

Research



**Cite this article:** Liu Z, Peach RL, Laumann F, Vallejo Mengod S, Barahona M. 2023 Kernel-based joint independence tests for multivariate stationary and non-stationary time series. *R. Soc. Open Sci.* **10**: 230857.  
<https://doi.org/10.1098/rsos.230857>

Received: 20 June 2023

Accepted: 3 November 2023

**Subject Category:**

Mathematics

**Subject Areas:**

applied mathematics/statistics

**Keywords:**

multivariate time-series analysis, joint independence test, higher-order interaction, kernel method

**Author for correspondence:**

Mauricio Barahona

e-mail: [m.barahona@imperial.ac.uk](mailto:m.barahona@imperial.ac.uk)

# Kernel-based joint independence tests for multivariate stationary and non-stationary time series

Zhaolu Liu<sup>1</sup>, Robert L. Peach<sup>2,3</sup>, Felix Laumann<sup>1</sup>, Sara Vallejo Mengod<sup>1</sup> and Mauricio Barahona<sup>1</sup>

<sup>1</sup>Department of Mathematics, Imperial College London, London SW7 2AZ, UK

<sup>2</sup>Department of Brain Sciences, Imperial College London, London W12 0NN, UK

<sup>3</sup>Department of Neurology, University Hospital Würzburg, Würzburg 97070, Germany

ZL, 0000-0002-8721-7506; MB, 0000-0002-1089-5675

Multivariate time-series data that capture the temporal evolution of interconnected systems are ubiquitous in diverse areas. Understanding the complex relationships and potential dependencies among co-observed variables is crucial for the accurate statistical modelling and analysis of such systems. Here, we introduce kernel-based statistical tests of joint independence in multivariate time series by extending the  $d$ -variable Hilbert–Schmidt independence criterion to encompass both stationary and non-stationary processes, thus allowing broader real-world applications. By leveraging resampling techniques tailored for both single- and multiple-realization time series, we show how the method robustly uncovers significant higher-order dependencies in synthetic examples, including frequency mixing data and logic gates, as well as real-world climate, neuroscience and socio-economic data. Our method adds to the mathematical toolbox for the analysis of multivariate time series and can aid in uncovering high-order interactions in data.

## 1. Introduction

Time series that record temporal changes in sets of system variables are ubiquitous across many scientific disciplines [1], from physics and engineering [2] to biomedicine [3,4], climate science [5,6], economics [7,8] and online human behaviour [9,10]. Many real-world systems are thus described as multivariate time series of (possibly) interlinked processes tracking the temporal evolution (deterministic or random) of groups of observables of interest. The relationships between the measured variables are often complex, in many cases displaying interdependencies among each other. For example, the spreading of COVID-19 in Indonesia

was dependent on weather conditions [11]; the sustainable development goals (SDGs) have extensive interlinkages [12]; there are strong interconnections between foreign exchange and cryptocurrencies [13]; and the brain displays multiple spatial and temporal scales of functional connectivity [14]. Driven by technological advances (e.g. imaging techniques in the brain sciences [15], or the increased connectivity of personal devices via the Internet of Things [16]), there is a rapid expansion in the collection and storage of multivariate time-series datasets, which underlines the need for mathematical tools to analyse the interdependencies within complex high-dimensional time-series data.

Characterizing the relationships between variables in a multivariate dataset often underpins the subsequent application of statistical and machine-learning methods. In particular, before further analyses can be performed, it is often crucial to determine whether the variables of interest are jointly independent [17]. Joint independence of a set of  $d$  variables means that no subset of the  $d$  variables are dependent. We need to look no further than ANOVA and  $t$ -tests to find classic statistical methods that assume joint independence of input variables, and the violation of this assumption can lead to incorrect conclusions [18]. Causal discovery methods, such as structural equation modelling, also require joint independence of noise variables [19]. Furthermore, joint independence has applications in uncovering higher-order networks, an emergent area highlighted in recent studies [20–24].

Kernel-based methods offer a promising framework for testing statistical independence. Notably, the  $d$ -variable Hilbert–Schmidt independence criterion (dHSIC) [19] can be used as a statistic to test the joint independence of  $d$  random variables. Developed as an extension of the pairwise HSIC [25], a statistical test that measures the dependence between two variables [25–27], dHSIC measures the dependence between  $d$  variables [19]. Specifically, dHSIC can be simply defined as the ‘squared distance’ between the joint distribution and the product of univariate marginals when they are embedded in a reproducing kernel Hilbert space (RKHS). Crucially, kernel methods do not make assumptions about the underlying distributions or type of dependencies (i.e. they are non-parametric). Yet, in its original form, dHSIC assumes the data to be *iid* (i.e. drawn from identical independent distributions). This is an unreasonable assumption in the case of time-series data, and it has precluded its application to temporal data.

To the best of our knowledge, dHSIC has not yet been extended to time-series data. The pairwise HSIC has been extended to deal with *stationary* random processes under two different test resampling strategies: shifting within time series [26] and the Wild Bootstrap method [27]. However, the assumption of stationarity, by which the statistical properties (e.g. mean, variance, autocorrelation) of the time series are assumed not to change over time, is severely restrictive in many real-world scenarios, as non-stationary processes are prevalent in many areas, e.g. stock prices under regime changes or weather data affected by seasonality or long-term trends. Hence, there is a need for independence tests that apply to both stationary and non-stationary processes. Recently, pairwise HSIC has been extended to non-stationary random processes by using random permutations over independent realizations of each time series, when available [28].

In this paper, we show how dHSIC can be applied to reject joint independence in the case of both stationary and non-stationary multivariate random processes. Following recent work [28], we adapt dHSIC so that it can be applied to stationary and non-stationary time-series data when multiple realizations are present. Additionally, we develop a new bootstrap method inspired by [26], which uses ‘shifting’ to deal with stationary time-series data when only one realization is available. Using these methodological advances, we then introduce statistical tests that rely on these two different resampling methods to generate appropriate null distributions: one for single-realization time series, which is only applicable to stationary random processes, and another for multiple realization time series, which is applicable to both stationary and non-stationary random processes. We show numerically that the proposed statistical tests based on dHSIC robustly and efficiently identify the lack of joint independence in synthetic examples with known ground truths. We further show how recursive testing from pairwise to  $d$ -order joint independence can reveal emergent higher-order dependencies in real-world socio-economic time series that cannot be explained by lower-order factorizations.

## 2. Preliminaries

### 2.1. Kernel-based tests for joint independence

**Definition (Joint independence of a set of variables).** The  $d$  variables  $X^j$ ,  $j = 1, \dots, d$ , with joint distribution  $\mathbb{P}_{X^1, \dots, X^d}$  are jointly independent if and only if the joint distribution is fully factorizable

into the product of its univariate marginals, i.e.  $\mathbb{P}_{X^1, \dots, X^d} = \prod_{j=1}^d \mathbb{P}_{X^j}$ , where the  $\mathbb{P}_{X^j}$  denote the marginals.

**Remark (Joint independence of subsets).** If  $d$  variables are jointly independent, then any subset of those  $d$  variables is also jointly independent, e.g.  $\mathbb{P}_{X^1, X^2, X^3} = \mathbb{P}_{X^1} \mathbb{P}_{X^2} \mathbb{P}_{X^3}$  implies  $\mathbb{P}_{X^1, X^2} = \mathbb{P}_{X^1} \mathbb{P}_{X^2}$ , which follows from marginalization with respect to  $X^3$  on both sides of the equality. Hence, by the contrapositive, lack of joint independence of a subset of variables implies lack of joint independence of the full set of variables.

A series of papers in the last two decades have shown how kernel methods can be used to test for independence of random variables (for details, see [19,25]). The key idea is to embed probability distributions in RKHSs [29] via characteristic kernels, thus mapping distributions uniquely to points in a vector space. For a summary of the key definitions and foundational results, see [30,31].

**Definition (RKHS and mean embedding for probability distributions [32,33]).** Let  $\mathcal{H}_k$  be a RKHS of functions  $f: \mathcal{X} \rightarrow \mathbb{R}$  endowed with dot product  $\langle \cdot, \cdot \rangle$ , and with a reproducing kernel  $k: \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ . Let  $\mathbb{P}$  be a distribution defined on a measurable space  $\mathcal{X}$ , then the mean embedding of  $\mathbb{P}$  in  $\mathcal{H}_k$  is an element  $\mu_{\mathbb{P}} \in \mathcal{H}_k$  given by  $\mu_{\mathbb{P}} := \int k(\mathbf{x}, \cdot) \mathbb{P}(d\mathbf{x})$ , with the property  $\langle f, \mu_{\mathbb{P}} \rangle = \mathbb{E}_{\mathbb{P}}[f] = \int f(\mathbf{x}) \mathbb{P}(d\mathbf{x})$ ,  $\forall f \in \mathcal{H}_k$ .

If the kernel is characteristic, the RKHS mapping is injective and this representation uniquely captures the information about each distribution. Based on such a mapping, statistics have been constructed to test for homogeneity (using the maximum mean discrepancy, MMD [33]) or independence (using the HSIC [25]) between two random variables.

**Remark.** An example of a characteristic kernel is the Gaussian kernel  $k_{\sigma}(\mathbf{x}, \mathbf{y}) = \exp(-\|\mathbf{x} - \mathbf{y}\|^2 / \sigma^2)$  where  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^p$ . The Gaussian kernel will be used throughout our applications below, but our results apply to any other characteristic kernel.

Recently, an extension of HSIC for  $d$  variables, denoted dHSIC, was introduced and used as a statistic for joint independence to test the null hypothesis  $H_0: \mathbb{P}_{X^1, \dots, X^d} = \prod_{j=1}^d \mathbb{P}_{X^j}$ .

