A Wavelet-Based Independence Test for Functional Data With an Application to MEG Functional Connectivity

Rui Miao, Xiaoke Zhang, and Raymond K. W. Wong

ABSTRACT

Measuring and testing the dependency between multiple random functions is often an important task in functional data analysis. In the literature, a model-based method relies on a model which is subject to the risk of model misspecification, while a model-free method only provides a correlation measure which is inadequate to test independence. In this paper, we adopt the Hilbert–Schmidt Independence Criterion (HSIC) to measure the dependency between two random functions. We develop a two-step procedure by first pre-smoothing each function based on its discrete and noisy measurements and then applying the HSIC to recovered functions. To ensure the compatibility between the two steps such that the effect of the pre-smoothing error on the subsequent HSIC is asymptotically negligible when the data are densely measured, we propose a new wavelet thresholding method for pre-smoothing and to use Besov-norm-induced kernels for HSIC. We also provide the corresponding asymptotic analysis. The superior numerical performance of the proposed method over existing ones is demonstrated in a simulation study. Moreover, in a magnetoencephalography (MEG) data application, the functional connectivity patterns identified by the proposed method are more anatomically interpretable than those by existing methods.

1. Introduction

In the recent decades, functional data analysis (FDA) has developed rapidly due to a huge and increasing number of datasets collected in the form of curves, surfaces and volumes. General introductions to the subject may be found in a few monographs (e.g., Ramsay and Silverman 2005; Ferraty and Vieu 2006). In many scientific fields, measurements are taken from multiple random functions per subject and the dependency between these functions is of interest. For instance, neuroscientists are interested in functional connectivity patterns between signals at multiple brain regions, which are measured over time in functional magnetic resonance imaging data. It is thus an important task in FDA to measure their dependency and to further test the significance of the dependency. Among extensive relevant research endeavors, most dependency test methods can be categorized as either model-based or model-free.

A model-based method typically infers the dependency between multiple functions by first assuming a functional regression model (see, e.g., Morris 2015, for a survey) which characterizes their structural relationship, and then testing the significance of the assumed model. See examples of model-based methods by Guo (2002), Huang, Wu, and Zhou (2002), Shen and Faraway (2004), and Antoniadis and Sapatinas (2007) for concurrent/varying-coefficient models and by Kokoszka et al. (2008) and Chen et al. (2020) for function-on-function regression models. The main disadvantage of a model-based method is its reliance on correct model specification. If the model is misspecified, the inference is not well grounded and might be inaccurate.

A model-free method can avoid the misspecification issue associated with model-based methods since it typically quantifies the dependency between random functions by a correlation measure, without assuming any particular model. As a natural extension of the canonical correlation for multivariate data, the functional canonical correlation is a popular correlation measure for functional data (e.g., Leurgans, Moyeed, and Silverman 1993; He, Müller, and Wang 2003; Eubank and Hsing 2008; Shin and Lee 2015). However, it is plagued by the involvement of inverting a covariance operator, which is an ill-posed problem and often requires proper regularizations. The dynamical correlation (Dubin and Müller 2005; Sang, Wang, and Cao 2019) and temporal correlation (Zhou, Lin, and Wang 2018) are two functional correlation measures without the aforementioned inverse problem. The former measures the angle between two random functions in the $L^2$ space. The latter essentially computes the Pearson correlation between two random functions at each time point and then averages all pointwise Pearson correlations over the time domain. However, since uncorrelatedness does not imply independence, these functional correlations are insufficient to test independence. Recently a few model-free approaches have been developed to test mean independence for functional data (e.g., Patilea, Sánchez-Sellero, and Saumard 2016; Lee, Zhang, and Shao 2020), but they can only test a weaker notion of independence.

In this article we develop a model-free independence test for functional data. Under the reproducing kernel Hilbert space (RKHS) framework, we propose to use the Hilbert–Schmidt Independence Criterion (HSIC; e.g., Gretton et al. 2005, 2008)
to measure the dependency between two random functions. An appealing property is that HSIC endowed with characteristic kernels is zero if and only if the two random functions are independent. However, the application of HSIC requires fully observed and noiseless functional data, while in practice functional data are always discretely measured and contaminated by noise. To tackle this problem, one may perform a two-step procedure: first pre-smooth the data, and then apply HSIC to the resulting functions. Clearly, pre-smoothing will affect the performance of HSIC. Indeed, the functional distance with respect to which the asymptotic convergence of the pre-smoothing procedure is measured is crucial, as HSIC is fundamentally based on a functional distance. Some common pre-smoothing procedures do not have existing convergence results on the required functional distance, and hence may not be compatible; namely, the pre-smoothing error may have a profound effect on the subsequent HSIC. See Section 3 for more discussion.

