A Consistent Independence Test for Multivariate Time-Series

Ronak Mehta [1], Cencheng Shen [2], Ting Xu [3], and Joshua T. Vogelstein [1,4]

[1] Department of Biomedical Engineering, Johns Hopkins University
[2] Department of Applied Economics and Statistics, University of Delaware
[3] Center for the Developing Brain, Child Mind Institute
[4] Institute for Computational Medicine, Kavli Neuroscience Discovery Institute, Johns Hopkins University

A fundamental problem in statistical data analysis is testing whether two phenomena are related. When the phenomena in question are time series, many challenges emerge. The first is defining a dependence measure between time series at the population level, as well as a sample level test statistic. The second is computing or estimating the distribution of this test statistic under the null, as the permutation test procedure is invalid for most time series structures. This work aims to address these challenges by combining distance correlation and multiscale graph correlation (`MGC`) from independence testing literature and block permutation testing from time series analysis. Two hypothesis tests for testing the independence of time series are proposed. These procedures also characterize whether the dependence relationship between the series is linear or nonlinear, and the time lag at which this dependence is maximized. For strictly stationary auto-regressive moving average (ARMA) processes, the proposed independence tests are proven valid and consistent. Finally, neural connectivity in the brain is analyzed using fMRI data, revealing linear dependence of signals within the visual network and default mode network, and nonlinear relationships in other regions. This work opens up new theoretical and practical directions for many modern time series analysis problems.

1 Introduction

Time series data are ubiquitous in fields such as neuroscience, finance, and sociology. For this reason, in many data analysis and machine learning settings, one wishes to determine the relationship between two jointly-observed time series (X_t, Y_t). While further analysis can involve predicting the value of one series given an observation of the other, or analyzing the geometry of the relationship, the first step is determining whether a discernible relationship exists [19]. In the case of phenomena X and Y for which independent and identically-distributed (i.i.d.) observations can be drawn, this question translates to whether the random variables X and Y are independent. In the time series setting, this question translates to whether to whether components of the time series X_t and Y_{t-j} spaced j timesteps apart are independent for various values of j.

Current approaches to general independence testing include kernel-methods such as Hilbert-Schmidt Information Criterion (`HSIC`) [6, 7], and distance based methods such as distance correlation (`DCorr`) [17, 18]. Sejdinovic et al. [14], Shen and Vogelstein [15] has shown distance and kernel-based methods are equivalent at both the population and sample levels, and only differ because their kernel or metric choice is different. Multiscale graph correlation (`MGC`) [16, 19] is an optimized local version of `DCorr` that is highly successful in terms of 1) applications to virtually any modality of data (such as Euclidean data, shapes, images, networks), 2) characterizing many types of geometric relationships, and 3) significant better power at low sample sizes and high dimensionality. However, almost all correlations and dependency measures (including `DCorr`, `HSIC`, and `MGC`) assume that observations from either modality are i.i.d., which is usually not true in general for time series.

For temporally-dependent data such as functional magnetic resonance images (fMRI), dynamic social networks, and financial indices, independence testing procedures are limited [9]. Researchers currently resort to measures of linear dependence such as autocorrelation and crosscorrelation, ruling out potential nonlinear relationships [9]. The goal of this work is to apply the principles of `DCorr` and `MGC` to time series (among other dependent data). Even after determining a reasonable test statistic to measure dependence between time series, computing or estimating the distribution of the statistic under the null hypothesis of independence presents an additional challenge. While the permutation test procedure is common for many hypothesis tests, it requires the exchangeability of observations in the sample. This property is not satisfied by time series, and the point is also addressed via a block permutation test approach.

This work proposes statistical independence tests for two multidimensional time series: cross-distance correlation (`DCorr-X`) and cross multiscale graph correlation (`MGC-X`), based on `DCorr` and `MGC`, respectively. A reasonable definition of dependence between time series is proposed both for the population and sample level, as well as a block permutation procedure to estimate the null distribution and p-value. Additionally, our procedure estimates the time lag at which this dependence is maximized, further characterizing the temporal nature of the relationship. Validity and consistency results are shown for the tests. Finally, an analysis neural connectivity via fMRI data and `MGC-X` is presented. These methods expand the scope of traditional independence testing to complex, non-i.i.d. settings, accelerating research capabilities in neuroscience, econometrics, sociology, and countless other fields.

2 Preliminaries

2.1 Notation

Let $\mathbb{Z}$ be the integers $\{..., -1, 0, 1, ...\}$, $\mathbb{N}$ be the non-negative integers $\{0, 1, 2, ...\}$, and $\mathbb{R}$ be the real line $(-\infty, \infty)$. Let F_X, F_Y, and $F_{X,Y}$ represent the marginal and joint distributions of random variables X and Y, whose realizations exist in $\mathcal{X}$ and $\mathcal{Y}$, respectively. Similarly, Let F_{X_t}, F_{Y_s}, and $F_{(X_t,Y_s)}$ represent the marginal and joint distributions of the time-indexed random variables X_t and Y_s at timesteps t and s. For this work, assume $\mathcal{X} = \mathbb{R}^p$ and $\mathcal{Y} = \mathbb{R}^q$ for $p, q \in \mathbb{Z}^+$. Finally, let $\{(X_t, Y_t)\}_{t=-\infty}^{\infty}$ represent the full, jointly-sampled time series, structured as a countably long list of observations (X_t, Y_t).

2.2 Problem Statement

To formalize the problem, consider a strictly stationary time series $\{(X_t, Y_t)\}_{t=-\infty}^{\infty}$, with the observed sample $\{(X_1, Y_1), ..., (X_n, Y_n)\}$. Choose some $M \in \mathbb{N}$, the "maximum lag" hyperparameter. We test the independence of two series via the following hypothesis.

$$
\begin{aligned}
H_0 &: F_{(X_t,Y_{t-j})} = F_{X_t}F_{Y_{t-j}} \text{ for each } j \in \{0, 1, ..., M\} \\
H_A &: F_{(X_t,Y_{t-j})} \neq F_{X_t}F_{Y_{t-j}} \text{ for some } j \in \{0, 1, ..., M\}
\end{aligned}
$$

The null hypothesis implies that for any $(M+1)$-length stretch in the time series, X_t is pairwise independent of present and past values Y_{t-j} spaced j timesteps away (including $j = 0$). A corresponding test for whether Y_t is dependent on past values of X_t is available by swapping the labels of each time series. Finally, the hyperparameter M governs the maximum number of timesteps in the past for which we check the influence of Y_{t-j} on X_t. This M can be chosen for computation considerations, as well as for specific subject matter purposes, e.g. a signal from one region of the brain might only influence be able to influence another within 20 time steps implies $M = 20$.

We frame the solution in three parts. First, we explore distributional parameters that are quantitative measures of dependence. These parameters are theoretical and can only be computed with knowledge of the distribution that governs (X_t, Y_t). Second, we investigate corresponding sample estimators for these population parameters, and from these estimators we propose a test statistic indicating the level of dependence between the time series. Finally, we estimate the sampling distribution of the test statistic under the null to determine statistical significance. A simple permutation test cannot be used, as the observed time series $\{X_t\}$ and $\{Y_t\}$ are not exchangeable, which motivates a modified procedure.

2.3 Distance covariance and correlation

First, we describe a suitable measure of dependence. Consider random variables $X \in \mathbb{R}^p$ and $Y \in \mathbb{R}^q$, with finite first and second moments. The *distance covariance* function is defined as:

$$
\begin{aligned}
d\text{cov}(X, Y) = &\ \mathbb{E}_{XY}\mathbb{E}_{X'Y'}\left[||X - X'||_2||Y - Y'||_2\right] \\
&+ \mathbb{E}_X\mathbb{E}_{X'}||X - X'||_2 \cdot \mathbb{E}_Y\mathbb{E}_{Y'}||Y - Y'||_2 \\
&- 2\mathbb{E}_{XY}\left[\mathbb{E}_{X'}||X - X'||_2 \cdot \mathbb{E}_{Y'}||Y - Y'||_2\right]
\end{aligned}
$$

X' and Y' are independent copies of X and Y respectively. $d\text{cov}(X, Y)$ is 0 if and only if X and Y are independent and non-zero otherwise [14, 18]. We can also represent this function as

$$
d\text{cov}(X, Y) = \int_{\mathbb{R}^p \times \mathbb{R}^q} |\mathbb{E}[g_{XY}(u, v)] - \mathbb{E}[g_X(u)]\mathbb{E}[g_Y(v)]]|^2 w(u, v)dudv
$$

where $\mathbb{E}[g_{XY}(u, v)] = \mathbb{E}[e^{i(u^TX+v^TY)}]$ is the joint characteristic function of (X, Y), $\mathbb{E}[g_X(u)] = \mathbb{E}[e^{iu^TX}]$ and $\mathbb{E}[g_Y(v)] = \mathbb{E}[e^{iv^TY}]$ represent the marginal characteristic functions with $i = \sqrt{-1}$. $w(u, v)$ is a weight function defined by:

$$
\begin{aligned}
w(u, v) &= (c_pc_q|u|^{1+p}|v|^{1+q})^{-1} \\
c_p &= \frac{\pi^{(1+p)/2}}{\Gamma((1 + p)/2)} \\
c_q &= \frac{\pi^{(1+q)/2}}{\Gamma((1 + q)/2)}
\end{aligned}
$$

In this representation, it is more apparent why independence would force the joint characteristic function to be equal to the product of the marginals, and the integral to be 0.