**Definition (dHSIC [19]).** Let us consider  $d$  random variables  $X^j$ ,  $j = 1, \dots, d$ , with joint distribution  $\mathbb{P}_{X^1, \dots, X^d}$ . For each  $X^j$ , let  $\mathcal{H}_{k^j}$  denote a separable RKHS with characteristic kernel  $k^j$ . The  $d$ -variable dHSIC, which measures the similarity between the joint distribution and the product of the marginals, is defined as

$$\text{dHSIC}(X^1, \dots, X^d) := \|\mu_{\mathbb{P}_{X^1, \dots, X^d}} - \mu_{\mathbb{P}_{X^1} \dots \mathbb{P}_{X^d}}\|_{\mathcal{H}}^2 \quad (2.1)$$

where  $\mathcal{H} := \mathcal{H}_{k^1} \otimes \dots \otimes \mathcal{H}_{k^d}$  and  $\otimes$  is the tensor product.

**Remark.** Given the definition (2.1), dHSIC is zero if and only if the variables are jointly independent, i.e. when the joint distribution is equal to the product of the marginals. This is the basis for using dHSIC to define the null hypothesis for statistical tests of joint independence.

**Remark (Emergent high-order dependencies).** As noted above, the rejection of joint independence for any subset of a set of  $d$  variables also implies the rejection of joint independence for the full set of  $d$  variables. Therefore, many observed rejections of joint independence at higher orders follow from rejections of joint independence at lower orders (i.e. within subsets of variables). To identify more meaningful high-order interactions, in some cases, we will also consider ‘first time rejections’ of  $d$ -way joint independence, i.e. when the joint independence of a set of  $d$  variables is rejected but the joint independence of each and all of its subsets of size  $d' < d$  cannot be rejected. We denote these as emergent high-order dependencies.

## 2.2. Time series as finite samples of stochastic processes

Our interest here is in the joint independence of time series, which we will view as finite samples of stochastic processes.

**Notation (Stochastic processes and sample paths).** We will consider a set of  $d$  stochastic processes  $\{X^j(t; \omega) : t \in \mathcal{T}\}$ ,  $j = 1, \dots, d$ , where  $t \in \mathcal{T}$  is defined over the index set, corresponding to time, and  $\omega \in \Omega$  is defined over the sample space. Below, we will also use the shorthand  $\{X_i^j\}$  to denote each stochastic process.

For each stochastic process, we may observe  $n$  independent realizations (or paths), which are samples from  $\Omega$  indexed by  $\omega_i$ :  $\{X^i(t; \omega_i) : t \in \mathcal{T}\}$ ,  $i = 1, \dots, n$ . Furthermore, each path is finite and sampled at times  $t = 1, \dots, T_j$ .

**Remark (Time series as data samples).** For each variable  $X^j$ , the data samples (time series) consist of  $n$  paths  $(X^j(1, \omega_i), \dots, X^j(T_j, \omega_i))$ ,  $i = 1, \dots, n$ , which we arrange as  $T_j$ -dimensional vectors  $\mathbf{x}_i^j = (x_{i,1}^j, \dots, x_{i,T_j}^j)$ ,  $i = 1, \dots, n$ , i.e. the components of the vector are given by  $x_{i,t_k}^j := X^j(t_k; \omega_i)$ .

**Definition (Independence of stochastic processes).** Two stochastic processes  $\{X_t^j\}$  and  $\{X_t^{j'}\}$  with the same index set  $\mathcal{T}$  are independent if for every choice of sampling times  $t_1, \dots, t_f \in \mathcal{T}$ , the random vectors  $(X^j(t_1), \dots, X^j(t_f))$  and  $(X^{j'}(t_1), \dots, X^{j'}(t_f))$  are independent. Independence is usually denoted as  $\{X_t^j\} \perp \{X_t^{j'}\}$ . Below, we will abuse notation and use the shorthand  $X_t^j \perp X_t^{j'}$ .

From this definition, it immediately follows that the realizations are independent.

**Remark (Independence of realizations).** Although the samples within a path  $(X^j(1, \omega_i), \dots, X^j(T_j, \omega_i)) = (x_{i,1}^j, \dots, x_{i,T_j}^j)$  are not necessarily independent across time, each variable is independent across realizations for any time  $t$ , i.e.  $x_{i,t}^j \perp x_{i',t}^j \forall t, \forall i \neq i'$ . In other words, the  $n$  time series are assumed to be *iid* samples,  $\{\mathbf{x}_i^j\}_{i=1}^n \stackrel{iid}{\sim} \mathbb{P}_{X^j}$ , where  $P_{X^j}$  is a finite-dimensional distribution of the stochastic process  $\{X_t^j\}$ .

**Definition (Stationarity).** A stochastic process is said to be stationary if all its finite-dimensional distributions are invariant under translations of time.

*Aim of the paper:* Here, we use kernels to embed finite-dimensional distributions of the  $d$  stochastic processes  $\{X_t^j\}$  and design tests for joint independence of time series thereof. Recent work has used HSIC to test for independence of pairs of stationary [27,34] and non-stationary [28] time series. Here, we extend this work to  $d > 2$  time series using tests based on dHSIC. We consider two scenarios:

- if we only observe a single time series ( $n = 1$ ) of each of the  $d$  variables, then we can only consider stationary processes;
- if we have access to several time series ( $n > 1$ ) of each of the  $d$  variables, then we can also study non-stationary processes.

### 3. dHSIC for joint independence of stationary time series

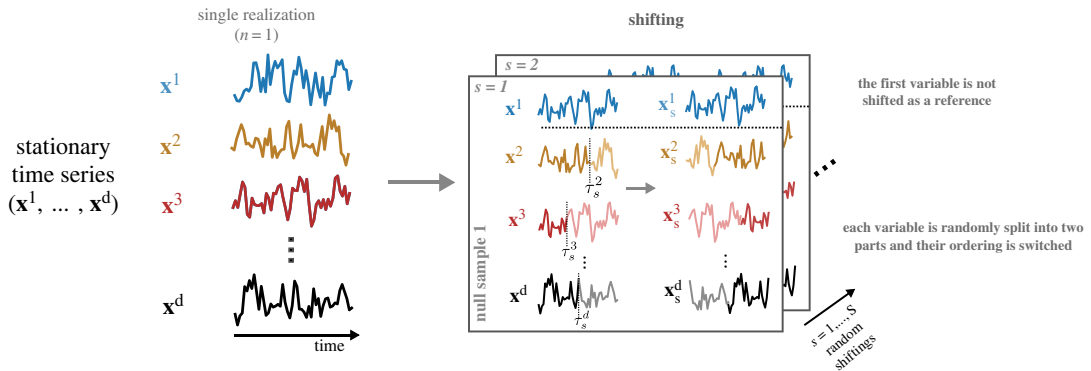
We first consider the scenario where we only have one time series ( $n = 1$ ) for each of the  $d$  variables  $X^j$ , which are all assumed to be stationary. Our dataset is then  $\{\mathbf{x}^j\}_{j=1}^d$ , and it consists of  $d$  time-series vectors  $\mathbf{x}^j = (x_1^j, \dots, x_T^j)$ , which we view as single realizations of the stationary stochastic processes  $\{X_t^j\}$ , all sampled at times  $t = 1, \dots, T$ . As will become clear below, the limited information provided by the single realization, together with the use of permutation-based statistical tests, means that the assumption of stationarity is necessary [26].

Let  $K^j \in \mathbb{R}^{T \times T}$  be kernel matrices with entries  $K_{ab}^j = k^j(x_a^j, x_b^j)$  where  $a, b \in \{1, \dots, T\}$ , and  $k^j : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$  is a characteristic kernel (e.g. Gaussian); hence, the matrix  $K^j$  captures the autocorrelational structure of variable  $X^j$ . In this case, dHSIC (2.1) can be estimated as the following expansion in terms of kernel matrices [19,35]:

$$\widehat{\text{dHSIC}}_{st}(\mathbf{x}^1, \dots, \mathbf{x}^d) := \frac{1}{T^2} \sum_{a=1}^T \sum_{b=1}^T \prod_{j=1}^d K_{ab}^j - \frac{2}{T^{d+1}} \sum_{a=1}^T \prod_{j=1}^d \sum_{b=1}^T K_{ab}^j + \frac{1}{T^{2d}} \prod_{j=1}^d \sum_{a=1}^T \sum_{b=1}^T K_{ab}^j. \quad (3.1)$$

The null hypothesis is  $H_0 : \mathbb{P}_{X^1, \dots, X^d} = \mathbb{P}_{X^1} \cdots \mathbb{P}_{X^d}$ , and we test (3.1) for statistical significance. To do so, we bootstrap the distribution under  $H_0$  using random shifting to generate  $S$  samples [26]. For each of the samples  $s = 1, \dots, S$ , we fix one time series ( $\mathbf{x}^1$  without loss of generality) and generate random shifting points  $\tau_s^j$ ,  $j = 2, \dots, d$  for each of the other  $d - 1$  time series where  $h < \tau_s^j < T$  and  $h$  is chosen to be the first index where the autocorrelation of  $\sum_{j=1}^d \mathbf{x}^j$  is less than 0.2 [26].