In this work, we carefully design our procedure to ensure that the two steps are compatible. For the first step, we propose a new wavelet thresholding method; while we use Besov-norm-induced kernels for HSIC in the second step. We can show that these choices in the two steps are theoretically compatible if the functional data are sufficiently densely measured. See Section 4 for details. Our work is motivated by the Human Connectome Project (HCP, https://www.humanconnectome.org) from which various brain imaging datasets are publicly accessible. In Section 7, the application of our method to a magnetoencephalography (MEG) dataset from HCP is capable of identifying anatomically interpretable functional connectivity patterns, suggesting a great potential of the proposed method in the study of functional connectivity between brain regions.

The main contribution of this article is three-fold. First, we design some suitable kernels such that the corresponding HSIC can identify the independence of a pair of random functions of which sample paths belong to Besov spaces, a larger class of functions than Sobolev spaces which are popular in RKHS modeling. We propose to use the Besov sequence norm for the wavelet coefficients of these random functions to induce such kernel, which is shown to be characteristic. Second, for dense functional data, we develop the asymptotic distribution of the empirical HSIC based on pre-smoothed functions by wavelet thresholding. To theoretically guarantee the compatibility between the pre-smoothing and empirical HSIC, we propose a new wavelet thresholding method that can efficiently reduce the pre-smoothing error measured by the Besov sequence norm used in the empirical HSIC when the noise is nearly independent. Since the asymptotic distribution involves many unknown quantities, we suggest a permutation test in practice and prove that not only can the test control the Type I error probability but also it is consistent. The theoretical results show that the two steps in our proposed procedure are compatible. Finally, we propose a data-adaptive approach to tuning the smoothness parameter by the Besov norm needed to induce the kernel for HSIC. It is numerically shown that this approach is able to enhance the sensitivity of HSIC to detecting dependencies at high frequencies.

The rest of the article proceeds as follows. Section 2 provides a brief introduction to HSIC. The two-step procedure for the proposed wavelet-based HSIC test is given in Section 3. Its asymptotic properties are presented in Section 4. Section 5 discusses tuning parameter selection. The numerical performance of the proposed method is illustrated in a simulation study in Section 6 and an MEG functional connectivity study in Section 7 where it is also compared with representative existing methods. Section 8 concludes the article. The code to implement the proposed method is publicly available on GitHub (https://github.com/rai-miao/wavHSIC).

2. Hilbert–Schmidt Independence Criterion

In this section, we give a brief introduction to HSIC. Let \( X \) and \( Y \) be two random functions of which sample paths belong to function spaces \( \mathcal{X} \) and \( \mathcal{Y} \), respectively, and \( \mathcal{H}(\kappa_{\mathcal{X}}) \) and \( \mathcal{H}(\kappa_{\mathcal{Y}}) \) be the RKHS equipped with kernels \( \kappa_{\mathcal{X}} \) and \( \kappa_{\mathcal{Y}} \) defined on \( \mathcal{X} \times \mathcal{X} \) and \( \mathcal{Y} \times \mathcal{Y} \) respectively.

HSIC requires that both \( \kappa_{\mathcal{X}} \) and \( \kappa_{\mathcal{Y}} \) are characteristic, in the sense that two probability measures \( P = P' \) if and only if \( \mathbf{P}^x(P) = \mathbf{P}^x(P') \) where \( \mathbf{P}^x(P) = EP_x(\kappa_Z(Z, \cdot)) \) for a random function \( Z \in \mathcal{Z} \) which follows \( P \) and \( (Z, Z) = (X, X) \) or \( (Y, Y) \) respectively. A characteristic kernel may be induced by a strong negative-type semi-metric (see Definition S1 and Proposition S1 in the supplementary material). Denote the joint probability measure of \( X \) and \( Y \) by \( P_{X\mathcal{Y}} \) and their marginal probability measures by \( P_X \) and \( P_Y \), respectively. Since \( \kappa_{\mathcal{X}} \) and \( \kappa_{\mathcal{Y}} \) are characteristic, \( P_X \) and \( P_Y \) are fully characterized by \( \mathbf{P}^x(P_X) = EP_x(\kappa_X(x, \cdot)) \) and \( \mathbf{P}^y(P_Y) = EP_y(\kappa_Y(y, \cdot)) \) respectively. Let \( \mathbf{P}^{x\otimes y}(P_{X\mathcal{Y}}) = EP_{X\mathcal{Y}}((\kappa_x \otimes \kappa_y)((x, y), (x', y')) = \kappa_x(x, x') \kappa_y(y, y') \), where the tensor product kernel \( \kappa_x \otimes \kappa_y \) is defined by \( (\kappa_x \otimes \kappa_y)((x, y), (x', y')) \).