This function can be normalized to the *distance correlation* function $d$corr as such:

$$
d\text{corr}(X, Y) = \begin{cases} \frac{d\text{cov}(X,Y)}{\sqrt{d\text{cov}(X,X)d\text{cov}(Y,Y)}} & d\text{cov}(X, X)d\text{cov}(Y, Y) > 0 \\ 0 & d\text{cov}(X, X)d\text{cov}(Y, Y) = 0 \end{cases}
$$

$d\text{corr}(X, Y)$ is bounded between 0 and 1, similar to the absolute value of Pearson's correlation. Unlike the case of Pearson's correlation, however, dependent but uncorrelated random variables will have nonzero values for $d\text{cov}(X, Y)$ and $d\text{corr}(X, Y)$. Take for example, $X \sim \mathcal{N}(0, 1)$ and $Y = X^2$. $\text{Cov}(X, Y) = 0$ while $d\text{corr}(X, Y) \approx 0.782$. Therefore, distance covariance and correlation make a desirable measure of dependence, whether linear or nonlinear.

Another common measure of dependence is the Hilbert-Schmidt Information Criterion (HSIC) whose estimator operates on the kernel matrices K_X and K_Y of the sample $\{(X_i, Y_i)\}_{i=1}^n$. These methods are equivalent with distance-based methods due to a bijective mapping between distance functions and kernels shown in Shen and Vogelstein [15]. Our task will be to apply this distance-based measure to a joint time series process.

2.4 Empirical estimates

We now show the sample estimates of these quantities. Given sample $\{(X_1, Y_1), ..., (X_n, Y_n)\}$, an empirical estimate $d\text{cov}_n(X, Y)$ is computed as follows. Generate two $n \times n$ distance matrices $[a_{ij}] = ||X_i - X_j||_2$ and $[b_{ij}] = ||Y_i - Y_j||_2$, respectively. Column center the matrices $[a_{ij}]$ and $[b_{ij}]$. This yields matrices A and B:

$$A_{ij} = a_{ij} - \frac{1}{n-1} \sum_{s=1}^{n} a_{sj}$$

Define B_{ij} equivalently. Finally, compute:

$$d\text{cov}_n(X, Y) = \frac{1}{n(n-1)} \sum_{i=1, j=1}^{n} A_{ij} B_{ji}$$

Column centering guarantees consistency of the estimator for the population $d\text{corr}$ [16]. Other centering schemes and their respective advantages are discussed in Shen et al. [16].

2.5 Multiscale Graph Correlation (`MGC`)

`MGC` builds on the distance correlation test statistic by retaining only the distances that are most informative toward the relationship between X and Y. Specifically, let A and B be the centered distance matrices above. Define G^k and H^l to be the k-nearest and l-nearest neighbor matrices, respectively. $G^k_{ij} = 1$ indicates that A_{ij} is within the smallest k values of the i-th row of A, and similarly for H^l [16, 19]. Define:

$$d\text{cov}_n^{k,l}(X, Y) = \sum_{i,j} A_{ij} G^k_{ij} B_{ij} H^l_{ij}$$

$$d\text{corr}_n^{k,l} = \frac{d\text{cov}_n^{k,l}(X, Y)}{\sqrt{d\text{cov}_n^{k,k}(X, X) \times d\text{cov}_n^{l,l}(Y, Y)}}$$

$$\text{mgc}_n(X, Y) = \text{smoothed max}_{k,l} \left\{ d\text{corr}_n^{k,l} \right\}$$

The test statistic $\text{mgc}_n(X, Y)$ is the smoothed maximum of the $d\text{corr}_n^{k,l}$ over k and l, giving this statistic better finite-sample performance, as opposed to dcorr [16], for many nonlinear settings. The test also returns k and l that achieve this maximum. If both k and l are equal to n, then the relationship is assumed to be linear [19]. That `MGC` provides information about the geometric nature of the relationship has proven to be a useful feature.

2.6 Linear dependence in time series

Because our problem of interest involves time series specifically, we describe other dependence measures for time series. For a stationary time series, $\{X_t\}_{t=-\infty}^{\infty}$, linear dependence is measured with the *autocovariance* function (ACVF) at lag $j \in \mathbb{N}$.

$$\text{ACVF}(j) = \text{Cov}(X_t, X_{t+j})$$

The normalized *autocorrelation* function (ACF) is defined as:

$$\text{ACF}(j) = \frac{\text{ACVF}(j)}{\text{ACVF}(0)}$$

To compute the empirical estimates, compute the empircal Pearson's correlation on the dataset $\{(X_1, X_{j+1}), \dots, (X_{n-j}, X_n)\}$.

Analogously, given two stationary time series $\{(X_t, Y_t)\}_{t=-\infty}^{\infty}$, the *crosscovariance* function (CCVF) at lead/lag $j \in \mathbb{Z}$ is defined as:

$$\mathsf{CCVF}(j) = \mathsf{Cov}(X_t, Y_{t+j})$$

The normalized *crosscorrelation* function (CCF) is defined as:

$$\mathsf{CCF}(j) = \frac{\mathsf{CCVF}(j)}{\sqrt{\mathsf{ACVF}_{X_t}(0)\mathsf{ACVF}_{Y_t}(0)}}$$

CCF measures pairwise linear dependence between components from either time series. The sample equivalents are computed when given $\{X_1, ..., X_{n-j}\}$ and $\{Y_{j+1}, ..., Y_n\}$ by taking the sample covariance and correlation. Plotting the estimates of this function is often the first resort for a researcher wishing to investigate the relationships within observations (ACF) or between observations (CCF). A result due to Bartlett gives 95% confidence bands $\pm\frac{1.96}{\sqrt{n}}$ for these estimates, and statistical programs such as $\mathtt{R}$ typically automatically overlay them on ACF and CCF plots made using built-in functions. However, this confidence interval is only accurate for linear time series, as in $X_t = \mu + \sum_{i=-\infty}^{\infty} \psi_i Z_{t-i}$ where $\mu, \phi_j \in \mathbb{R}$ and $Z_j \sim \mathcal{N}(0, \tau^2)$ for $\tau^2 > 0$. For nonlinear models, these bands can be misleading, and ad-hoc analysis of the ACF or CCF plot can lead to false conclusions [13]. The well-known Ljung-Box statistic makes use of these functions in testing for autocorrelatedness of time series.

$$\mathsf{LB} = n(n+2)\sum_{j=1}^{M} \frac{\mathsf{ACF}^2(j)}{n-j}$$

While suitable for tests for linear dependence, as a test for general dependence, LB performs poorly [3]. We are therefore interested in estimating a function similar to the CCVF/CCF that will capture nonlinear dependencies.

2.7 Other dependence measures in time series

Recent literature addresses nonlinear forms of dependence in time series via adaptations of frequency-domain methods. The *standardized spectral density* of a stationary time series is defined as

$$h(\omega) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \mathsf{ACF}(j) e^{-ij\omega},$$

the Fourier transform of the ACF. In white noise settings (the time series in uncorrelated with itself at various lags), $h(\omega)$ is uniform, as all frequencies are represented equally in the spectrum. Deviation from uniformity implies autocorrelatedness of the time series. Similarly, Hong [10] defines the *generalized spectral density*

$$f(\omega, u, v) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \sigma_j(u, v) e^{-ij\omega}$$

with $\sigma_j(u, v) = |\mathbb{E}[g_{X_t, X_{t-j}}(u, v)] - \mathbb{E}[g_{X_t}(u)]\mathbb{E}[g_{X_{t-j}}(v)]|^2$. Deviations of the generalized spectral density from uniformity imply dependence [10]. To generate a test statistic based on this observation, Fokianos and Pitsillou [5] use:

$$\int_{-\pi}^{\pi} ||\hat{f}_n(\omega, u, v) - \hat{f}_0(\omega, u, v)||_w^2 d\omega$$

where $\hat{f}_n(\omega, u, v)$ is an estimate of the generalized spectral density, and $\hat{f}_0(\omega, u, v) = \frac{1}{2\pi}\hat{\sigma}_0(u, v)$ is an estimate of the uniform density under the assumption of independence, and $|| \cdot ||_w$ is a weighted L_2-norm with respect to some weight function $w(u, v)$. Fokianos and Pitsillou [5] gives kernel-type estimators for the spectrum (as is common in spectral estimation [1]), and using the weight function described in Section 2.3, the statistic simplifies to a Ljung-Box type:

$$\int_{-\pi}^{\pi} ||\hat{f}_n(\omega, u, v) - \hat{f}_0(\omega, u, v)||_w^2 d\omega = \frac{2}{\pi} \sum_{j=1}^{n-1} \left(1 - \frac{j}{n}\right) \kappa^2 \left(\frac{j}{p}\right) d\text{cov}_n(X_t, X_{t-j})$$

where $\kappa(\cdot)$ is a kernel function, and p is a bandwidth parameter. Fokianos and Pitsillou [5] offer consistency results for a test of serial dependence based on this test statistic, when $p = cn^\lambda$ for $c > 0$ and $\lambda \in (0, 1)$. We can adapt this methodology, as well as that of `MGC`, for testing of dependence *between* (rather than *within*) stationary time series.