Each time series is then shifted by  $\tau_s^j$  so that  $x_{s,t}^j = x_{(t+\tau_s^j) \bmod T}^j$ . This shifting procedure, which is illustrated in figure 1, breaks the dependence across time series yet retains the local temporal dependence within each time series. In this way, we produce  $S$  randomly shifted datasets  $(\mathbf{x}_s^1, \dots, \mathbf{x}_s^d)$ ,  $s = 1, \dots, S$ , and the estimated dHSIC is computed for each shifting:  $\widehat{\text{dHSIC}}_{st}(\mathbf{x}_s^1, \dots, \mathbf{x}_s^d)$ . The p-value is computed by Monte Carlo approximation [19]. Given a significance level  $\alpha$ , the null hypothesis  $H_0$  is rejected if  $\alpha > \text{p-value}$ . We note that although an



**Figure 1.** Shifting strategy for random sampling of single-realization stationary time series. The shifting method is used for stationary time series when only one realization of each variable is available. For each null sample  $s$ , the first time series  $\mathbf{x}^1$  is kept fixed and a random shifting point  $\tau_s^j$  is chosen for each of the other time series  $\mathbf{x}^j$ ,  $j = 2, \dots, d$  so that the sections before and after  $\tau_s^j$  (darker and lighter shades of colour) are switched. This process generates  $S$  randomly shifted samples that are used to bootstrap the null distribution.

alternative to shifting called Wild Bootstrap has been proposed [27,36], it has been reported to produce large false positive rates [37]. We therefore use shifting (and not the Wild Bootstrap) in this manuscript.

## 3.1. Numerical results

### 3.1.1. Validation on synthetic stationary multivariate systems with a single realization

To validate our approach, we apply the dHSIC test for joint independence to datasets consisting of  $d = 3$  time series of length  $T$  with  $n = 1$  realizations (i.e. one time series per variable). We use three stationary models with a known dependence structure (ground truth), the strength of which can be varied. For each test, we use  $S = 1000$  randomly shifted samples and we take  $\alpha = 0.05$  as the significance level. We then generate 200 such datasets for every model and combination of parameters  $(T, \lambda)$ , and compute either the test power (i.e. the probability that the test correctly rejects the null hypothesis when there is dependence) or the type I error (i.e. the probability that the test mistakenly rejects the true null hypothesis when there is independence) for the 200 datasets.

*Model 1.1: Three-way dependence ensuing from pairwise dependencies.* The first stationary example [38] has a three-way dependence that follows from the presence of two simultaneous two-way dependencies:

$$X_t = \frac{1}{2}X_{t-1} + \epsilon_t, \quad Y_t = \frac{1}{2}Y_{t-1} + \eta_t, \quad Z_t = \frac{1}{2}Z_{t-1} + \zeta_t + \lambda(X_t + Y_t), \quad (3.2)$$

where  $\epsilon_t$ ,  $\eta_t$ ,  $\zeta_t$  and  $\theta_t$  are generated as *iid* samples from a normal distribution  $\mathcal{N}(0, 1)$ , and the dependence coefficient  $\lambda$  regulates the magnitude of the dependence between variables, i.e. for  $\lambda = 0$  we have joint independence of  $(X, Y, Z)$  and the dependence grows as  $\lambda$  is increased. Figure 2a shows the result of our test for  $d = 3$  variables applied to time series of length  $T = [100, 300, 600, 900, 1200]$  and increasing values of the dependence coefficient  $0 \leq \lambda \leq 1$  generated from model (3.3). As either  $\lambda$  or  $T$  increase, it becomes easier to reject the null hypothesis of joint independence. Full test power can already be reached for  $\lambda = 0.5$  across all lengths of time series. Our test also rejects pairwise independence between the  $(X, Z)$  and  $(Y, Z)$  pairs, and fails to reject independence between  $(X, Y)$ , as expected from the ground truth.

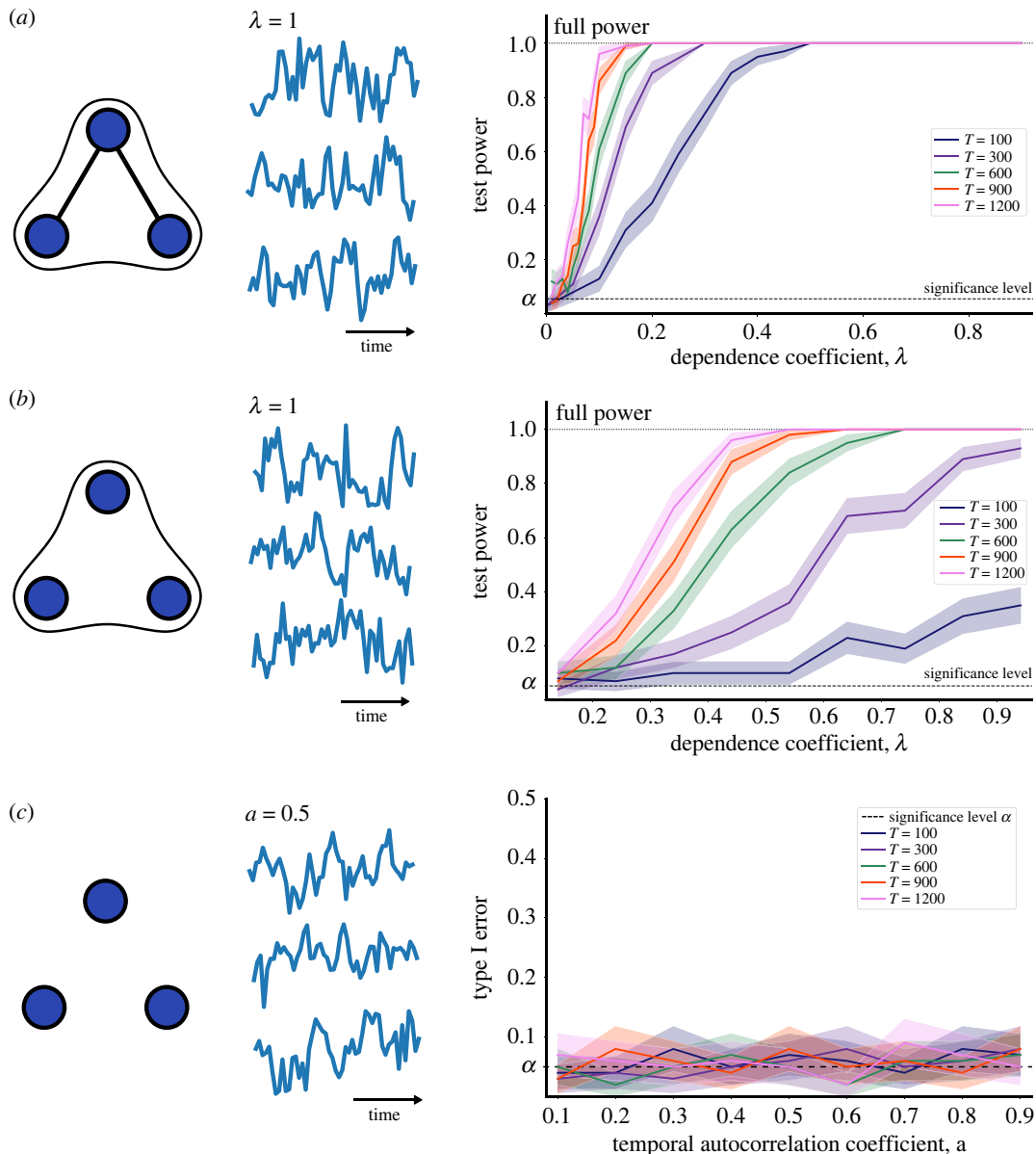
*Model 1.2: Pure three-way dependence.* Our second stationary example, also from [38], includes a three-way dependence without any underlying pairwise dependence:

$$X_t = \frac{1}{2}X_{t-1} + \epsilon_t, \quad Y_t = \frac{1}{2}Y_{t-1} + \eta_t, \quad Z_t = \frac{1}{2}Z_{t-1} + \zeta_t + \lambda|\theta_t|\text{sign}(X_t Y_t), \quad (3.3)$$

where  $\epsilon_t$ ,  $\eta_t$ ,  $\zeta_t$  and  $\theta_t$  are *iid* samples from  $\mathcal{N}(0, 1)$ , and the coefficient  $\lambda$  regulates the three-way dependence. Figure 2b shows that the test rejects the null hypothesis as either  $\lambda$  or  $T$  increase, although the test power is lower relative to (3.2), as there are no two-way dependencies present in this case, i.e. this is a three-way emergent dependency.

*Model 1.3: Joint independence.* As a final validation, we use a jointly independent example [38]:

$$X_t = aX_{t-1} + \epsilon_t, \quad Y_t = aY_{t-1} + \eta_t, \quad Z_t = aZ_{t-1} + \zeta_t, \quad (3.4)$$

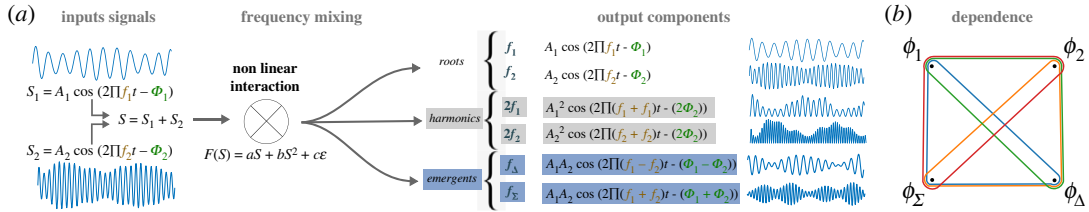


**Figure 2.** Stationary systems with a single realization. Left: a visualization of the ground truth dependencies where edges represent rejection of pairwise dependence and three-way hyperedges represent the rejection of three-way joint independence. Middle: an example of the three-variable time series for which dHSIC was computed. Right: test power (for  $a, b$ ) and type I error (for  $c$ ) computed by applying the dHSIC test to 200 datasets generated for each model with different dependence coefficient  $\lambda$  (autocorrelation coefficient  $a$  for  $c$ ) and length of time series  $T$ . The lines represent the average over the 200 datasets and the shaded areas correspond to confidence intervals. The systems are taken from [38]: (a) three-way dependence ensuing from pairwise dependencies (3.2); (b) emergent three-way dependence with no underlying pairwise dependencies (3.3); and (c) joint independence (3.4).

where  $\epsilon_t$ ,  $\eta_t$  and  $\zeta_t$  are *iid* samples from  $\mathcal{N}(0, 1)$ . Figure 2c shows that in this case we do not reject the null hypothesis of joint independence across a range of values of the autocorrelation parameter  $a$ . Note that the type I error of the test remains controlled around the significance  $\alpha = 0.05$  for all values of  $T$  and  $a$ .