Sejdinovic et al. (2013) showed that \( X \) and \( Y \) are independent, that is, \( P_{X\mathcal{Y}} = P_X P_Y \), if and only if \( \mathbf{P}^{x\otimes y}(P_{X\mathcal{Y}}) = \mathbf{P}^x(P_X) \mathbf{P}^y(P_Y) \), although \( \kappa_{\mathcal{X}} \otimes \kappa_{\mathcal{Y}} \) is not characteristic for all probability measures on \( \mathcal{H}(\kappa_{\mathcal{X}}) \times \mathcal{H}(\kappa_{\mathcal{Y}}) \). Therefore, to test the independence between \( X \) and \( Y \), it suffices to study the difference between \( \mathbf{P}^{x\otimes y}(P_{X\mathcal{Y}}) \) and \( \mathbf{P}^x(P_X) \mathbf{P}^y(P_Y) \). Since \( \mathbf{P}^{x}(P_X) \in \mathcal{H}(\kappa_{\mathcal{X}}), \mathbf{P}^{y}(P_Y) \in \mathcal{H}(\kappa_{\mathcal{Y}}) \) and \( \mathbf{P}^{x\otimes y}(P_{X\mathcal{Y}}) \in \mathcal{H}(\kappa_{X} \otimes \kappa_{Y}) \) where \( \mathcal{H}(\kappa_{\mathcal{X}} \otimes \kappa_{\mathcal{Y}}) \) is the RKHS equipped with \( \kappa_{X} \otimes \kappa_{Y} \), HSIC may be used to measure this difference under the norm of \( \mathcal{H}(\kappa_{X} \otimes \kappa_{Y}) \).

Definition (HSIC). Suppose that \( \int_X \kappa_X(x, x) dP_X(x) < \infty \) and \( \int_Y \kappa_Y(y, y) dP_Y(y) < \infty \). The HSIC of \( P_{X\mathcal{Y}} \) is defined by

\[
\gamma(P_{X\mathcal{Y}}, \kappa_{\mathcal{X}}, \kappa_{\mathcal{Y}}) = ||\mathbf{P}^{x\otimes y}(P_{X\mathcal{Y}}) - \mathbf{P}^x(P_X) \mathbf{P}^y(P_Y)||_{\mathcal{H}(\kappa_{\mathcal{X}} \otimes \kappa_{\mathcal{Y}})}
= 4 \int_{X \times Y} \int_{X \times Y} \kappa_x(x, x') \kappa_y(y, y') dP_{X\mathcal{Y}} - P_X P_Y(x, x', y, y') \times \text{d}(P_{X\mathcal{Y}} - P_X P_Y(x, y) dP_{X\mathcal{Y}} - P_X P_Y(x', y')).
\]

In practice with \( \{(X_i, Y_i) : i = 1, \ldots, n\} \) which are independently and identically distributed (iid) copies of \( (X, Y) \), the sample versions of \( \mathbf{P}^{x}(P_n) \), \( \mathbf{P}^y(P_n) \) and \( \mathbf{P}^{x \otimes y}(P_n) \) are defined by \( \mathbf{P}^x(P_n) = \frac{1}{n} \sum_{i=1}^{n} \kappa_X(X_i, \cdot) \), \( \mathbf{P}^y(P_n) = \frac{1}{n} \sum_{i=1}^{n} \kappa_Y(Y_i, \cdot) \), and \( \mathbf{P}^{x \otimes y}(P_n) = \frac{1}{n} \sum_{i=1}^{n} \{(\kappa_{X_i} \otimes \kappa_{Y_i})((X_i, Y_i), (\cdot, \cdot))\} \). Obviously \( \mathbf{P}^{x}(P_n) \in \mathcal{H}(\kappa_{X}), \mathbf{P}^y(P_n) \in \mathcal{H}(\kappa_{Y}) \) and \( \mathbf{P}^{x \otimes y}(P_n) \in \mathcal{H}(\kappa_{X} \otimes \kappa_{Y}) \), so we can obtain a sample version of HSIC as follows.
Definition 2 (Empirical HSIC). Under the same setting in Definition 1, the empirical HSIC, which is an estimator of HSIC, is defined by

