017). *Analysis of repeated measures*. New York: Routledge. doi:10.1201/9781315137421

Davidian, M. (2017). *Nonlinear models for repeated measurement data*. New York: Routledge. doi:10.1201/9780203745502

Davis, C. S. (2002). *Statistical methods for the analysis of repeated measurements*. New York: Springer. doi:10.1007/B97287

Dimitriev, D. A., Saperova, E. V., Dimitriev, A. D., Indeikina, O. S., and Remizova, N. M. (2019). The effect of breathing at the resonant frequency on the nonlinear dynamics of heart rate. *Hum. Physiol.* 45, 54–61. doi:10.1134/S0362119719010067

Donges, J. F., Zou, Y., Marwan, N., and Kurths, J. (2009). Complex networks in climate dynamics. *Eur. Phys. J. Special Top.* 174 (1), 157–179. doi:10.1140/epjst/e2009-01098-2

Elstad, M., O'Callaghan, E. L., Smith, A. J., Ben-Tal, A., and Ramchandra, R. (2018). Cardiorespiratory interactions in humans and animals: rhythms for life. *Am. J. Physiology-Heart Circulatory Physiology* 315, H6–H17. doi:10.1152/ajpheart.00701.2017

Faes, L., Gómez-Extremera, M., Pernice, R., Carpena, P., Nollo, G., Porta, A., et al. (2019a). Comparison of methods for the assessment of nonlinearity in short-term heart rate variability under different physiopathological states. *Chaos An Interdiscip. J. Nonlinear Sci.* 29, 123114. doi:10.1063/1.5115506

Faes, L., Mijatovic, G., Antonacci, Y., Pernice, R., Barà, C., Sparacino, L., et al. (2022). A new framework for the time- and frequency-domain assessment of high-order interactions in networks of random processes. *IEEE Trans. Signal Process.* 70, 5766–5777. doi:10.1109/TSP.2022.3221892

Faes, L., Nollo, G., and Porta, A. (2011). Information domain approach to the investigation of cardio-vascular, cardio-pulmonary, and vasculo-pulmonary causal couplings. *Front. Physiology* 2, 80. doi:10.3389/fphys.2011.00080

Faes, L., Pereira, M. A., Silva, M. E., Pernice, R., Busacca, A., Javorka, M., et al. (2019b). Multiscale information storage of linear long-range correlated stochastic processes. *Phys. Rev. E* 99, 032115. doi:10.1103/PhysRevE.99.032115

Faes, L., and Porta, A. (2014). *Conditional entropy-based evaluation of information dynamics in physiological systems*. Berlin, Heidelberg: Springer Berlin Heidelberg, 61–86. doi:10.1007/978-3-642-54474-3_3

Faes, L., Porta, A., Javorka, M., and Nollo, G. (2017). Efficient computation of multiscale entropy over short biomedical time series based on linear state-space models. *Complexity* 2017, 1–13. doi:10.1155/2017/1768264

Faes, L., Porta, A., and Nollo, G. (2008a). Mutual nonlinear prediction as a tool to evaluate coupling strength and directionality in bivariate time series: comparison among different strategies based on *k* nearest neighbors. *Phys. Rev. E* 78, 026201. doi:10.1103/PhysRevE.78.026201

Faes, L., Porta, A., and Nollo, G. (2015). Information decomposition in bivariate systems: theory and application to cardiorespiratory dynamics. *Entropy* 17, 277–303. doi:10.3390/e17010277

Faes, L., Zhao, H., Chon, K. H., and Nollo, G. (2008b). Time-varying surrogate data to assess nonlinearity in nonstationary time series: application to heart rate variability. *IEEE Trans. Biomed. Eng.* 56, 685–695. doi:10.1109/TBME.2008.2009358

Fang, Y., Sun, J.-T., Li, C., Poon, C.-S., and Wu, G.-Q. (2008). Effect of different breathing patterns on nonlinearity of heart rate variability. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* 2008, 3220–3223. doi:10.1109/IEMBS.2008.4649889

Hénon, M. (1976). A two-dimensional mapping with a strange attractor. *Commun. Math. Phys.* 50, 69–77. doi:10.1007/BF01608556

Ivanov, P. C. (2021). The new field of network physiology: building the human physiolome. *Front. Netw. Physiology* 1, 711778. doi:10.3389/fnetp.2021.711778