3 The `DCorr-X` and `MGC-X` independence tests

Recall the problem statement in Section 2.2, that is given $\{(X_1, Y_1), ..., (X_n, Y_n)\} \sim F_{\{(X_t,Y_t)\}_{t=1}^n}$, we test:

$$\begin{aligned} H_0 &: F_{(X_t, Y_{t-j})} = F_{X_t} F_{Y_{t-j}} \quad \text{for all } j \in \{0, 1, ..., M\}, \\ H_A &: F_{(X_t, Y_{t-j})} \neq F_{X_t} F_{Y_{t-j}} \quad \text{for some } j \in \{0, 1, ..., M\} \end{aligned}$$

Our proposed procedure comprises of a suitable measure of dependence, its estimator used to derive a test statistic, and resampling procedure as a way to estimate the p-value of the observed test statistic.

The test statistics

Define the *cross-distance correlation* at lag j as

$$d\text{corr}(j) := d\text{corr}(X_t, Y_{t-j}).$$

Where $d\text{corr}(\cdot, \cdot)$ is the distance correlation function described in Section 2.3. Assuming strict stationarity of $\{(X_t, Y_t)\}$ is important in even defining $d\text{corr}(j)$, as the parameter depends only on the spacing j, and not the timestep t of X_t and Y_{t-j}. Similarly, let $d\text{corr}_n(j)$ be its estimator, with $\text{mgc}_n(j)$ being the `MGC` test statistic evaluated for $\{X_t\}$ and $\{Y_{t-j}\}$. The $\texttt{DCorr-X}^M$ test statistic is

$$d\text{corr-X}_n^M = \sum_{j=0}^{M} \left(\frac{n-j}{n}\right) \cdot d\text{corr}_n(j).$$

Similarly, the `MGC-X` test statistic is

$$\text{mgc-X}_n^M = \sum_{j=0}^{M} \left(\frac{n-j}{n}\right) \cdot \text{mgc}_n(j).$$

While `MGC-X` is more computationally intensive than `DCorr-X`, `MGC-X` employs multiscale analysis to achieve better finite-sample power in high-dimensional, nonlinear, and structured data settings [11, 16, 19].

The block permutation test

Let T_n represent either of the test statistics above. To compute the p-value, one need to estimate the null distribution of T_n, namely its distribution under indepdendence pair of data. A typical permutation test would permute the indices $\{1, 2, 3, ..., n\}$, reorder the series $\{Y_t\}$ according to this permutation, and T_n would be computed on $\{X_t\}$ and the reordered $\{Y_t\}$. This procedure would be repeated K times, generating K samples of the test statistic under the null. This permutation test requires exchangeability of the sequence $\{Y_t\}$, which would be true in the i.i.d. case, but is generally violated in the time series case. Instead, a block permutation captures the dependence between elements of the series, as described in [13]. Letting $\lceil \cdot \rceil$ be the ceiling function, this procedure partitions the list of indices into size b “blocks", and permutes the $\lceil \frac{n}{b} \rceil$ blocks in order to generate samples of the test statistic under the null. Specifically,

1. Choose a random permutation of the indices $\{0, 1, 2, ..., \lceil \frac{n}{b} \rceil\}$.
2. From index i in the permutation, produce block $B_i = (Y_{bi+1}, Y_{bi+2}, ..., Y_{bi+b})$, which is a section of the series $\{Y_t\}$.
3. Let the series $\{Y_{\pi(1)}, ..., Y_{\pi(n)}\} = (B_1, B_2, ..., B_{\frac{n}{b}})$, where π maps indices $\{1, 2, ..., n\}$ to the new, block permuted indices.
4. Compute $T_n^{(r)}$ on the series $\{(X_t, Y_{\pi(t)})\}_{t=1}^n$ for replicate r.

Repeat this procedure K times (typically $K = 100$ or 1000), and let $T_n^{(0)} = T_n$, with:

$$p\text{-value}(T_n) = \frac{1}{K+1} \sum_{r=0}^{K} \mathbb{I}\{T_n^{(r)} \geq T_n\}$$

where $\mathbb{I}\{\cdot\}$ is the indicator function.

Estimating the optimal lag

To give more information as to the nature of the relationship between the two time series, we estimate the lag j that maximizes the dependence between X_t and Y_{t-j} (the “optimal lag"). Denote this M^*. Both procedures yield estimators $\hat{M}$:

$$\hat{M}_T^n = \underset{j}{\text{argmax}} \left(\frac{n-j}{n} \right) \cdot T_n(j)$$

4 Theoretical Results

The following assumptions are shared across the consequent theorems. Proofs are in the Appendix. The assumptions and results are listed once again below.

Assumption 1. $\left\{ \begin{bmatrix} X_t \\ Y_t \end{bmatrix} \right\}_t$ *is a strictly stationary process.*

Assumption 2. $\left\{ \begin{bmatrix} X_t \\ Y_t \end{bmatrix} \right\}_t$ *has finite second moment.*

Assumption 3. *For the block permutation test procedure, we choose a block size b_n and number of* *permutations K such that as $n \to \infty$,*

$$b_n \to \infty,$$
$$\frac{b_n}{n} \to 0,$$
$$K \to \infty$$

Assumption 4. *The following hold.*

$$\sup |F_{X_t,X_{t-j}} - F_{X_t}F_{X_{t-j}}| \to 0 \text{ as } j \to \infty$$
$$\sup |F_{Y_t,Y_{t-j}} - F_{Y_t}F_{Y_{t-j}}| \to 0 \text{ as } j \to \infty$$
$$\sup |F_{X_t,Y_{t-j}} - F_{X_t}F_{Y_{t-j}}| \to 0 \text{ as } j \to \infty$$

Assumption 5. $\left\{\begin{bmatrix} X_t \\ Y_t \end{bmatrix}\right\}_t$ *is nonconstant, and a continuous random variable for all t.*

Next, we adopt the notation from Shen et al. [16], in that $d\text{corr}^{\rho_k,\rho_l}(X,Y)$ is the population local correlation defined as follows. Let $B(c,r)$ be an r-ball about point c, and $\bar{\cdot}$ denote the complex conjugate. Let $(\rho_k,\rho_l) \in [0,1] \times [0,1]$ represent the normalized scale. Finally, let (X,Y) and (X',Y') be sampled independently from F_{XY}, and $\boldsymbol{I}(\cdot)$ be the indicator function.

$$\boldsymbol{I}^{\rho_k}_{X,X'} = \boldsymbol{I}\left(\int_{B(X,||X-X'||)} dF_X(u) \leq \rho_k\right)$$
$$\boldsymbol{I}^{\rho_l}_{Y',Y} = \boldsymbol{I}\left(\int_{B(Y',||Y-Y'||)} dF_Y(u) \leq \rho_l\right)$$
$$h^{\rho_k}_X(u) = (g_X(u)\overline{g_{X'}(u)} - g_X(u)\overline{g_{X''}(u)})\boldsymbol{I}^{\rho_k}_{X,X'}$$
$$h^{\rho_l}_{Y'}(v) = (g_{Y'}(v)\overline{g_Y(v)} - g_{Y'}(v)\overline{g_{Y'''}(v)})\boldsymbol{I}^{\rho_l}_{Y',Y}$$
$$d\text{cov}^{\rho_k,\rho_l}(X,Y) = \int_{\mathbb{R}^p \times \mathbb{R}^q} \left\{\mathbb{E}[h^{\rho_k}_X(u)\overline{h^{\rho_l}_{Y'}(v)}] - \mathbb{E}[h^{\rho_k}_X(u)]\mathbb{E}[\overline{h^{\rho_l}_{Y'}(v)}]\right\} w(u,v)dudv$$
$$d\text{corr}^{\rho_k,\rho_l}(X,Y) = \begin{cases} \frac{d\text{cov}^{\rho_k,\rho_l}(X,Y)}{\sqrt{d\text{cov}^{\rho_k,\rho_l}(X,X)d\text{cov}^{\rho_k,\rho_l}(Y,Y)}} & d\text{cov}^{\rho_k,\rho_l}(X,X)d\text{cov}^{\rho_k,\rho_l}(Y,Y) > 0 \\ 0 & d\text{cov}^{\rho_k,\rho_l}(X,X)d\text{cov}^{\rho_k,\rho_l}(Y,Y) = 0 \end{cases}$$