\[\gamma(P_{XY}, \kappa_X, \kappa_Y) := \left| \mathbf{P}_{X \otimes Y}^{\mathbf{P}_{XY}}(P_{XY}) \right|^2 \mathbb{H}(k_X \otimes k_Y) \]

\[= 4 \int_{X \times Y} \int_{X \times Y} \kappa_X(x, x') \kappa_Y(y, y') d(P_{XY} - P_{X}P_{Y})(x, y) \times d(P_{XY} - P_{X}P_{Y})(x', y').\]

By Sejdinovic et al. (2013), the empirical HSIC can be rewritten as follows:

\[\gamma(P_{XY}, \kappa_X, \kappa_Y) = n^{-2} \text{tr}(\mathbf{R}^{-X} \mathbf{R}^{-Y} \mathbf{H}),\]

where \(\mathbf{R} = (\kappa_X(X_i, X_j))_{i,j=1}^{n} \) and \(\mathbf{R}' = (\kappa_Y(Y_i, Y_j))_{i,j=1}^{n} \) are Gram matrices, and \(\mathbf{H} = \mathbf{I}_n - n^{-1} \mathbf{I}_n \mathbf{I}_n^\top \) is the centering matrix with the \(n \times n\) identity matrix \(\mathbf{I}_n\) and \(\mathbf{I}_n = (1, \ldots, 1)^\top\) of dimension \(n\).

3. Methodology

Suppose that bivariate functional data \((X_i, Y_i) : i = 1, \ldots, n\) collected from \(n\) subjects are iid copies of a pair of random functions \((X, Y), \) which, without loss of generality, is defined on the domain \([0, 1] \times [0, 1]\). Let the sample paths of \(X\) and \(Y\) belong to function spaces \(\mathcal{X}\) and \(\mathcal{Y}\), respectively. In many applications such as brain imaging analysis, the measurements of each function are sampled at a discrete and regular grid and subject to noise contamination. Hence, we assume that the observations are \((\hat{X}_i, \hat{Y}_i) := (X_i(T_l) + \epsilon_i^X(T_l), Y_i(T_l) + \epsilon_i^Y(T_l)) : i = 1, \ldots, n; l = 1, \ldots, m\), where \(\{T_l = (l-1)/m : l = 1, \ldots, m\}\) is a regular grid with \(m = 2^{l+1}\) for some integer \(l > 0\) and the two sets of mean-zero random noise, \((\epsilon_i^X)_{i=1}^n = (\epsilon_1^X, \ldots, \epsilon_m^X)\) and \((\epsilon_i^Y)_{i=1}^n = (\epsilon_1^Y, \ldots, \epsilon_m^Y)\), are independent of each other and of \((X_i, Y_i) : i = 1, \ldots, n\). The error terms in each set are further assumed to be identically distributed, independent across subjects, but possibly dependent within each subject. We defer the discussion on the error dependence structures to Section 4. Our goal is to formulate an HSIC-based test for the independence between \(X\) and \(Y\) via \((\hat{X}_i, \hat{Y}_i) : i = 1, \ldots, n; l = 1, \ldots, m\). For simplicity we assume that all functions share the same measurement grid and \(m = 2^{l+1}\), but the proposed method is applicable with minor modifications if the grid is irregular, the functions are measured at different grids, or \(m \neq 2^{l+1}\) (see Remark 1).

Due to the success of existing HSIC-based independence tests for multivariate data, it is tempting to treat the discretized observations as multivariate data and directly apply existing methods. However, there are two issues with this approach. First, in order to capture enough information, \(m\) should be large enough, which naturally leads to high-dimensional data. Without reasonable structure across these \(m\) dimensions, HSIC does not perform well. In the FDA literature, modeling the sample paths with certain form of smoothness has been shown an empirically successful strategy in many applications. It is beneficial to incorporate smoothness structure during the design of a tailor-made HSIC method. Second, the discretized observations are contaminated by noise. Hence, these raw observations are indeed not “smooth” but the noiseless ones are.