### 3.1.2. Synthetic frequency mixing data

As a further illustration linked more closely to real-world applications, we have generated a dataset based on frequency mixing of temporal signals. Frequency mixing is a well-known phenomenon in electrical engineering, widely used for heterodyning, i.e. shifting signals from one frequency range to another. Applying a nonlinear function (e.g. a quadratic function or a rectifier) to the sum of two



**Figure 3.** Frequency mixing. (a) Two independent input signals with root frequencies  $f_1=7$  Hz and  $f_2=18$  Hz are mixed via a quadratic function with noise to generate the signal  $F$ . This signal has components at the root frequencies, harmonics (double of the root frequencies) and emergents (sum and difference of the root frequencies), as shown by the output waveforms (right). Time series of the instantaneous phases  $\phi_1, \phi_2, \phi_\Delta, \phi_\Sigma$  are extracted from the output components at  $f_1, f_2, f_\Delta, f_\Sigma$  and dHSIC is applied to them. (b) In this case, the dHSIC test does not reject pairwise independence between any pair of variables (i.e. there are no pairwise dependencies), but rejects the joint independence between any three of the four variables (i.e. four three-way emergent dependencies are present), as shown by the triangles, and consequently also rejects the joint independence between the four variables (square).

signals with distinct frequencies generates new signals with emergent frequencies at the sum and difference of the input signals (figure 3a–c). It has previously been shown that the instantaneous phases of the emergents display a unique three-way dependence, without any pairwise dependencies [39–41]. Importantly, given sufficiently long time series, the instantaneous phase can be considered a stationary signal [39]. Hence, we can apply our test to this system.

Here, we generated a dataset using the sum of two sinusoidal functions with frequencies  $f_1=7$  Hz and  $f_2=18$  Hz as input, to which we applied a quadratic function plus weighted Gaussian noise  $\epsilon$ . This produces a signal  $F$  that contains components at input (root) frequencies ( $f_1=7$  Hz and  $f_2=18$  Hz), second harmonics ( $2f_1=14$  Hz and  $2f_2=36$  Hz) and emergent frequencies ( $f_\Delta=f_2-f_1=11$  Hz and  $f_\Sigma=f_1+f_2=25$  Hz). See figure 3a and [39] for further details. We then computed a wavelet transform and extracted the instantaneous phases for frequencies  $f_1, f_2, f_\Sigma$  and  $f_\Delta$ , which we denoted  $\phi_1, \phi_2, \phi_\Sigma$  and  $\phi_\Delta$ . These phases can be considered as stationary time series. The ground truth is that there should be no pairwise dependencies between any of those phases, but there are higher-order interactions involving three-way and four-way dependencies [39].

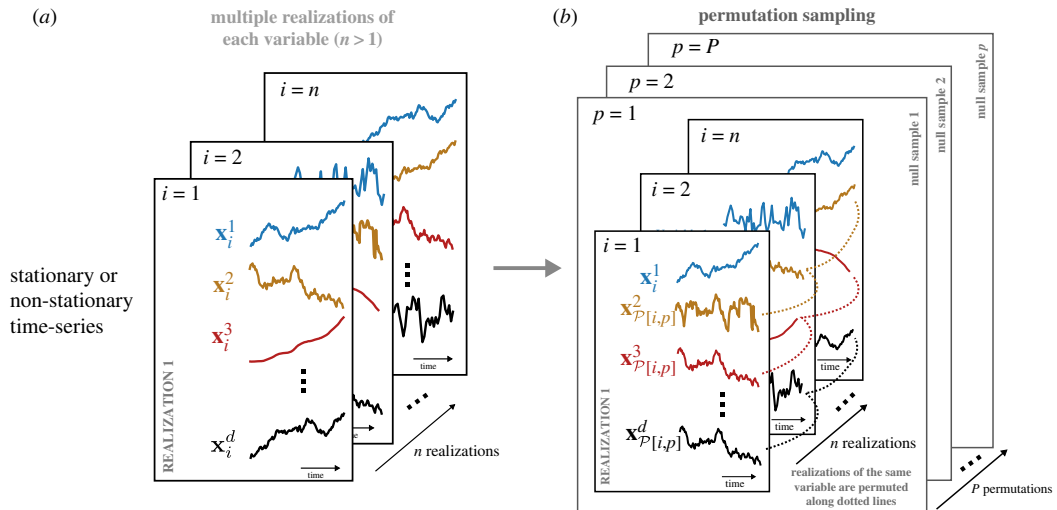
We applied dHSIC with shifting to all possible groupings of  $d$  phases (for  $d=2, 3, 4$ ) from the set  $\{\phi_1, \phi_2, \phi_\Sigma, \phi_\Delta\}$ . The phases consisted of time series with length  $T=1000$ , and we used  $S=1000$  shiftings for our bootstrap. We found that the null hypothesis of independence could not be rejected for any of the six phase pairs ( $d=2$ ), whereas joint independence was rejected for all four phase triplets ( $d=3$ ) and for the phase quadruplet ( $d=4$ ). The rejection of all the three-way and four-way joint independence hypotheses, without rejection of any of the pairwise independence hypotheses, thus recovers the ground truth expected structure (figure 3b).

### 3.1.3. Application to climate data

As an application to real-world data, we used the PM2.5 air quality dataset, which contains four variables: hourly measurements of particulate matter with a diameter of 2.5 microns or less (PM2.5) recorded by the US Embassy in Beijing between 2010 and 2014, and three concurrent meteorological variables (dew point, temperature, air pressure) measured at Beijing Capital International Airport [42]. Non-stationary trends and yearly seasonal effects were removed by taking differences of period 1 and period 52 in the averaged weekly data. Stationarity of the de-trended series was verified by an Adfuller test [43]. As expected, we found that the null hypotheses (joint independence) were rejected for all groups of  $d=2, 3, 4$  variables, implying that PM2.5, dew point, temperature and air pressure are all dependent on each other.

## 4. dHSIC for joint independence of non-stationary time series with multiple realizations

When we have multiple independent observations of the  $d$  variables, these can be viewed as *iid* samples of a multivariate probability distribution. By doing so, the requirements of stationarity and same point-in-time measurements across all variables can be loosened.



**Figure 4.** Random permutation sampling of multivariate time series with multiple realizations. A permutation strategy similar to the one developed for *iid* data [19] can be applied when multiple realizations of either stationary or non-stationary time-series data are available. Each null sample  $p = 1, \dots, P$  is generated by randomly permuting the time series for variables  $j = 2, \dots, d$  across realizations  $i = 1, \dots, n$ , as indicated by the dotted lines, while the first variable remains unchanged. Null distributions are generated from independent samples of this process.

Consider the case when we have access to  $n > 1$  observations of the set of variables  $(X^1, \dots, X^d)$ , where each observation  $i = 1, \dots, n$  consists of  $d$  time series  $X^j$ , which we write as vectors  $\mathbf{x}_i^j = (x_{i,1}^j, \dots, x_{i,T_i}^j)$  of length  $T_j$ . Each of the  $n$  observations thus consists of a set  $\{\mathbf{x}_i^j\}_{j=1}^d$ , which can be viewed as an independent (*iid*) realization of a finite-dimensional multivariate distribution  $\mathbb{P}_{X^1, \dots, X^d}$ . To simplify our notation, we compile the  $n$  observations of each  $X^j$  as rows of a  $n \times T_j$  matrix  $\mathbf{X}^j$ , so that  $\mathbf{X}_{[i,:]}^j = \mathbf{x}_i^j$ .

Let  $\kappa^j: \mathbb{R}^{T_j} \times \mathbb{R}^{T_j} \rightarrow \mathbb{R}$  be a characteristic kernel (e.g. Gaussian) that captures the similarity between a pair of time series of variable  $X^j$ . We then define the set of kernel matrices  $\mathcal{K}^j \in \mathbb{R}^{n \times n}$  with entries  $\mathcal{K}_{\alpha\beta}^j = \kappa^j(\mathbf{x}_\alpha^j, \mathbf{x}_\beta^j)$  where  $\alpha, \beta \in \{1, \dots, n\}$ . Therefore, the matrix  $\mathcal{K}^j$  captures the similarity structure between the time series of variable  $X^j$  across the  $n$  observations. This set-up thus allows us not to require stationarity in our variables, since the  $n$  observations capture the temporal behaviour of the  $d$  variables concurrently. In this case, dHSIC for the set of observations  $(\mathbf{X}^1, \dots, \mathbf{X}^d)$  can be estimated as [19]

$$\widehat{\text{dHSIC}}_{\text{mult}}(\mathbf{X}^1, \dots, \mathbf{X}^d) := \frac{1}{n^2} \sum_{\alpha=1}^n \sum_{\beta=1}^n \prod_{j=1}^d \mathcal{K}_{\alpha\beta}^j - \frac{2}{n^{d+1}} \sum_{\alpha=1}^n \prod_{j=1}^d \sum_{\beta=1}^n \mathcal{K}_{\alpha\beta}^j + \frac{1}{n^{2d}} \prod_{j=1}^d \sum_{\alpha=1}^n \sum_{\beta=1}^n \mathcal{K}_{\alpha\beta}^j. \quad (4.1)$$