For our purposes, let $d\text{corr}^{\rho_k,\rho_l}(j)$ be the population distance correlation evaluated at evaluated at (X_t, Y_{t-j}) (with corresponding independent copy (X'_t, Y'_{t-j})). $d\text{corr}^{k,l}_n(j)$ represents the sample local correlation as before evaluated for time series X_t and Y_{t-j}, with $\text{mgc}_n(j) = \text{smoothed max}_{k,l}\left\{d\text{corr}^{k,l}_n(j)\right\}$. $\pi : \{1,2,...,n\} \to \{1,2,...,n\}$ is a mapping that represents the permutation. Note that $d\text{corr}^{1,1}(j) = d\text{corr}(j)$. In the following statements, we drop the argument j, keeping in mind that the sample and population local correlation can be applied to a shifted time series Y_{t-j}.

Theorem 1. *For any $\epsilon > 0$, there exists a sufficiently large n such that*

$$\mathbb{E}\left[d\text{cov}^{k,l}_n\right] = d\text{cov}^{\rho_k,\rho_l} + \mathcal{O}\left(\frac{1}{n}\right) + \mathcal{O}(\epsilon)$$
$$\text{Var}\left[d\text{cov}^{k,l}_n\right] = \mathcal{O}\left(\frac{1}{n}\right) + \mathcal{O}(\epsilon)$$

Theorem 2. *For each $j \in \{0, ..., M\}$, $d$cov$_n^{k,l}(j) \overset{n\to\infty}{\to} dcov^{\rho_k,\rho_l}(j)$ in probability.*

Theorem 1 elucidates the bias and variance of our estimate of the population local correlation as a function of the sample size n, allowing Theorem 2 to confirm that the estimate approaches the population parameter as the sample size increases. This has implications for the validity and consistency of the test.

Next, Theorem 3 shows the validity of the block permutation procedure for estimating the p-value. Let T_n^* be the observed test statistic for either `DCorr-X` or `MGC-X`. Let $T_{n,K,\alpha}$ be the top $100(1-\alpha)$-th percentile of the K values of T_n evaluated at $(X_t, Y_{\pi(t)})$ for each index mapping π.

Theorem 3. *Under the null, $P(T_n^* \geq T_{n,K,\alpha}) \to \alpha$ as $n \to \infty$. In other words, `DCorr-X` and `MGC-X` using block permutation are asymptotically* **valid**.

This means that using this block permutation procedure, `DCorr-X` and `MGC-X` correctly controls the Type I error for large sample size.

Theorem 4. *Under the alternative, $P(T_n^* \geq T_{n,K,\alpha}) \to 1$ as $n, K \to \infty$. In other words, `DCorr-X` and `MGC-X` with block bootstrap are* **consistent** *against any dependency.*

Similarly, Theorem 4 states that as the sample size increases, under any alternative, the probability or correctly rejecting the null approaches 1.

Finally, Theorem 5 inspects the optimal lag estimation technique, in that the returned $\hat{M}$ finds the lag $M^* = \text{argmax}_{j:0\leq j\leq M}\, d\text{corr}(j)$, at which $d$corr(j) or `MGC`(j) is highest.

Theorem 5. *$\hat{M} \to M^*$ in probability as $n \to \infty$.*

Thus, $\hat{M}$ is a consistent estimator for M^*, correctly detecting the lag of maximal dependence.

5 Simulation Results

5.1 Power Curves

The following simulations estimates the power of either test at varying sample sizes. ϵ_t represents the noise on time series $\{X_t\}$ and η_t represents the noise on time series $\{Y_t\}$, both generated as standard normal random variables.

No Dependence The process below represents two independent, stationary autoregressive time series with lag one, commonly denoted AR(1). In this case, the independence test should control the rejection probability by some desired significance level α. While the block bootstrap procedure is known to have results in terms of asymptotic validity for many problems, Figure 1A shows validity for this process at sample sizes as low as n=70 with 100 replicates.

$$\begin{bmatrix} X_t \\ Y_t \end{bmatrix} = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix} \begin{bmatrix} X_{t-1} \\ Y_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_t \\ \eta_t \end{bmatrix}$$

Linear Dependence This process presents a case of linear correlation, detectable by the CCF. Figure 1B shows `DCorr-X` always using the global scale (which is optimal here) while `MGC-X` might choose a local scale (not (n,n)) due to random variation, explaining a slight decrease in power, as expected.

$$\begin{bmatrix} X_t \\ Y_t \end{bmatrix} = \begin{bmatrix} 0 & 0.5 \\ 0.5 & 0 \end{bmatrix} \begin{bmatrix} X_{t-1} \\ Y_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_t \\ \eta_t \end{bmatrix}$$

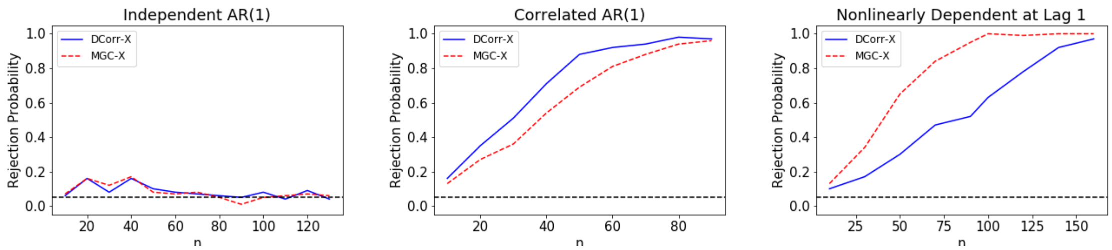

Figure 1: The three processes above are simulated for n timesteps, $R = 100$ times, with $K = 100$ bootstrap replicates, with the percent of series for which either test (`DCorr-X` or `MGC-X`) rejects the null hypothesis recorded. The black dashed line represents $\alpha = 0.05$. In column 1, the tests show empirical validity, as they accepts at rate α. In columns 2 and 3, the power converges to 1 for either test. These simulations support the claim of asymptotic validity and consistency for a larger class of time series than specified by Section 4.

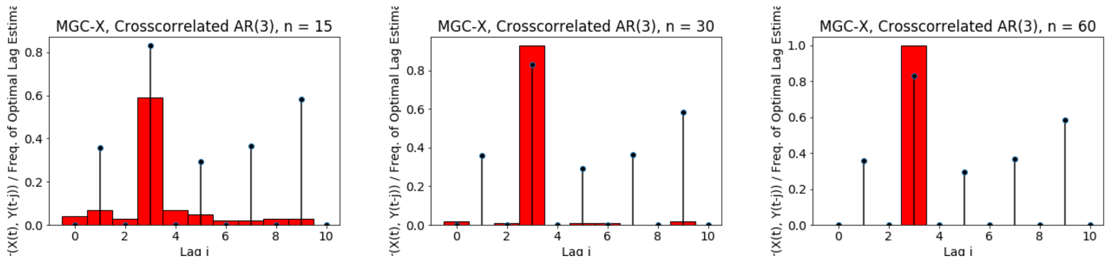

Figure 2: For the above linear time series simulations, $\phi_1 = 0.1, \phi_3 = 0.8, \epsilon_t, \eta_t \sim \mathcal{N}(0, 1)$. The stems correspond to the true cross-correlation, CCF(j), while the bars represent the empirical distribution of the optimal lag estimate for $R = 100$ trials computed with `MGC-X`.

Nonlinear Dependence The final process presents a simple case where the time series are dependent, but not correlated. Figure 1C shows `MGC-X`, which estimates a local scale to increase power for each value of n, achieves greater power than `DCorr-X` until n is large enough that `DCorr-X` is consistent as well.

$$\begin{bmatrix} X_t \\ Y_t \end{bmatrix} = \begin{bmatrix} \epsilon_t Y_{t-1} \\ \eta_t \end{bmatrix}$$

5.2 Optimal Lag Estimation

Consider the process

$$\begin{bmatrix} X_t \\ Y_t \end{bmatrix} = \begin{bmatrix} 0 & \phi_1 \\ \phi_1 & 0 \end{bmatrix} \begin{bmatrix} X_{t-1} \\ Y_{t-1} \end{bmatrix} + \begin{bmatrix} 0 & \phi_3 \\ \phi_3 & 0 \end{bmatrix} \begin{bmatrix} X_{t-3} \\ Y_{t-3} \end{bmatrix} + \begin{bmatrix} \epsilon_t \\ \eta_t \end{bmatrix}.$$

For $\phi_3 \gg \phi_1$, there exists a stronger dependence between the X_t and Y_{t-3} than X_t and Y_{t-1}. Figure 2 shows that `MGC-X` estimate the correct lag more often as n grows.