Ivanov, P. C., Liu, K. K., Lin, A., and Bartsch, R. P. (2017). Network physiology: from neural plasticity to organ network interactions. *Springer Proc. Phys.* 191, 145–165. doi:10.1007/978-3-319-47810-4_12

Jamšek, J., Paluš, M., and Stefanovska, A. (2010). Detecting couplings between interacting oscillators with time-varying basic frequencies: instantaneous wavelet bispectrum and information theoretic approach. *Phys. Rev. E* 81, 036207. doi:10.1103/PhysRevE.81.036207

Javorka, M., Krohova, J., Czippelova, B., Turianikova, Z., Lazarova, Z., Wiszt, R., et al. (2018). Towards understanding the complexity of cardiovascular oscillations: insights from information theory. *Comput. Biol. Med.* 98, 48–57. doi:10.1016/j.compbiomed.2018.05.007

Komaee, A. (2020). Mutual information rate between stationary Gaussian processes. *Results Appl. Math.* 7, 100107. doi:10.1016/j.rinam.2020.100107

Koutlis, C., Kimiskidis, V. K., and Kugiumtzis, D. (2021). Comparison of causality network estimation in the sensor and source space: simulation and application on eeg. *Front. Netw. Physiology* 1, 706487. doi:10.3389/fnetp.2021.706487

Kozachenko, L. F., and Leonenko, N. N. (1987). Sample estimate of the entropy of a random vector. *Probl. Peredachi Inf.* 23, 9–16.

Krakovská, A., Jakubík, J., Chvosteková, M., Coufal, D., Jajcay, N., and Paluš, M. (2018). Comparison of six methods for the detection of causality in a bivariate time series. *Phys. Rev. E* 97, 042207. doi:10.1103/PHYSREVE.97.042207

Kraskov, A., Stögbauer, H., and Grassberger, P. (2004). Estimating mutual information. *Phys. Rev. E* 69, 066138. doi:10.1103/PhysRevE.69.066138

Lancaster, G., Iatsenko, D., Pidde, A., Ticcinelli, V., and Stefanovska, A. (2018). Surrogate data for hypothesis testing of physical systems. *Phys. Rep.* 748, 1–60. doi:10.1016/j.physrep.2018.06.001

Lehnertz, K., Bröhl, T., and Rings, T. (2020). The human organism as an integrated interaction network: recent conceptual and methodological challenges. *Front. Physiology* 11, 598694. doi:10.3389/fphys.2020.598694

Li, K., Guo, L., Nie, J., Li, G., and Liu, T. (2009). Review of methods for functional brain connectivity detection using fmri. *Comput. Med. Imaging Graph.* 33, 131–139. doi:10.1016/j.compmedimag.2008.10.011

Lindner, M., Vicente, R., Priesemann, V., and Wibral, M. (2011). Trentool: a matlab open source toolbox to analyse information flow in time series data with transfer entropy. *BMC Neurosci.* 12, 119. doi:10.1186/1471-2202-12-119

Lizier, J. T. (2012). *The local information dynamics of distributed computation in complex systems*. Springer Science and Business Media.

Lucio, J. H., Valdés, R., and Rodríguez, L. R. (2012). Improvements to surrogate data methods for nonstationary time series. *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.* 85, 056202. doi:10.1103/PhysRevE.85.056202

Martins, A., Pernice, R., Amado, C., Rocha, A. P., Silva, M. E., Javorka, M., et al. (2020). Multivariate and multiscale complexity of long-range correlated cardiovascular and respiratory variability series. *Entropy* 22, 315. doi:10.3390/e22030315

Mary, M. C. H., Singh, D., and Deepak, K. K. (2018). Impact of respiration on cardiovascular coupling using granger causality analysis in healthy subjects. *Biomed. Signal Process. Control* 43, 196–203. doi:10.1016/J.BSPC.2018.03.008

Miao, H., Zhang, F., and Tao, R. (2020). Mutual information rate of nonstationary statistical signals. *Signal Process.* 171, 107531. doi:10.1016/j.sigpro.2020.107531

Mijatovic, G., Antonacci, Y., Loncar-Turukalo, T., Minati, L., and Faes, L. (2021). An information-theoretic framework to measure the dynamic interaction between neural spike trains. *IEEE Trans. Biomed. Eng.* 68, 3471–3481. doi:10.1109/tbme.2021.3073833

Palus, M. (1996). Nonlinearity in normal human eeg: cycles, temporal asymmetry, nonstationarity and randomness, not chaos. *Biol. Cybern.* 75, 389–396. doi:10.1007/s004220050304