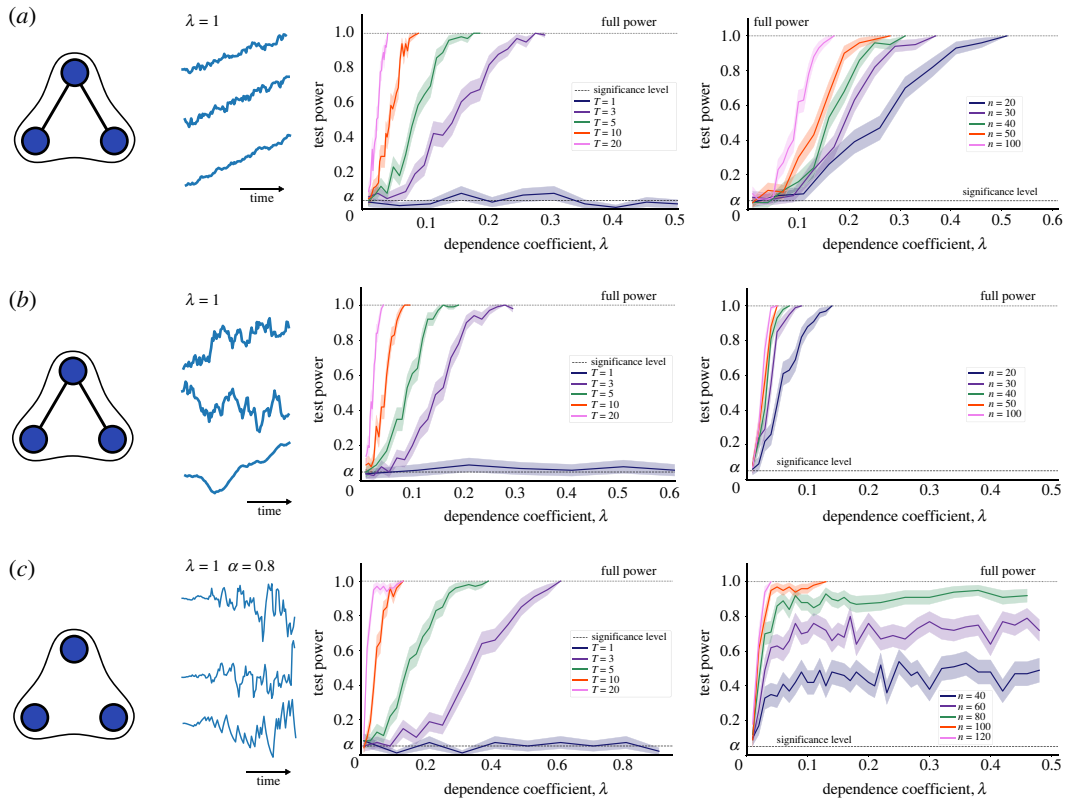
Similarly to §3, the null hypothesis is  $H_0: \mathbb{P}_{X^1, \dots, X^d} = \mathbb{P}_{X^1} \cdots \mathbb{P}_{X^d}$  and we test (4.1) for statistical significance. Due to the availability of multiple realizations, however, we use a different resampling method (standard permutation test) to bootstrap the distribution of (4.1) under  $H_0$  (figure 4). For each of the samples  $p = 1, \dots, P$ , we fix one variable ( $X^1$  without loss of generality), and we randomly permute the rest of the variables *across realizations* to create the permuted sample  $\{\mathbf{x}_p^j\}_{j=1}^d = \{\mathbf{x}_{P[i,p]}^j\}_{j=1}^d$ , where  $P[i, p]$  indicates a random permutation between realizations, and  $\mathbf{x}_p^1 = \mathbf{x}_i^1, \forall p$ . In this way, we produce  $P$  permuted datasets  $(\mathbf{X}_p^1, \dots, \mathbf{X}_p^d)$ ,  $p = 1, \dots, P$ , with  $\mathbf{X}_p^1 = \mathbf{X}^1$ . The estimated dHSIC (4.1) is then computed for each permutation  $p$ . Given a significance level  $\alpha$ , the null hypothesis  $H_0$  is rejected if  $\alpha > \text{p-value}$  where the p-value is computed by Monte Carlo approximation [19].

## 4.1. Numerical results

### 4.1.1. Validation on simple non-stationary multivariate systems

The dHSIC test is applied to datasets consisting of  $n$  observations of non-stationary time series of length  $T$  of three variables  $(X, Y, Z)$ , with ground truth dependencies that can be made stronger by increasing a dependence coefficient  $\lambda$ . For every model and combination of parameters  $(n, T, \lambda)$ , we generate 200 datasets and compute the test power, i.e. the probability that the test correctly rejects the null hypothesis in our 200 datasets. Figure 5 shows our numerical results for two non-stationary models: the first model (shown in figure 5*a,b* with two non-stationary trends) has a three-way dependence





**Figure 5.** Three-variable non-stationary systems with multiple realizations. Left: a visualization of the ground truth dependencies, where edges represent pairwise dependence and 3-edges represent a three-way dependence. Middle left: an example of a realization of a three-variable time series. Middle right: test power computed by applying the dHSIC test to 200 datasets generated from a model at varying values of the dependence coefficient  $\lambda$  and the length of time series  $T$ , with a fixed number of realizations  $n = 100$ . Right: test power computed by applying the dHSIC test to 200 datasets generated from a model at varying values of the dependence coefficient  $\lambda$  and the number of realizations  $n$ , with a fixed length of time series  $T = 20$ . The lines represent the average over the 200 datasets and the shaded areas correspond to confidence intervals. The systems are: (a,b) three-way dependence ensuing from two-way dependencies (4.2): (a) linear trend, (b) a complex dependence term; and (c) three-way dependence with no underlying pairwise dependencies and a non-stationary trend (4.3).

ensuing from two-way dependencies; the second model (shown in figure 5d for a non-stationary trend) has an emergent three-way dependence with no pairwise dependencies.

*Model 2.1: Three-way dependence ensuing from pairwise dependencies with non-stationarity.* The first model has the same dependence structure as (3.2), i.e. two simultaneous pairwise dependencies and an ensuing three-way dependence, but in this case with non-stationary trends:

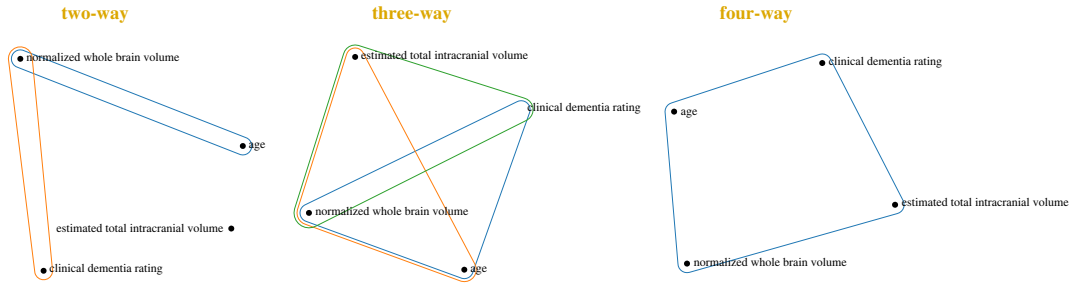
$$X_t = g_1(t) + X_{t-1} + \epsilon_t, \quad Y_t = g_2(t) + Y_{t-1} + \eta_t, \quad Z_t = g_3(t) + Z_{t-1} + \zeta_t + \lambda(X_t + Y_t), \quad (4.2)$$

where  $\epsilon_t$ ,  $\eta_t$  and  $\zeta_t$  are *iid* samples from a normal distribution  $\mathcal{N}(0, 1)$ ;  $\lambda$  regulates the strength of the dependence ( $\lambda = 0$  means joint independence); and  $g_1(t)$ ,  $g_2(t)$ ,  $g_3(t)$  are non-stationary trends as follows:

- linear trend (figure 5a):  $g_1(t) = g_2(t) = g_3(t) = t$
- complex nonlinear trend (figure 5b):  $g_1(t) = \sin^2(t)/\log(1+t)$ ,  $g_2(t) = \cos^2(t)/\log(1+t)$ ,  $g_3(t) = \sin(t)\cos(t)/\log(1+t)$ .

Figure 5a,b shows that the dHSIC test is able to reject the null hypothesis of joint independence for (4.2) even for short time series and low values of the dependence coefficient  $\lambda$ . The test power increases rapidly as the length of the time series  $T$  or the number of realizations  $n$  are increased. As expected, the null hypothesis cannot be rejected for  $T = 1$ , since the temporal dependence is no longer observable.

*Model 2.2: Emergent three-way dependence with non-stationarity.* The second model has the same dependence structure as (3.4) (i.e. an emergent three-way dependence without two-way dependencies)



**Figure 6.** High-order dependencies between four variables in MRI and Alzheimer’s data containing multiple realizations of time-series data. The hyperedges represent rejections of the respective joint independence tests. We find 2 (out of 6) pairwise dependencies and 3 (out of 4) three-way dependencies, as well as the four-way dependence between all variables. There are no emergent dependencies in this example.

but with non-stationary trends

$$X_t = aX_{t-1} + \epsilon_t + t \sin(t), \quad Y_t = aY_{t-1} + \eta_t + t \cos(t), \quad Z_t = aZ_{t-1} + \zeta_t + \lambda t \text{sign}(X_t Y_t), \quad (4.3)$$

where, again,  $\epsilon_t$ ,  $\eta_t$  and  $\zeta_t$  are *iid* samples from  $\mathcal{N}(0, 1)$ , and  $\lambda$  regulates the strength of the dependence. We set  $a = 0.8$ , the point at which the data becomes non-stationary according to an Adfuller test. Figure 5c shows good performance of the test, which is able to reject joint independence for low values of  $\lambda$ , with increasing test power as the length of the time series  $T$  and the number of realizations  $n$  are increased (figure 5c).