Similarly, consider the simple nonlinear process, which has clear dependence at lag 3:

$$\begin{bmatrix} X_t \\ Y_t \end{bmatrix} = \begin{bmatrix} \epsilon_t Y_{t-3} \\ \eta_t \end{bmatrix}$$

Figure 3 shows similar phenomenon in the distribution of optimal lag estimates, as center more tightly around the true optimal lag $j = 3$ for even the relatively small sample size of $n = 60$.

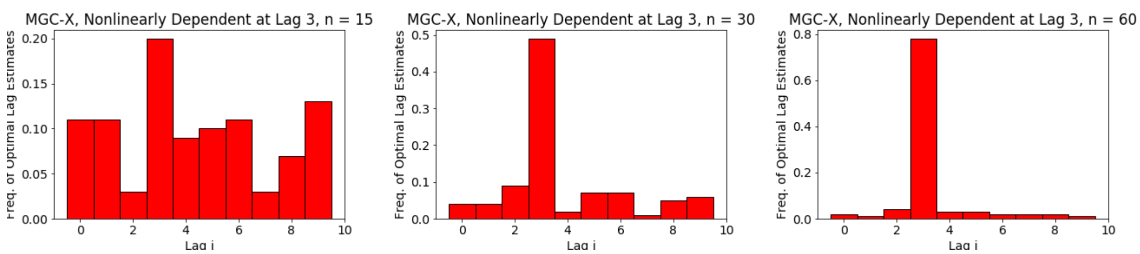

Figure 3: Similar to the correlated example, the bars represent the empirical distribution of the optimal lag estimate for $R = 100$ trials computed with `MGC-X`.

Shorthand	Networks
dAtt	Dorsal Attention Network
vAtt	Ventral Attention Network
DMN	Default Mode Network
SM	Somatomotor Network
Visual	Visual Network
Limbic	Limbic Network
FP	FrontoParietal Network

Table 1: Networks representing functional areas of the brain. Their shorthand notation is used in the following figures.

6 Analyzing Connectivity in the Human Brain

The following example is from an individual (Subject ID: 100307) of the Human Connectome Project (HCP) available for download from: https://www.humanconnectome.org/study/hcp-young-adult/data-releases. The human cortex is parcellated into 180 parcels as per hemisphere using HCP multi-modal parcellation atlas [4]. 22 parcels are selected as regions-of-interest (ROIs) representing various locations across the cortex. The parcels are denoted $X^{(1)}, ..., X^{(22)}$. Each parcel consists of contiguous set of vertices whose fMRI signal is projected on the cortical surface. Averaging those vertices within a parcel yields a univariate time series $X^{(u)} = (X_1^{(u)}, ..., X_n^{(u)})$. In this case $n = 1200$. The network names and their shorthand notation are listed in Table 1. The selected ROIs, their parcel number in the HCP multi-modal parcellation [4], and assigned network are listed in Table 2.

Figure 4 shows the negative logarithm base 10 p-value of `MGC-X` performed on each pair of parcels. Note that the test is asymmetric, resulting in an asymmetric matrix. Even with the maximum lag set to 1, there are many dependencies within and between ROIs.

Figure 5 shows the optimal lag for each interdependency, maximum lag parameter $M = 10$. Within the visual ("Visual") and default mode networks ("DMN"), signals depend on each other most strongly at the current timestep. On the other hand, between regions many signals can depend on other signals from up to 10 timesteps in the past.

Finally, Figure 6 shows the optimal scale at which the local correlation is maximized for both the first time series $X = X^{(u)}$ and the second $Y = X^{(v)}$. With many of the time series within the default mode network and visual network having a normalized optimal scale of (1,1) when tested for dependence, it would appear that the signals in these networks are linearly dependent. In many other networks, nonlinear dependencies are observed among their signals.

ROI ID	Network	Parcel Key	Parcel Name
1	Visual	1	V1
2	Visual	23	MT
3	Visual	18	FFC
4	SM	53	3a
5	SM	24	A1
6	dAtt	96	6a
7	dAtt	117	API
8	dAtt	50	MIP
9	dAtt	143	PGp
10	vAtt	109	MI
11	vAtt	148	PF
12	vAtt	60	p32pr
13	vAtt	38	23c
14	Limbic	135	TF
15	Limbic	93	OFC
16	FP	83	p9-46v
17	FP	149	PFm
18	DMN	150	PGi
19	DMN	65	p32pr
20	DMN	161	32pd
21	DMN	132	TE1a
22	DMN	71	9p

Table 2: The first column displayed the numeric order in which parcels appear in the following figures. The second displayed the network to which these parcels belong, while the last displays the parcel number from the et al. [4] parcellation.

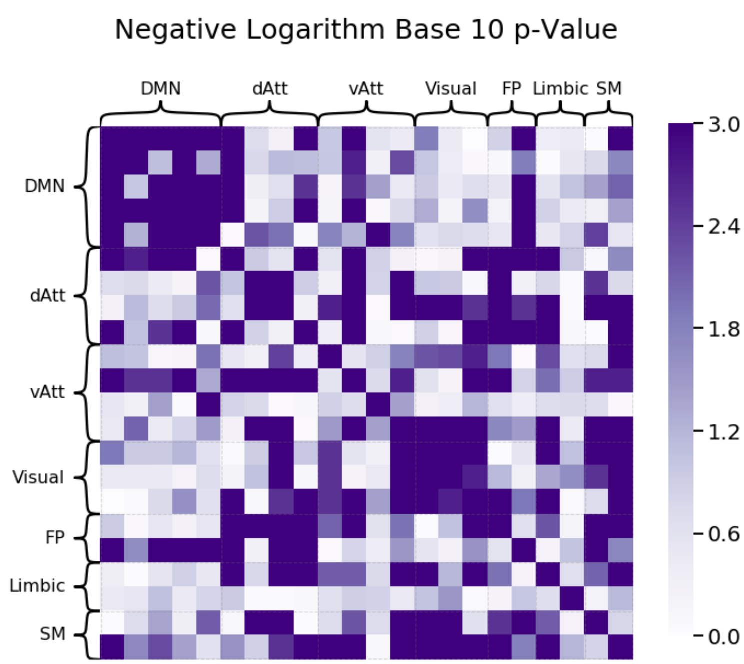

Figure 4: For the above matrix P, P_{uv} is the resulting p-value of MGC-X applied to $\{X^{(u)}\}$ and $\{X^{(v)}\})$ with maximum lag $M = 1$. Any value above approximately 1.3 is significant at $\alpha = 0.05$.