The proposed method is directly based on the definition of HSIC (Definition 1) when applied to random functions. Clearly, the application of such HSIC requires the trajectories of all random functions to be fully observed and noiseless. Thus, with discrete and noisy measurements in practice, a natural idea is to perform pre-smoothing to recover these trajectories followed by applying HSIC to random functions. However, the compatibility of these two steps is generally unclear. Namely, it is nontrivial to know whether the pre-smoothing error (measured in a certain norm) would have a profound effect on the subsequent HSIC-based test. For instance, if the sample paths of all random functions are assumed to belong to a Sobolev space, it is seemingly reasonable to pre-smooth each trajectory by a smoothing spline followed by the HSIC based on Sobolev-norm-induced kernels. However, the compatibility of the two steps is unknown since there is no theoretical result to guarantee that the pre-smoothing error under a Sobolev norm converges to zero, although the corresponding results with respect to the \(L^2\) or empirical norm exist.

To address this compatibility issue, we propose to use HSIC based on Besov-norm-induced kernels for testing independence under the assumption that the sample paths of all random functions belong to Besov spaces, a larger class of functions than Sobolev spaces. To recover each trajectory, we develop a new wavelet thresholding method for pre-smoothing. Its theoretical compatibility with the proposed HSIC is given in Section 4. In the rest of this section, we first introduce wavelets (e.g., Ogden 1997; Vidakovic 2009; Morettin, Pinheiro, and Vidakovic 2017) together with other related results and then details the proposed two-step procedure.

3.1. Wavelets and Besov Sequence Norms

Following the Cohen-Daubechies-Jawerth-Vial (CDJV) construction (Cohen et al. 1993), let father and mother wavelets be \(\phi, \psi \in C^R[0, 1]\), respectively, with \(D\) vanishing moments (e.g., Daubechies 1992) where \(C^R[0, 1]\) is the space of all functions on \([0, 1]\) with \(R\)-order continuous derivatives. We consider a Besov space \(B^s_{p,q}[0, 1]\) with norm \(\|f\|_{B^s_{p,q}[0, 1]}\) of which smoothness parameter \(s\) satisfies \(1/p < s < \min(R, D)\) such that \(B^0_{p,q}[0, 1]\) can be embedded continuously in \(C[0, 1]\). Formal definitions of \(B^s_{p,q}[0, 1]\) and its norm \(\|f\|_{B^s_{p,q}[0, 1]}\) are given in Section S1.2 in the supplementary material. Then for any function \(f \in B^s_{p,q}[0, 1] \cap L^2[0, 1]\) and a fixed coarse scale \(L\), we have the following decomposition:

\[f(t) = \sum_{k=0}^{2^L-1} \xi_k [2^{l/2} \phi(2^lt - k)] + \sum_{j=L}^{2^L-1} \sum_{k=0}^{2^j-1} \theta_{jk} [2^{l/2} \psi(2^lt - k)], \quad t \in [0, 1].\]  

(1)

Denote \(\theta_{jk} = \xi_{2^j+k} \leq j < L, 0 \leq k < 2^j\) and \(\theta_{-1,0} = \xi_0\). Based on the wavelet coefficients of \(f\), \(\theta^l = (\theta_{-1,1}^l, \theta_{0,1}^l, \ldots, \theta_{L,1}^l, \theta_{L+1,1}^l, \ldots)\), where \(\theta^l = (\theta_{0,0}, \theta_{1,1}, \ldots, \theta_{L,0})^\top\) and \(\theta^l = \theta_{1,0}\), the Besov sequence norm \(\|f\|_{B^s_{p,q}}\) (e.g., Donoho et al. 1995; Johnstone and Silverman 2005)
is defined by
\[ ||\theta^j||_{p,q} = \left( \sum_{\beta=1}^{2^{jq}} ||\theta^j_\beta||_{p,q}^2 \right)^{1/q}, \quad s = \alpha + 1/2 - 1/p, \quad (2) \]
where \( ||\cdot||_p \) is the \( \ell_p \)-norm for vectors. Denote the corresponding vector by \( b^p,q_a = \{ a : ||a||_{p,q,a} < \infty \} \). Note that the two norms \( ||\cdot||_{p,q,a} \) and \( ||\cdot||_{p,q,a} \) are equivalent (e.g., DeVore and Lorentz 1993; Donoho et al. 1995) and obviously \( b^p,q_a \subset b^p,q \) if \( \beta \leq \alpha \). In practice, if \( f \) is observed at \( m = 2^{j+1} \) dyadic time points \( \{0/m, 1/m, \ldots, (m-1)/m\} \), the discrete wavelet transformation can be used to calculate the wavelet coefficients \( \theta^j \) with \( \theta^j = 0 \) when \( j > J \). Then, we can denote \( \theta^j = ((\theta^j)_{0}^\top, (\theta^j)_1^\top, \ldots, (\theta^j)_p^\top) \).