#### 4.1.2. Synthetic XOR dependence

The exclusive OR (XOR) gate (denoted  $\oplus$ ) is a logical device with two Boolean (0–1) inputs and one Boolean output, which returns a 1 when the number of ‘1’ inputs is odd. Here, we consider a system with three Boolean variables  $X$ ,  $Y$ ,  $W$  driven by noise, which get combined via XOR gates to generate another Boolean variable  $Z$ :

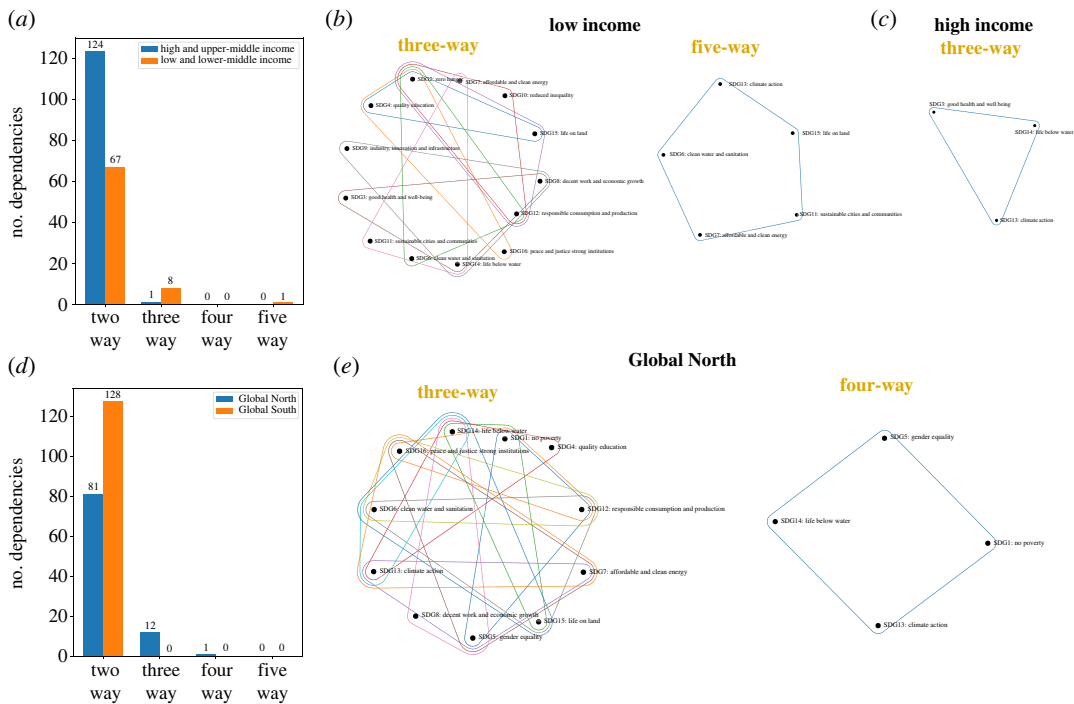
$$\text{and} \quad \begin{aligned} X_t &= \begin{cases} X_{t-1}, & \text{if } \epsilon_t \geq 0.5 \\ 1 - X_{t-1}, & \text{otherwise} \end{cases} & Y_t &= \begin{cases} Y_{t-1}, & \text{if } \eta_t \geq 0.5 \\ 1 - Y_{t-1}, & \text{otherwise} \end{cases} \\ W_t &= \begin{cases} W_{t-1}, & \text{if } \zeta_t \geq 0.5 \\ 1 - W_{t-1}, & \text{otherwise} \end{cases} & Z_t &= \begin{cases} X_t \oplus Y_t \oplus W_t, & \text{if } \theta_t \geq 0.5 \\ 1 - X_t \oplus Y_t \oplus W_t, & \text{otherwise} \end{cases} \end{aligned} \quad (4.4)$$

where  $\epsilon_t$ ,  $\eta_t$ ,  $\zeta_t$  and  $\theta_t$  are *iid* samples from  $\mathcal{U}[0, 1)$ , a uniform distribution between 0 and 1, and  $X_0$ ,  $Y_0$  and  $W_0$  are initialized as random Boolean variables. The dependence in this system is high-order: it only appears when considering the four variables, with no three-way or two-way dependencies. We find that our test does not reject joint independence for  $d = [2, 3]$  variables, but does reject joint independence of the four-variable case.

#### 4.1.3. Application to MRI and Alzheimer’s data

As a first application to data with multiple realizations, we apply our test to a magnetic resonance imaging (MRI) and Alzheimer’s longitudinal dataset [44], which comprises demographic and MRI data collected from subjects over several visits. Here, we consider  $n = 56$  subjects, each with at least three visits ( $T = 3$ ), and we assume that the subjects constitute *iid* realizations—a reasonable assumption since this is a well-designed population study with representative samples. We then perform dHSIC tests to find dependencies between four key variables: age, normalized whole brain volume (nWBV), estimated total intracranial volume (eTIV) and clinical dementia rating (CDR). The first three variables are clinical risk factors, whereas CDR is a standardized measure of disease progression.

Our findings are displayed as hypergraphs in figure 6 where nodes represent variables and hyperedges represent rejections of joint independence from the two-way, three-way and four-way dHSIC tests. In this case, we find only two pairwise dependencies (age–nWBV and nWBV–CDR), while eTIV is seemingly disconnected to the rest of the variables. Note that the possible emergent three-way interaction (age–eTIV–CDR) is not present, although eTIV shows the expected three-way and four-way dependencies with CDR, nWBV and age. This example highlights how our method can be used to reveal the different higher-order dependencies beyond pairwise interactions. To understand



**Figure 7.** Emergent high-order dependencies between SDGs. (a–c) Comparison of SDG dependencies in low- and high-income countries. (a) There is a higher number of emergent higher-order dependencies in low-income countries. The  $d > 2$  dependencies are mapped onto  $d$ -order hypergraphs for (b) low-income and (c) high-income countries. (d,e) Comparison of SDG dependencies in Global North and Global South countries. (d) Emergent high-order dependencies are found in the Global North, whereas the Global South displays only two-way dependencies. (e) The  $d > 2$  dependencies for the Global North are mapped onto  $d$ -order hypergraphs.

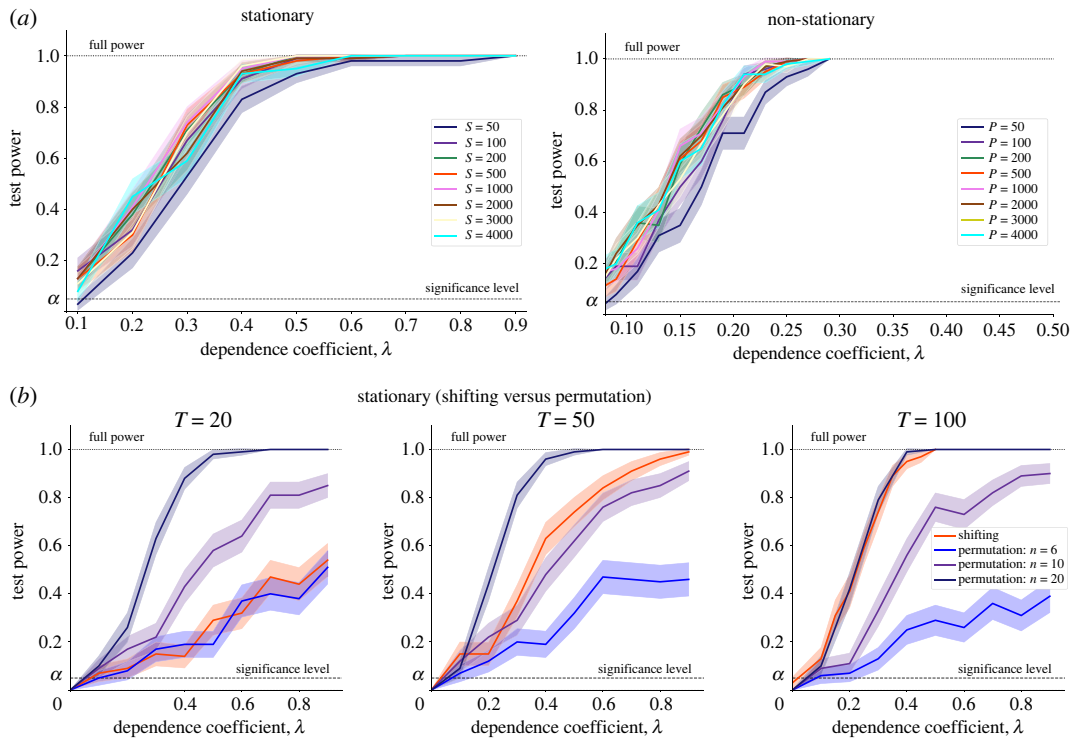
the complex high-order interactions of the incomplete factorizations, methods based on Streitberg and Lancaster interaction can be explored in future work [45].

#### 4.1.4. Application to socio-economic data

As a final illustration in a different domain area, we test for joint independence between the United Nations SDGs [46]. This dataset consists of a time series of a large number of socio-economic indicators forming the 17 SDGs (our variables  $X^j$ ,  $j = 1, \dots, 17$ ) measured yearly between 2000 and 2019 ( $T = 20$ ) for all 186 countries in the world (see [12] for details on the dataset). We take the countries to be *iid* realizations, as in [12], although this assumption is less warranted here than for the dementia dataset in §4.1.3 due to moderate correlations between countries due to socio-economic and political relationships.