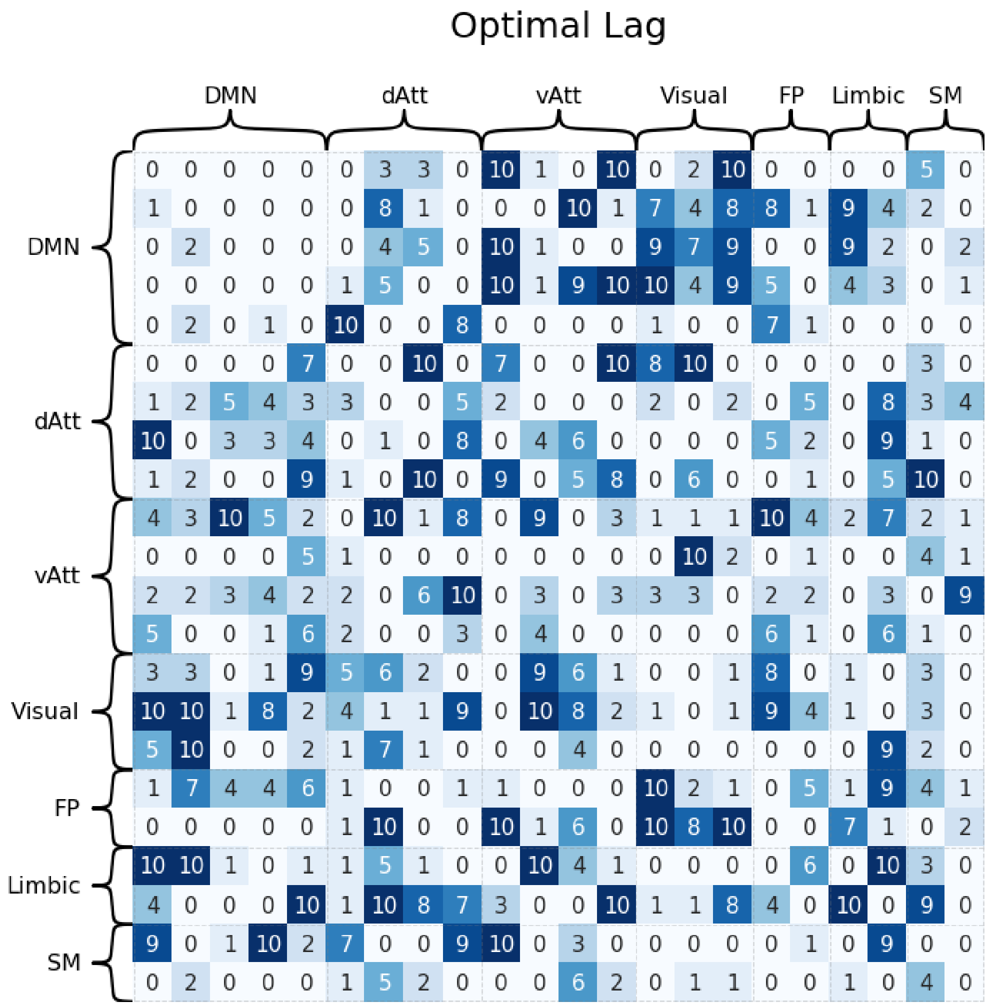

Figure 5: For the above matrix Q, Q_{uv} is the resulting optimal lag of `MGC-X` applied to $\{X^{(u)}\}$ and $\{X^{(v)}\}$ with maximum lag $M = 10$. Dark blue values represent large values of optimal lag, which imply that $\{X^{(u)}\}$ depends strongly on past values of $\{X^{(v)}\}$)

.

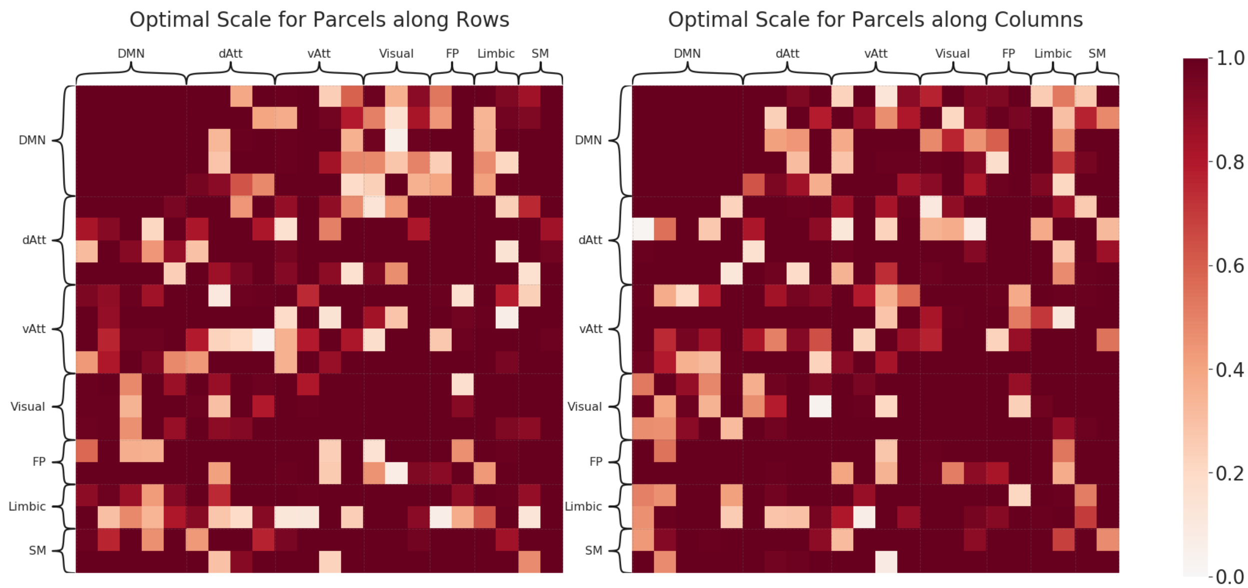

Figure 6: For the left matrix P, P_{uv} is the optimal scale of $\{(X_t^{(u)}\}$ (divided by n) when `MGC-X` is applied to $\{(X_t^{(u)}\}$ and $\{X_t^{(v)})\})$ with maximum lag $M = 1$. For the right matrix Q, Q_{uv} is the optimal scale of $\{(X_t^{(v)}\}$ (divided by n) when `MGC-X` is applied to $\{(X_t^{(u)}\}$ and $\{X_t^{(v)})\})$ with maximum lag $M = 1$. The full optimal scale is read as the pair (P_{uv}, Q_{uv}). A normalized optimal scale of (1,1) implies a linear relationship between $\{(X_t^{(u)}\}$ and $\{X_t^{(v)})\})$, as present within many of the time series of the visual network and default mode network. Deviations from (1,1) imply nonlinear relationships, which are ubiquitous within and between the other networks.

7 Discussion

The results regarding DCorr-X and MGC-X prompt further analysis both in theory and application. Theoretically, next steps involve extending validity and consistency results for a growing M, introducing consistency against all alternatives. From a methodological viewpoint, the ideas of independence testing and block permutation can be readily extended to other forms of dependent data, specifically spatial statistics and geographical data [8] [2] [20].

In light of Shen and Vogelstein [15], this independence testing procedure is directly applicable to the K-sample testing of time series problem as well. Considering the expansion of these methods, MGC-X motivates research in many applications, especially for biological time series, for which finite-sample testing power is crucial. In neuroscience specifically, the Human Connectome Project data can be used to characterize differential connectivity across individuals and mental tasks/states. These analyses can easily be extended to the population level, investigating multiple brains.

The functions to perform the tests are implemented in the pip installable Python package mgcpy [12], inviting researchers of all disciplines to use DCorr-X and MGC-X as a data analysis tool for multiple time series analysis. The theoretical, methodological, and practical impact of DCorr-X and MGC-X expand the reach of statistical principles in making scientific discovery possible.

Data and Code Availability Statement The analysis and visualization of the data was performed using open-source software packages MGC (https://neurodata.io/mgc) and GraSPy (https://neurodata.io/graspy/).

Acknowledgements The authors would like to thank Hayden Helm, Ben Pedigo, Jaewon Chung, and Bijan Varjavand for their helps and discussions in preparation of the paper. This work was partially supported by the National Science Foundation award DMS-1921310, and the Defense Advanced Research Projects Agency (DARPA) Lifelong Learning Machines program through contract FA8650-18-2-7834.

References

[1] Jonathan D. Cryer and Kung sik Chan. 14.9: Other methods of spectral estimation.âĂİ time series analysis with applications in r. *Springer*, 2011.

[2] Nordman D, S Lahiri, and B Fridley. Optimal block size for variance estimation by a spatial block bootstrap method. *SankhyÄĄ: The Indian Journal of Statistics*, 2007.

[3] Dominic Edelmann, Konstantinos Fokianos, and Maria Pitsillou. An updated literature review of distance correlation and its applications to time series. *International Statistical Review*, 2018.

[4] Glasser et al. A multi-modal parcellation of human cerebral cortex. *Nature*, 2016.

[5] K Fokianos and M Pitsillou. Consistent testing for pairwise dependence in time series. *Technometrics*, 59(2), 2017.

[6] A. Gretton and L. Gyorfi. Consistent nonparametric tests of independence. *Journal of Machine Learning Research*, 11:1391–1423, 2010.

[7] A. Gretton, R. Herbrich, A. Smola, O. Bousquet, and B. Scholkopf. Kernel methods for measuring independence. *Journal of Machine Learning Research*, 6:2075–2129, 2005.

[8] Y Guan and J Meng. A thinned block bootstrap variance estimation procedure for inhomogeneous spatial point patterns. *Journal of the American Statistical Association*, 2007. doi: 10.1198/016214507000000879.

[9] Wang Guochang, Wai Keung Li, and Ke Zhu. New hsic-based tests for independence between two stationary multivariate time series. *ArXiv E-prints*, 2018.

[10] Yongmiao Hong. Hypothesis testing in time series via the empirical characteristic function: A generalized spectral density approach. *Journal of the American Statistical Association*, 94(448): 1201, 1999. doi: 10.2307/2669935.

[11] Y. Lee, C. Shen, C. E. Priebe, and J. T. Vogelstein. Network dependence testing via diffusion maps and distance-based correlations. *Biometrika*, 2019.

[12] S. Panda, S. Palaniappan, J. Xiong, A. Swaminathan, S. Ramachandran, E. Bridgeford, C. Shen, and J. Vogelstein. Mgc: Comprehensive high dimensional independence testing python package. *https://arxiv.org/abs/1907.02088*, 2019.