We can show that some Besov sequence norm can induce a characteristic kernel, which is required by HSIC.

**Theorem 1.** For \( 0 < q' < q \leq p \leq 2, \quad 0 \leq \alpha \leq \alpha' \) and \( \alpha' > 1/p \), let the semi-metric \( r_{p,q}(f,g) = ||\theta^j - \theta^k||_{p,q} \) for \( f, g \in B^p,q_a [0,1] \). The wavelet coefficients of \( f \) and \( g \) respectively. The function induced by \( r_{p,q} \), which is \( \kappa (z, \zeta) = r_{p,q}(z,0) + r_{p,q}(\zeta,0) - r_{p,q}(z,\zeta), \quad z, \zeta \in B^p,q_a[0,1] \), is a characteristic kernel.

The proof of Theorem 1 is given in Section S2.1 in the supplementary material. By Theorem 1, we can define HSIC properly based on kernels induced by Besov sequence norms. For simplicity, hereafter we focus on popular choices of \( p = q = 2 \) and \( q' = 1 \). Accordingly, we abbreviate \( B^2,q_a(0,1 \} \) and \( B^2_a \) to \( B^2 \) and \( b^2 \), respectively, and the kernel functions are
\[ \kappa (z_1, z_2) = ||\theta^{z_1}||_{b^2} + ||\theta^{z_2}||_{b^2} - ||\theta^{z_1} - \theta^{z_2}||_{b^2}, \]
\[ z_1, z_2 \in Z, 0 \leq \beta \leq \alpha, \]
for \( (Z, Z) = (X, X) \) and \( (Y, Y) \).

### 3.2. Two-Step Procedure

Let \( Z = Z \) or \( Y \). Under the setting in Section 3.1, we assume \( Z \in B^{2\gamma} \) where \( 1/2 < \alpha < \min\{R, D\} \). Note that \( B^{2\gamma} \subset b^{2\gamma} \) for \( 0 < \beta \gamma < \alpha \gamma \) so \( Z \in B^{2\gamma} \) as well. To test the independence between \( X \) and \( Y \) based on their discretely measured and noisy observations, we propose to first denote each function and then apply HSIC to the recovered functions. The two-step procedure is explicitly stated as follows:

**Step 1.** By the decomposition (1) and the resolution limitation due to a finite number of measurements \( m = 2^{j+1} \) taken for each subject, we obtain the initial wavelet coefficient estimates for each \( z_i \), denoted by \( \hat{\theta}^{z_i} = (\hat{\theta}^{z_i}_{-1})^\top, (\hat{\theta}^{z_i}_0)^\top, \ldots, (\hat{\theta}^{z_i}_p)^\top)^\top \), via the discrete wavelet transformation with the coarse scale \( L_2 \). The coarse scale \( L_2 \) may be selected by cross-validation or domain knowledge. We propose to denote \( \theta^{z_i} \) and accordingly obtain \( \hat{\theta}^{z_i} = ((\hat{\theta}^{z_i}_{-1})^\top, (\hat{\theta}^{z_i}_0)^\top, \ldots, (\hat{\theta}^{z_i}_p)^\top)^\top \) as follows. First, we let \( \hat{\theta}^{z_i}_j = \hat{\theta}^{z_i}_j \) for \( j = -1, \ldots, L_2 - 1 \). Moreover, we apply the following penalized least square to obtain \( \theta^{z_i}_j = \hat{\theta}^{z_i}_j, j = L_2, \ldots, J \):
\[ \theta^{z_i}_j = \text{arg min}_{\theta} \left\{ ||\theta^{z_i}_j - \theta||_2^2 + \delta_j^2 \text{pen}_\theta(||\theta^{z_i}_j||_0) \right\}, \]
\[ (3) \]
where \( ||\cdot||_2 \) denotes the Euclidean norm, \( ||\cdot||_0 \) denotes the number of nonzero elements, \( \delta_j = 2^{q/2}\delta \) is the noise standard deviation at the resolution level \( j \) with \( \delta > 0 \), and the penalty term \( \text{pen}_\theta(k) = k \log (\sigma_j m_\theta k)^{1/2} \) that depends on \( \tau_j > 1, \quad \tau_j = 1/2^{2q-2\gamma(j-2)} \). The proposed procedure in (3) is comparable to denoising a certain type of correlated noise (see technical assumptions in Theorem 2 in Section 4). Compared to the penalty (12.34) in Johnstone (2019), we employ a different \( \tau \) in the penalty (3) such that the pre-smoothing error measured by the Besov sequence norm used in the empirical HSIC in Step 2 below converges to zero if \( m \) diverges to infinity (see Theorem 2 in Section 4). This can guarantee the compatibility between this and the next steps.