As an illustration of the differences in data dependencies across country groupings, we consider two classic splits: (i) a split based on income level ( $n = 74$  countries with low and lower-middle income, and  $n = 105$  countries with high and upper-middle income); and (ii) a split based on broad geography and socio-economic development ( $n = 49$  countries in the Global North and  $n = 137$  countries in the Global South). This dataset highlights the difficulties of examining high-order dependencies as the number of variables grows, e.g.  $d = 17$  in this case.

The results of applying this recursive scheme to the SDG dataset are shown in figure 7. The comparison between low- and high-income countries (figure 7a–c) shows that the latter have strong pairwise dependencies (124 rejections of two-way independence out of a total of 136 pairs) and only 1 emergent three-way interaction (figure 7c), whereas the former have more emergent higher-order dependencies (eight three-way and one five-way) (figure 7b). These results suggest that the interdependencies between SDGs are more complex for lower-income countries, whereas most of the high-order dependencies in high-income countries are explained by the pairwise dependencies between indicators. Given that many analyses of SDG interlinkages consider only pairwise relationships, this implies the need to consider high-order interactions to capture relationships in lower-income countries where policy actions targeting pairwise interlinkages could be less effective. The comparison between the Global North and Global South



**Figure 8.** Robustness and efficiency of shifting and permutation resampling strategies. (a) Test power for the same model in its stationary version (Model 1.1 (3.2)), using shifting resampling, (left) and non-stationary version (Model 2.1 (4.2) with linear trend, using permutation resampling, (right)). Relatively few null samples ( $S, P > 100$ ) are enough to attain high test power for both schemes. (b) For stationary time series with multiple realizations, both shifting and permutation can be employed. Shifting is preferred if the number of time observations ( $T$ ) is large relative to the number of realizations ( $n$ ); conversely, permutation is preferred if  $n$  is large relative to  $T$ . For Model 1.1 (3.2) with  $T = 20$ , the permutation scheme with  $n = 6$  already reaches comparable performance to shifting, whereas for  $T = 100$  we need  $n = 20$  for permutation resampling to reach comparable performance to shifting. We use  $S = P = 1000$  for all tests in (b).

(figure 7*d,e*) shows that the latter has exclusively two-way dependencies, whereas the former has emergent three-way interactions (12) and four-way interactions (1) (figure 7*e*). Interestingly, two SDGs, climate action and life below water, consistently appear in emergent high-order dependencies in lower- and higher-income countries, and in Global North groupings, suggesting their potential for further studies. In addition, the hypergraphs of emergent high-order interactions for different country groupings can be studied using network science techniques, including the computation of centrality measures to rank the importance of SDGs within the system of interdependent SDG objectives and the use of community detection algorithms to extract clusters of highly interdependent SDGs [12].

## 5. Discussion

In this paper, we present dHSIC tests for joint independence in both stationary and non-stationary time-series data. For single realizations of stationary time series, we employ a random shifting method as a resampling technique. In the case of multiple realizations of either stationary or non-stationary time series, we consider each realization as an independent sample from a multivariate probability distribution, enabling us to use random permutation as a resampling strategy. To validate our approach, we conducted experiments on diverse synthetic examples, successfully recovering ground truth relationships, including in the presence of a variety of non-stationary behaviours. As illustrated by applications to climate, SDGs, and MRI and Alzheimer's data, the testing framework could be applicable to diverse scientific areas in which stationary or non-stationary time series are the norm.

There are some computational considerations that need to be taken into account for different applications. In our numerical experiments, we have evaluated the impact of several parameters, including the length of the time series  $T$  and the number of observations  $n$ , on the computational efficiency and statistical power of our test. In general, the test statistic can be computed in  $\mathcal{O}(dT^2)$  or  $\mathcal{O}(dn^2)$ , where  $d$  is the number of variables and  $T^2$  or  $n^2$  are the sizes of the kernel matrices [19].

Hence, the computational cost increases with the number of variables and/or number of realizations and length of the time series. The computational cost also grows linearly with the number of resamplings ( $S$  or  $P$ ) used to approximate the null distribution, but our findings show that the test is robust even for low numbers of resamplings. Figure 8a shows that the test power does not improve substantially beyond 100 resamplings (permutations)—a result that has been previously discussed for *iid* data [47]. Therefore, achieving a balance between test power and computational efficiency is crucial, particularly when dealing with large multivariate datasets.

It is worth noting that for stationary data with multiple independent realizations, both resampling schemes (shifting and permutation) can be employed to sample the null distribution. If the number of realizations ( $n$ ) is much larger than the length of the time series ( $T$ ), the permutation strategy provides more efficient randomization as long as the realizations are diverse. Conversely, when  $n$  is smaller than  $T$ , time shifting allows to better exploit the observed temporal dynamics. As an illustration of this point for Model 1.1 (3.2) with multiple realizations, figure 8b shows that if we have  $T = 20$  time points available, then the permutation-based approach has the same performance as the shifting approach when the number of realizations reaches  $n = 6$ . However, if  $T = 100$  time points are available, both performances become similar when the number of realizations is  $n = 20$ . These resampling alternatives must also be evaluated in conjunction with the study of different kernels that can more effectively capture the temporal structure within and across time series (e.g. signature kernels). We leave the investigation of these areas as an avenue of future research.

The interest in higher-order networks, such as hypergraphs or simplicial complexes, has been steadily growing [24] with applications across scientific fields [22,48–51]. Higher-order networks can be natural formalizations of relational data linking  $d$  entities [52,53]. However, there is a scarcity of research and a lack of consensus on how to *construct* higher-order networks from observed *iid* or time-series data [54], and the joint independence methods proposed here could serve to complement approaches based on information measures [20]. By iteratively testing from pairwise independence up to  $d$ -order joint independence, our approach can uncover emergent dependencies not explained by lower-order relationships. This framework presents a direction for the development of higher-order networks, bridging the gap between observed data and the construction of meaningful higher-order network representations.

**Ethics.** Ethics approval was not required for this study.

**Data accessibility.** Climate data: <http://dx.doi.org/10.24432/C5JS49>, SDG data: <https://datacatalog.worldbank.org/dataset/sustainable-development-goals>, and MRI and Alzheimer's data: <https://www.kaggle.com/datasets/jboysen/mri-and-alzheimers>. Synthetic examples: the code to generate the synthetic data and the algorithms to implement the tests are available at [https://github.com/barahona-research-group/dHSIC\\_ts](https://github.com/barahona-research-group/dHSIC_ts).

**Declaration of AI use.** We have not used AI-assisted technologies in creating this article.

**Authors' contributions.** Z.L.: conceptualization, formal analysis, investigation, methodology, visualization, writing—original draft, writing—review and editing; R.L.P.: supervision, visualization, writing—review and editing; F.L.: data curation, writing—review and editing; S.V.M.: writing—review and editing; M.B.: conceptualization, supervision, writing—review and editing.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

**Conflict of interest declaration.** We declare we have no competing interests.

**Funding.** M.B. acknowledges support by EPSRC grant no. EP/N014529/1 funding the EPSRC Centre for Mathematics of Precision Healthcare at Imperial, and by the Nuffield Foundation under the project 'The Future of Work and Well-being: The Pissarides Review'. R.L.P. acknowledges funding from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) Project-ID 424778381-TRR 295.

**Acknowledgements.** We thank Jianxiong Sun for valuable discussions and Asem Alaa for help in maintaining the GitHub repository.

## References

1. Fulcher BD, Jones NS. 2017 htcsa: a computational framework for automated time-series phenotyping using massive feature extraction. *Cell Syst.* **5**, 527–531.e3. (doi:10.1016/j.cels.2017.10.001)
2. Champion K, Lusch B, Kutz JN, Brunton SL. 2019 Data-driven discovery of coordinates and governing equations. *Proc. Natl Acad. Sci. USA* **116**, 22 445–22 451. (doi:10.1073/pnas.1906995116)
3. Liu Z, Barahona M. 2021 Similarity measure for sparse time course data based on Gaussian processes. In *Uncertainty in Artificial Intelligence*, pp. 1332–1341. PMLR.
4. Saavedra-Garcia P *et al.* 2021 Systems level profiling of chemotherapy-induced stress resolution in cancer cells reveals druggable trade-offs. *Proc. Natl Acad. Sci. USA* **118**, e2018229118. (doi:10.1073/pnas.2018229118)
5. Duchon C, Hale R. 2012 *Time series analysis in meteorology and climatology: an introduction*. UK: John Wiley & Sons.
6. Katz RW, Skaggs RH. 1981 On the use of autoregressive-moving average processes to model meteorological time series. *Mon Weather Rev.* **109**, 479–484. (doi:10.1175/1520-0493(1981)109<0479:otuoam>2.0.co;2)