Pincus, S. M. (1991). Approximate entropy as a measure of system complexity. *Proc. Natl. Acad. Sci.* 88, 2297–2301. doi:10.1073/pnas.88.6.2297

Pinto, H., Antonacci, Y., Pernice, R., Barà, C., Javorka, M., Faes, L., et al. (2023). Decomposing the mutual information rate of heart period and respiration variability series to assess cardiorespiratory interactions. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* 2023, 1–4. doi:10.1109/EMBC40787.2023.10341174

Popivanov, D., and Mineva, A. (1999). Testing procedures for non-stationarity and non-linearity in physiological signals. *Math. Biosci.* 157, 303–320. doi:10.1016/S0025-5564(98)10088-3

Porta, A., and Faes, L. (2013). Assessing causality in brain dynamics and cardiovascular control. *Philosophical Trans. R. Soc. A Math. Phys. Eng. Sci.* 371, 20120517. doi:10.1098/rsta.2012.0517

Prichard, D., and Theiler, J. (1994). Generating surrogate data for time series with several simultaneously measured variables. *Phys. Rev. Lett.* 73, 951–954. doi:10.1103/PhysRevLett.73.951

Radhakrishna, R. K., Dutt, D. N., and Yeragani, V. K. (2000). Nonlinear measures of heart rate time series: influence of posture and controlled breathing. *Aut. Neurosci. Basic Clin.* 83, 148–158. doi:10.1016/s1566-0702(00)00173-9

Saul, J. P., Berger, R. D., Chen, M. H., and Cohen, R. J. (1989). Transfer function analysis of autonomic regulation. ii. respiratory sinus arrhythmia. *Am. J. Physiology-Heart Circulatory Physiology* 256, H153–H161. doi:10.1152/ajpheart.1989.256.1. H153

Sausedo-Solorio, J., and Pisarchik, A. (2011). Dynamics of unidirectionally coupled bistable hénon maps. *Phys. Lett. A* 375, 3677–3681. doi:10.1016/j. physleta.2011.07.057

Schreiber, T., and Schmitz, A. (1996). Improved surrogate data for nonlinearity tests. *Phys. Rev. Lett.* 77, 635–638. doi:10.1103/PhysRevLett.77.635

Schreiber, T., and Schmitz, A. (2000). Surrogate time series. *Phys. D. Nonlinear Phenom.* 142, 346–382. doi:10.1016/S0167-2789(00)00043-9

Searle, S. R., Speed, F. M., and Milliken, G. A. (1980). Population marginal means in the linear model: an alternative to least squares means. *Am. Statistician* 34, 216–221. doi:10.1080/00031305.1980.10483031

Shannon, C. E. (1948). A mathematical theory of communication. *Bell Syst. Tech. J.* 27, 379–423. doi:10.1002/j.1538-7305.1948.tb01338.x

Song, H. S., and Lehrer, P. M. (2003). The effects of specific respiratory rates on heart rate and heart rate variability. *Appl. Psychophysiol. Biofeedback* 28, 13–23. doi:10.1023/A:1022312815649

Theiler, J., Eubank, S., Longtin, A., Galdrikian, B., and Doyne Farmer, J. (1992). Testing for nonlinearity in time series: the method of surrogate data. *Phys. D. Nonlinear Phenom.* 58, 77–94. doi:10.1016/0167-2789(92)90102-S

Theiler, J., and Prichard, D. (1996). Constrained-realization monte-carlo method for hypothesis testing. *Phys. D. Nonlinear Phenom.* 94, 221–235. doi:10.1016/0167-2789(96)00050-4

Varley, T. F., Pope, M., Puxeddu, M. G., Faskowitz, J., and Sporns, O. (2023). Partial entropy decomposition reveals higher-order information structures in human brain activity. *Proc. Natl. Acad. Sci. U. S. A.* 120, e2300888120. doi:10.1073/pnas.2300888120

Victor, J. D. (2002). Binless strategies for estimation of information from neural data. *Phys. Rev. E* 66, 051903. doi:10.1103/PhysRevE.66.051903

Widjaja, D., Montalto, A., Vlemincx, E., Marinazzo, D., Van Huffel, S., and Faes, L. (2015). Cardiorespiratory information dynamics during mental arithmetic and sustained attention. *PLOS ONE* 10, e0129112–e0129114. doi:10.1371/journal.pone.0129112