**Step 2.** Since the wavelet coefficient estimates \( \hat{\theta}^{z_i} \in b^{2\gamma} \subset b^{2\gamma} \) and \( \hat{\theta}^{z_i} \in b^{2\gamma} \subset b^{2\gamma} \), \( i = 1, \ldots, n \), for \( \beta \gamma < \alpha \gamma \), \( \beta \gamma < \alpha \gamma \), we may apply HSIC to the denoised functions where the kernels \( k_X \) and \( k_Y \) are defined by \( k_{X_X} \) and \( k_{Y_Y} \), respectively, as defined in Theorem 1. Explicitly, we have \( \gamma(P_{n,\hat{X},\hat{Y},\kappa_X,\kappa_Y}) = n^{-2} \text{tr} (G^{\hat{X}} H^\hat{Y} H G^{\hat{X}}) \), where
\[ G^{\hat{X}} = \left( ||\theta^{z_i}_j||_{b^{2\gamma}} + ||\theta^{z_j}_j||_{b^{2\gamma}} - ||\theta^{z_i}_j - \theta^{z_j}_j||_{b^{2\gamma}} \right)_{1 \leq j \leq n}, \]
and
\[ G^{\hat{Y}} = \left( ||\theta^{z_i}_j||_{b^{2\gamma}} + ||\theta^{z_j}_j||_{b^{2\gamma}} - ||\theta^{z_i}_j - \theta^{z_j}_j||_{b^{2\gamma}} \right)_{1 \leq j \leq n}. \]
By adopting \( \rho_{b\gamma} \) and \( \rho_{b\gamma} \), which \( \beta \gamma < \alpha \gamma \) and \( \beta \gamma < \alpha \gamma \) to construct kernels, we are able to make the pre-smoothing step theoretically compatible with the HSIC. As revealed in Theorems 2 and 3 in Section 4, if the observations of all functions are sufficiently dense, the denoising error is asymptotically negligible in the asymptotic distribution of the HSIC. This is a key benefit of using wavelets and Besov norms for pre-smoothing.

In Section 4, the asymptotic distribution of \( \gamma(P_{n,\hat{X},\hat{Y},\kappa_X,\kappa_Y}) \) is developed in Theorem 3 under the independence hypothesis. Despite its theoretical appeal, the asymptotic distribution unfortunately involves many unknown quantities. Therefore, we suggest using permutations to perform the independence test which, as shown in Theorem 4, can control the Type I error probability and is also consistent.