[13] Dimitris N Politis. The impact of bootstrap methods on time series analysis. *Statist. Sci.*, 18(2): 219–230, 2003.

[14] D. Sejdinovic, B. Sriperumbudur, A. Gretton, and K. Fukumizu. Equivalence of distance-based and rkhs-based statistics in hypothesis testing. *Annals of Statistics*, 41(5):2263–2291, 2013.

[15] C. Shen and J. T. Vogelstein. The exact equivalent of distance and kernel methods for hypothesis testing. *https://arxiv.org/abs/1806.05514*, 2019.

[16] C. Shen, C. E. Priebe, and J. T. Vogelstein. From distance correlation to multiscale graph correlation. *Journal of the American Statistical Association*, 2019.

[17] G. Szekely and M. Rizzo. Brownian distance covariance. *Annals of Applied Statistics*, 3(4):1233–1303, 2009.

[18] G. Szekely, M. Rizzo, and N. Bakirov. Measuring and testing independence by correlation of distances. *Annals of Statistics*, 35(6):2769–2794, 2007.

[19] J. T. Vogelstein, Q. Wang, E. Bridgeford, C. E. Priebe, M. Maggioni, and C. Shen. Discovering and deciphering relationships across disparate data modalities. *eLife*, 8:e41690, 2019.

[20] J Zhu and G Morgan. Comparison of spatial variables over subregions using a block bootstrap. *Journal of Agricultural, Biological, and Environmental Statistics*, 2003. doi: https://doi.org/10.1198/1085711043154.

Appendix

The assumptions and results are listed once again below.

Assumption 1. $\left\{\begin{bmatrix} X_t \\ Y_t \end{bmatrix}\right\}_t$ *is a strictly stationary process.*

Assumption 2. $\left\{\begin{bmatrix} X_t \\ Y_t \end{bmatrix}\right\}_t$ *has finite second moment.*

Assumption 3. *For the block permutation test procedure, we choose a block size b_n and number of permutations K such that as $n \to \infty$,*

$$b_n \to \infty,$$
$$\frac{b_n}{n} \to 0,$$
$$K \to \infty$$

Assumption 4. *The following hold.*

$$\sup |F_{X_t, X_{t-j}} - F_{X_t} F_{X_{t-j}}| \to 0 \text{ as } j \to \infty$$
$$\sup |F_{Y_t, Y_{t-j}} - F_{Y_t} F_{Y_{t-j}}| \to 0 \text{ as } j \to \infty$$
$$\sup |F_{X_t, Y_{t-j}} - F_{X_t} F_{Y_{t-j}}| \to 0 \text{ as } j \to \infty$$

Assumption 5. $\left\{ \begin{bmatrix} X_t \\ Y_t \end{bmatrix} \right\}_t$ *is nonconstant, and a continuous random variable for all t.*

Theorem 1. *For any $\epsilon > 0$, there exists a sufficiently large n such that*

$$\mathbb{E}\left[d\text{cov}_n^{k,l}\right] = d\text{cov}^{\rho_k,\rho_l} + \mathcal{O}\left(\frac{1}{n}\right) + \mathcal{O}(\epsilon)$$
$$\text{Var}\left[d\text{cov}_n^{k,l}\right] = \mathcal{O}\left(\frac{1}{n}\right) + \mathcal{O}(\epsilon)$$

Proof. In the case that X and Y represent i.i.d. random variables, as in Shen et al. [16], then

$$d\text{cov}^{\rho_k,\rho_l}(X, Y) = \mathbb{E}\left[A_{ij} B_{ji} \boldsymbol{I}(R_{ij}^A \leq k) \boldsymbol{I}(R_{ji}^B \leq l)\right] + \mathcal{O}\left(\frac{1}{n}\right) \tag{1}$$

for any $i \neq j$. For time series, we defined

$$d\text{cov}^{\rho_k,\rho_l}(X_t, Y_t) = \int_{\mathbb{R}^p \times \mathbb{R}^q} \left\{ \mathbb{E}[h_{X_t}^{\rho_k}(u) \overline{h_{Y_t'}^{\rho_l}(v)}] - \mathbb{E}[h_{X_t}^{\rho_k}(u)] \mathbb{E}[\overline{h_{Y_t'}^{\rho_l}(v)}] \right\} w(u,v) du dv$$

for independent copies of the series (X_t, Y_t) and (X_t', Y_t'). For a finite time series sample, observations X_i and X_j may not be independent, and 1 may not hold. However, due to the continuity of $\mathbb{E}\left[A_{ij} B_{ji} \boldsymbol{I}(R_{ij}^A \leq k) \boldsymbol{I}(R_{ji}^B \leq l)\right]$ over the set of joint distributions (X_i, Y_i) and (X_j, Y_j) and Assumption 4, we have that there is a $\gamma \in \mathbb{N}$ (dependent on ϵ) such that

$$|\mathbb{E}\left[A_{ij} B_{ji} \boldsymbol{I}(R_{ij}^A \leq k) \boldsymbol{I}(R_{ji}^B \leq l)\right] - d\text{cov}^{\rho_k,\rho_l}| < \epsilon$$

for $|i - j| > \gamma$. Thus, when computing

$$\mathbb{E}\left[d\text{cov}_n^{k,l}\right] = \mathbb{E}\left[\frac{1}{n(n-1)} \sum_{i,j}^n A_{ij} B_{ji} \boldsymbol{I}(R_{ij}^A \leq k) \boldsymbol{I}(R_{ji}^B \leq l)\right]$$
$$= \frac{1}{n(n-1)} \sum_{i,j}^n \mathbb{E}\left[A_{ij} B_{ji} \boldsymbol{I}(R_{ij}^A \leq k) \boldsymbol{I}(R_{ji}^B \leq l)\right]$$

for index pairs (i,j) such that $|i-j| > \gamma$, we can apply Theorem 5 of Shen et al. [16] to simplify the expectation to $d\text{corr}^{\rho_k,\rho_l}$ and be off by at most $\mathcal{O}(\epsilon)$. We count the fraction of index pairs for which $|i-j| \leq \gamma$.

$$\frac{n + 2\sum_{i=1}^{\gamma}(n-i)}{n(n-1)} = \frac{(1+2\gamma)n + \gamma(1+\gamma)}{n(n-1)} = \mathcal{O}\left(\frac{1}{n}\right)$$

Then, the fraction of index pairs for which $|i-j| > \gamma$ is equal to:

$$\frac{n^2 - (1+2\gamma)n + \gamma(1+\gamma)}{n^2}$$

For these index pairs, we can apply the argument from Theorem 5 of Shen et al. [16], and we have that for sufficiently large n:

$$\begin{aligned}\mathbb{E}\left[d\text{cov}_n^{k,l}\right] &= \frac{n^2 - (1+2\gamma)n + \gamma(1+\gamma)}{n^2} \cdot \left[d\text{cov}^{\rho_k,\rho_l} + \mathcal{O}(\epsilon) + \mathcal{O}\left(\frac{1}{n}\right)\right] + \mathcal{O}\left(\frac{1}{n}\right) \\ &= d\text{cov}^{\rho_k,\rho_l} + \mathcal{O}(\epsilon) + \mathcal{O}\left(\frac{1}{n}\right)\end{aligned}$$

We argue very similarly for the variance.

$$\begin{aligned}\text{Var}\left[d\text{cov}_n^{k,l}\right] &= \text{Var}\left[\frac{1}{n(n-1)}\sum_{i,j}^{n} A_{ij}B_{ji}\boldsymbol{I}(R_{ij}^A \leq k)\boldsymbol{I}(R_{ji}^B \leq l)\right] \\ &= \frac{1}{n^2(n-1)^2}\text{Var}\left[\sum_{i,j}^{n} A_{ij}B_{ji}\boldsymbol{I}(R_{ij}^A \leq k)\boldsymbol{I}(R_{ji}^B \leq l)\right]\end{aligned}$$

$$= \frac{1}{n^2(n-1)^2}\text{Cov}\left[\sum_{i,j}^{n} A_{ij}B_{ji}\boldsymbol{I}(R_{ij}^A \leq k)\boldsymbol{I}(R_{ji}^B \leq l), \sum_{s,t}^{n} A_{st}B_{ts}\boldsymbol{I}(R_{st}^A \leq k)\boldsymbol{I}(R_{ts}^B \leq l)\right] \tag{2}$$

Given an ϵ, invoking the continuity of 2 with respect to the joint distribution of (X_i, Y_i), (X_j, Y_j), (X_s, Y_s), and (X_t, Y_t), we are concerned with sets of indices i, j, s, t being beyond $\gamma(\epsilon)$ distance of one another, as as these are the indices for which the components can be considered approximately independent. To such sets of indices, we have n choices for the first index, at most $n - 2\gamma - 1$ choices for the second, and so on. Thus, the fraction of these "approximately independent" indices is:

$$\frac{n(n-2\gamma-1)(n-4\gamma-2)(n-6\gamma-3)}{n^2(n-1)^2}$$

Similarly, the number of elements for which there is dependence between time series components in elements of either time series is at most:

$$\frac{n^4 - n(n-2\gamma-1)(n-4\gamma-2)(n-6\gamma-3)}{n^4} = \mathcal{O}\left(\frac{1}{n}\right)$$