7. Mills TC. 1990 *Time series techniques for economists*. Cambridge, UK: Cambridge University Press.
8. Siami-Namini S, Namin AS. 2018 Forecasting economics and financial time series: ARIMA vs. LSTM. (<http://arxiv.org/abs/1803.06386>)
9. Peach RL, Yaliraki SN, Lefevre D, Barahona M. 2019 Data-driven unsupervised clustering of online learner behaviour. *npj Sci. Learn.* **4**, 1–11. (doi:10.1038/s41539-019-0054-0)
10. Peach RL, Greenbury SF, Johnston IG, Yaliraki SN, Lefevre DJ, Barahona M. 2021 Understanding learner behaviour in online courses with Bayesian modelling and time series characterisation. *Sci. Rep.* **11**, 2823. (doi:10.1038/s41598-021-81709-3)
11. Tosepu R, Gunawan J, Effendy DS, Lestari H, Bahar H, Asfian P. 2020 Correlation between weather and COVID-19 pandemic in Jakarta, Indonesia. *Sci. Total Environ.* **725**, 138436. (doi:10.1016/j.scitotenv.2020.138436)
12. Laumann F, von Kügelgen J, Uehara THK, Barahona M. 2022 Complex interlinkages, key objectives, and nexuses among the Sustainable Development Goals and climate change: a network analysis. *Lancet Planet. Health* **6**, e422–e430. (doi:10.1016/s2542-5196(22)00070-5)
13. Baumöhl E. 2019 Are cryptocurrencies connected to Forex? A quantile cross-spectral approach. *Finance Res. Lett.* **29**, 363–372. (doi:10.1016/j.frl.2018.09.002)
14. Mokhtari F, Laurienti PJ, Rejeski WJ, Ballard G. 2019 Dynamic functional magnetic resonance imaging connectivity tensor decomposition: a new approach to analyze and interpret dynamic brain connectivity. *Brain Connect.* **9**, 95–112. (doi:10.1089/brain.2018.0605)
15. Guan S, Jiang R, Bian H, Yuan J, Xu P, Meng C, Biswal B. 2020 The profiles of non-stationarity and non-linearity in the time series of resting-state brain networks. *Front. Neurosci.* **14**, 493. (doi:10.3389/fnins.2020.00493)
16. Miorandi D, Sicari S, De Pellegrini F, Chlamtac I. 2012 Internet of things: vision, applications and research challenges. *Ad Hoc Netw.* **10**, 1497–1516. (doi:10.1016/j.adhoc.2012.02.016)
17. Vogelstein JT, Bridgeford EW, Wang Q, Priebe CE, Maggioni M, Shen C. 2019 Discovering and deciphering relationships across disparate data modalities. *Elife* **8**, e41690. (doi:10.7554/elife.41690)
18. Mishra P, Singh U, Pandey CM, Mishra P, Pandey G. 2019 Application of student's t-test, analysis of variance, and covariance. *Ann. Card. Anaesth.* **22**, 407–411. (doi:10.4103/aca.aca\_94\_19)
19. Pfister N, Bühlmann P, Schölkopf B, Peters J. 2018 Kernel-based tests for joint independence. *J. R. Stat. Soc. B (Statistical Methodol.)* **80**, 5–31. (doi:10.1111/rssb.12235)
20. Rosas FE, Mediano PA, Gastpar M, Jensen HJ. 2019 Quantifying high-order interdependencies via multivariate extensions of the mutual information. *Phys. Rev. E* **100**, 032305. (doi:10.1103/physreve.100.032305)
21. Battiston F *et al.* 2021 The physics of higher-order interactions in complex systems. *Nat. Phys.* **17**, 1093–1098. (doi:10.1038/s41567-021-01371-4)
22. Arnaudon A, Peach RL, Petri G, Expert P. 2022 Connecting Hodge and Sakaguchi-Kuramoto through a mathematical framework for coupled oscillators on simplicial complexes. *Commun. Phys.* **5**, 211. (doi:10.1038/s42005-022-00963-7)
23. Nuriso M, Arnaudon A, Lucas M, Peach RL, Expert P, Vaccarino F, Petri G. 2023 A unified framework for Simplicial Kuramoto models. (<http://arxiv.org/abs/2305.17977>)
24. Battiston F, Cencetti G, Iacopini I, Latora V, Lucas M, Patania A, Young JG, Petri G. 2020 Networks beyond pairwise interactions: structure and dynamics. *Phys. Rep.* **874**, 1–92. (doi:10.1016/j.physrep.2020.05.004)
25. Gretton A, Fukumizu K, Teo C, Song L, Schölkopf B, Smola A. 2007 A Kernel statistical test of independence. In *Advances in neural information processing systems*, vol. 20.
26. Chwiałkowski K, Gretton A. 2014 A Kernel independence test for random processes. In *Int. Conf. on Machine Learning*, pp. 1422–1430. PMLR.
27. Chwiałkowski KP, Sejdinovic D, Gretton A. 2014 A wild bootstrap for degenerate kernel tests. In *Advances in neural information processing systems*, vol. 27.
28. Laumann F, Kügelgen J, Barahona M. 2021 Kernel two-sample and independence tests for nonstationary random processes. *Eng. Proc.* **5**, 31. (doi:10.3390/engproc2021005031)
29. Muandet K, Fukumizu K, Sriperumbudur B, Schölkopf B. 2017 Kernel mean embedding of distributions: a review and beyond. *Found. Trends Mach. Learn.* **10**, 1–141. (doi:10.1561/9781680832891)
30. Sriperumbudur BK, Gretton A, Fukumizu K, Schölkopf B, Lanckriet GR. 2010 Hilbert space embeddings and metrics on probability measures. *J. Mach. Learn. Res.* **11**, 1517–1561.
31. Sriperumbudur BK, Fukumizu K, Lanckriet GR. 2011 Universality, characteristic kernels and RKHS embedding of measures. *J. Mach. Learn. Res.* **12**, 2389–2410.
32. Fukumizu K, Gretton A, Sun X, Schölkopf B. 2007 Kernel measures of conditional dependence. In *Advances in neural information processing systems*, vol. 20.
33. Gretton A, Borgwardt KM, Rasch MJ, Schölkopf B, Smola A. 2012 A Kernel two-sample test. *J. Mach. Learn. Res.* **13**, 723–773.
34. Besserve M, Logothetis NK, Schölkopf B. 2013 Statistical analysis of coupled time series with kernel cross-spectral density operators. In *Advances in neural information processing systems*, vol. 26.
35. Sejdinovic D, Gretton A, Bergsma W. 2013 A kernel test for three-variable interactions. In *Advances in neural information processing systems*, vol. 26.
36. Shao X. 2010 The dependent wild bootstrap. *J. Am. Stat. Assoc.* **105**, 218–235. (doi:10.1198/jasa.2009.tm08744)
37. Mehta R, Chung J, Shen C, Xu T, Vogelstein JT. 2019 Independence testing for multivariate time series. (<http://arxiv.org/abs/1908.06486>)
38. Rubenstein PK, Chwiałkowski KP, Gretton A. 2016 A Kernel test for three-variable interactions with random processes. (<http://arxiv.org/abs/1603.00929>)
39. Luff C, Peach RL, Mallas EJ, Rhodes E, Laumann F, Boyden ES, Sharp DJ, Barahona M, Grossman N. 2023 The neuron mixer and its impact on human brain dynamics. *bioRxiv*. (doi:10.1101/2023.01.05.522833)
40. Haufler D, Paré D. 2019 Detection of multiway gamma coordination reveals how frequency mixing shapes neural dynamics. *Neuron* **101**, 603–614. (doi:10.1016/j.neuron.2018.12.028)
41. Kleinfeld D, Mehta SB. 2006 Spectral mixing in nervous systems: experimental evidence and biologically plausible circuits. *Progress Theor. Phys. Suppl.* **161**, 86–98. (doi:10.1143/ptps.161.86)
42. Liang X, Zou T, Guo B, Li S, Zhang H, Zhang S, Huang H, Chen SX. 2015 Assessing Beijing's PM<sub>2.5</sub> pollution: severity, weather impact, APEC and winter heating. *Proc. R. Soc. A* **471**, 20150257. (doi:10.1098/rspa.2015.0257)
43. Said SE, Dickey DA. 1984 Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika* **71**, 599–607. (doi:10.1093/biomet/71.3.599)
44. Boysen J. 2017 MRI and Alzheimers. [https://www.kaggle.com/datasets/jboysen/mri-and-alzheimers/data?select=oasis\\_longitudinal.csv](https://www.kaggle.com/datasets/jboysen/mri-and-alzheimers/data?select=oasis_longitudinal.csv) (accessed 28 January 2020).
45. Liu Z, Peach RL, Mediano PA, Barahona M. 2023 Interaction measures, partition lattices and Kernel tests for high-order interactions. In *Advances in neural information processing systems*, vol. 37.
46. World Bank. Sustainable Development Goals. 2020. <https://datacatalog.worldbank.org/dataset/sustainable-development-goals> (accessed 28 January 2020).
47. Rindt D, Sejdinovic D, Steinsaltz D. 2021 Consistency of permutation tests of independence using distance covariance, HSIC and dHSIC. *Stat* **10**, e364. (doi:10.1002/sta4.364)
48. Gao T, Li F. 2018 Studying the utility preservation in social network anonymization via persistent homology. *Comput. Secur.* **77**, 49–64. (doi:10.1016/j.cose.2018.04.003)
49. Arenzon JJ, de Almeida RM. 1993 Neural networks with high-order connections. *Phys. Rev. E* **48**, 4060. (doi:10.1103/physreve.48.4060)
50. Sonntag M, Teichert HM. 2004 Competition hypergraphs. *Discrete Appl. Math.* **143**, 324–329. (doi:10.1016/j.dam.2004.02.010)
51. Klamt S, Haus UU, Theis F. 2009 Hypergraphs and cellular networks. *PLoS Comput. Biol.* **5**, e1000385. (doi:10.1371/journal.pcbi.1000385)
52. White JG, Southgate E, Thomson JN, Brenner S. 1986 The structure of the nervous system of the nematode *Caenorhabditis elegans*. *Phil. Trans. R. Soc. Lond. B* **314**, 1–340. (doi:10.1098/rstb.1986.0056)
53. Atkin RH. 1974 *Mathematical structure in human affairs*. London: Heinemann Educational Books.
54. Schneidman E, Still S, Berry MJ, Bialek W. 2003 Network information and connected correlations. *Phys. Rev. Lett.* **91**, 238701. (doi:10.1103/PhysRevLett.91.238701)