In the i.i.d. case, Shen et al. [16] found $\text{Var}\left[d\text{cov}_n^{k,l}\right] = \mathcal{O}\left(\frac{1}{n}\right)$. Thus, we can use this for the approximately independent elements, being off by $\mathcal{O}(\epsilon)$ and we have:

$$\begin{aligned}\text{Var}\left[d\text{cov}_n^{k,l}\right] &= \frac{n(n-2\gamma-1)(n-4\gamma-2)(n-6\gamma-3)}{n^2(n-1)^2} \cdot \left[\mathcal{O}\left(\frac{1}{n}\right) + \mathcal{O}(\epsilon)\right] + \mathcal{O}\left(\frac{1}{n}\right) \\ &= \mathcal{O}\left(\frac{1}{n}\right) + \mathcal{O}(\epsilon)\end{aligned}$$

$\square$

Theorem 2. *For each $j \in \{0, ..., M\}$, $d\text{cov}_n^{k,l}(j) \overset{n\to\infty}{\to} d\text{cov}^{\rho_k,\rho_l}(j)$ in probability.*

Proof. The claim follows from Theorem 1 and Chebyshev's inequality. Take any $a > 0$ and any j.

$$
\begin{aligned}
P\left[|d\text{cov}_n^{k,l} - d\text{cov}^{\rho_k,\rho_l}| > a\right] &= P\left[|d\text{cov}_n^{k,l} - \mathbb{E}[d\text{cov}_n^{k,l}] + \mathcal{O}\left(\frac{1}{n}\right) + \mathcal{O}(\epsilon)| > a\right] \\
&\leq P\left[|d\text{cov}_n^{k,l} - \mathbb{E}[d\text{cov}_n^{k,l}]| + |\mathcal{O}\left(\frac{1}{n}\right)| + |\mathcal{O}(\epsilon)| > a\right] \\
&\leq P\left[|d\text{cov}_n^{k,l} - \mathbb{E}[d\text{cov}_n^{k,l}]| > a - |\mathcal{O}\left(\frac{1}{n}\right)| - |\mathcal{O}(\epsilon)|\right] \\
&\leq \frac{\text{Var}\left[d\text{cov}_n^{k,l}\right]}{(a - |\mathcal{O}\left(\frac{1}{n}\right)| - |\mathcal{O}(\epsilon)|)^2}
\end{aligned}
$$

For a fixed a, this probability approaches 0 as $n \to \infty$. □

To inspect validity, we first show that the block permuted $Y_{\pi(t)}$ is distributed the same as $\{Y_t\}$ asymptotically for every t, conditioned on previous observations. We then show that X_t is asymptotically independent of $Y_{\pi(t-j)}$ for all $|j| \leq M$. With these two facts in place, the approximately independently sampled processes, (X_t) and $(Y_{\pi(t)})$, are used to evaluate observations of the test statistic under the null. From these observations, a valid test is produced.

Theorem 3. *Under the null, $P(T_n^* \geq T_{n,K,\alpha}) \to \alpha$ as $n \to \infty$. In other words, `DCorr-X` and `MGC-X` using block permutation are asymptotically* **valid***.*

Proof. We first show that the distribution of $Y_{\pi(t)|\pi(t-1),\pi(t-2),...}$ is asymptotically the same as that of $Y_t \mid Y_{t-1}, Y_{t-1}, ...$ For any index t, this will be true if $Y_{\pi(t)}$ "close" to the components on which it is dependent. Specifically, given any ϵ consider γ such that:

$$\sup |F_{Y_t|Y_{t-1},Y_{t-2},...,Y_{t-j}} - F_{Y_t|Y_{t-1},Y_{t-2},...}| < \epsilon \text{ for } j > \gamma$$

where $Y_{t-1}, Y_{t-2}, ...$ represents the infinite sequence of past values. Such a γ exists due to Assumption 4. Let $b_n > \gamma$. The condition above is only violated when t is within the first γ indices of a block. The fraction of such indices is $\frac{\gamma}{b_n}$, which approaches 0 as n gets large. Because ϵ can be taken arbitrarily small, $F_{Y_{\pi(t)|\pi(t-1),\pi(t-2),...}}$ approaches $F_{Y_t|Y_{t-1},Y_{t-1},...}$ for large n with high probability for any t.

Next, we show that for each t, X_t and $Y_{\pi(t)}$ are asymptotically sampled independently. By assumption, this requires t and $\pi(t)$ to be sufficiently far. Given any ϵ, choose γ such that:

$$\sup |F_{X_t,Y_{t-j}} - F_{X_t}F_{Y_{t-j}}| < \epsilon \text{ for } j \geq \gamma$$

For asymptotic independence, $P(|t - \pi(t-j)| < \gamma \text{ for some } |j| < M)$ must approach 0. Let $L = \max\{\gamma, M\}$, and n be large enough such that $b_n > L$. If index t is in the first L indices of any block, then it vanishes due to the same argument as above. If t is in the remaining $b_n - L$ indices, then the permuted block that contains $\pi(t)$ must have $\pi(t) = t$ (the block containing the original indices surrounding t). This is one block among $\frac{n}{b_n}$ choices.

$$P(|t - \pi(t-j)| \leq \gamma \text{ for some } |j| \leq M) \leq \frac{L}{b_n} \cdot 1 + \frac{b_n - L}{b_n} \cdot \frac{b_n}{n} \to 0 \text{ as } n \to \infty$$

by Assumption 3. The resamples $(Y_{\pi(t)})$ are asymptotically independent from the series (X_t), as the indices t are far from any of the block permuted indices $\pi(t)$ with high probability. The statistic $T_n(X_t, Y_{\pi(t)})$ indeed follows its distribution under the null for growing n.

Finally, by the Glivenko-Cantelli theorem, the empirical cumulative distribution function (CDF) $\hat{F}_{K,T_n}$ taken with K observations of the statistic T_n, converges to the true CDF F_{T_n} in the sense that

$$\sup_{s \in \mathbb{R}} |\hat{F}_{K,T_n}(s) - F_{T_n}(s)| \to 0 \text{ almost surely as } K \to \infty.$$

Therefore, by letting the number K of resampled series $(Y_{\pi(t)})$ grow with n, say $K = n$, the estimate of the distribution of T_n is consistent, implying that $P(T_n^* \geq T_{n,K,\alpha}) \to \alpha$. `DCorr-X` and `MGC-X` with block permutation are asymptotically **valid** tests. □

Theorem 4. *Under the alternative, $P(T_n^* \geq T_{n,K,\alpha}) \to 1$ as $n, K \to \infty$. In other words, `DCorr-X` and `MGC-X` with block bootstrap are* **consistent** *against any dependency.*

Proof. Due to Theorem 2, we have the convergence of $d\text{cov}_n^{k,l}(j) \to d\text{cov}^{\rho_k,\rho_l}(j)$ as $n \to \infty$. Because the test statistic is a finite sum from $j = 0$ to M, the test statistic will converge in probability to its population counterpart - either $\sum_{j=0}^{M} d\text{cov}(j)$ for `DCorr-X` or $\sum_{j=0}^{M} \max_{\rho_k,\rho_l} d\text{cov}^{\rho_k,\rho_l}(j)$ for `MGC-X`. This value will be nonzero under the alternative [16]. Similarly, due to the argument of Theorem 3, $T_{n,K,\alpha} \to 0$ in probability as $n \to \infty$. Thus, under the alternative, $P(T_n^* \geq T_{n,K,\alpha}) \to 1$ as $n \to \infty$. □

Theorem 5. *$\hat{M} \to M^*$ in probability as $n \to \infty$.*

Proof. For each $j : 0 \leq j \leq M$, $d\text{corr}_n^{k,l}(j) \to d\text{corr}^{\rho_k,\rho_l}(j)$ in probability. Thus, from the finiteness of the search space for the maximum:

$$\hat{M} = \underset{j}{\text{argmax}} \left(\frac{n-j}{n}\right) \cdot d\text{corr}_n(j) \to \underset{j}{\text{argmax}}\, d\text{corr}(j) = M^*$$
$$\hat{M} = \underset{j}{\text{argmax}} \left(\frac{n-j}{n}\right) \cdot \text{mgc}_n(j) \to \underset{j}{\text{argmax}} \max_{\rho_k,\rho_l} d\text{corr}^{\rho_k,\rho_l}(j) = M^*$$

for `DCorr-X` and `MGC-X` respectively. □