In this section, we show that the proposed two-step procedure converges. First, the assumption encompasses both short- and long-range dependences of the noise process when it is a white noise model by allowing correlation among noise terms to some extent. First, the assumption encompasses both short- and long-range dependences of the noise process when it is a stationary and Gaussian (Johnstone and Silverman 1997). For the short-range dependence case where \( \varsigma_Z = 0 \), there is no variance inflation with the increase of level \( j \). For the long-range dependence case, when \(-1/2 < \varsigma_Z < 0\), the process \( e^\theta_z(t) = m - 1 \sum_{l=1}^{[m]} e^\theta_z l \) can be approximated by a fractional Brownian motion \( B^H(t), H = 1/2 - \varsigma_Z \) (Taqqu 1975), which is widely used for modeling long-range dependence. Then the convergence rate (with \( \hat{\varsigma}_Z \) replaced by \( \hat{\varsigma}_Z - 2H = \hat{\varsigma}_Z - 2/2 - \varsigma_Z \) in Theorem 2) is asymptotically minimax up to a constant. When \( \beta_Z = 0 \) in particular, this rate coincides with those of Wang (1996) and Johnstone and Silverman (1997). Second, when \( \varsigma_Z > 0, \varsigma_Z \) captures noise amplification as reflected in the noise level \( \hat{\varsigma}_Z = 2\hat{\varsigma}_Z \), which is common in the linear inverse problem (Abramovich and Silverman 1998; Johnstone and Paul 2014), for example, \( \varsigma_Z = 1/2 \) for the two-dimensional Radon transformation (Donoho 1995).

Since the HSIC is constructed based on the kernels induced by \( \rho_{\phi_X} \) and \( \rho_{\phi_Y} \), the same norms used to evaluate the denoising error as in Theorem 2, the compatibility between the pre-smoothing by wavelet soft-thresholding and HSIC is theoretically guaranteed. As shown in Theorem 3, the effect of the denoising error on the distribution of the HSIC is asymptotically negligible for dense functional data.

To develop the asymptotic distribution of \( \gamma(P_{\tilde{\kappa}_{X,Y}, \kappa_X, \kappa_Y}) \), we further define the centered kernel for \( \kappa_X \) by \( \tilde{\kappa}_X(X, X') = (\kappa_X(X, \cdot) - \mathbf{P}^X(P_X)), \kappa_Y(X', \cdot) - \mathbf{P}^Y(P_X))_{H(\kappa_X)} \). Further, define an integral kernel operator \( S_{\kappa_X} : \mathbf{H}(\kappa_X) \to \mathbf{H}(\kappa_X) \) by \( S_{\kappa_X}(g) = \int_X \tilde{\kappa}_X(x, \cdot) g(x) dP_X(x) \) for any \( g \in \mathbf{H}(\kappa_X) \). An integral kernel operator \( S_{\kappa_X} \) for \( Y \) can be similarly defined.
Theorem 3. Under the same assumptions of Theorem 2, if $m$ satisfies
\[ m^{-\frac{1}{2}+\alpha/2e^2}o(n^{-1}), \]
for both $Z = X$ and $Z = Y$, then
\[ n\gamma(P_{n,X}, X, Y) \sim \sum_{u=1}^{n} \sum_{v=1}^{n} \mu_{uv}N_{uv}^2, \]
where \( \sim \) represents weak convergence, $N_{uv} \sim N(0,1)$, $u, v \geq 1$ are iid and \( \mu_{uv} = e^{\alpha/2} \) are eigenvalues of $S_{X, Y}$, and $S_{X, Y}$, respectively.

The proof of Theorem 3 is given in Section S2.3 in the supplementary material. The asymptotic distribution of $\gamma(P_{n,X}, X, Y)$ in Theorem 3 is the same as that for fully observed $(X, Y)$ : $X_i \in B^{f_1}, Y_i \in B^{f_2}, i = 1, \ldots, n$ (Sejdinovic et al. 2013). The requirement (4) ensures that the error due to the denoising procedure is asymptotically negligible under $\beta^2$ norm if the measurements are sufficiently dense. In general, for fixed $\alpha$, $\beta$ and $\kappa$, the order of $m$ should be higher than $n^{1/\gamma}$ when $r = (\alpha - \beta^2)/(2\alpha x + 2\delta^2 + 1)$ which, for example, is $n^{10/3}$ if $(\alpha, \beta, \kappa) = (2, 1/2, 0)$ and $n^4$ if $(\alpha, \beta, \kappa) = (3, 1, 1/2)$.

Since the asymptotic reference distribution of $\gamma(P_{n,X}, X, Y)$ when $X$ and $Y$ are assumed independent involves many unknown quantities, in practice we perform the test by permutation. As shown in Theorem 4, the permutation test can control the Type I error probability and is also consistent.

Theorem 4 (Permutation Test). Let the level of significance be $\alpha \in (0, 1)$. If the null hypothesis that $X$ and $Y$ are independent is true, then the permutation test of $\gamma(P_{n,X}, X, Y)$ based on a finite number of permutations rejects the null hypothesis with probability at most $\alpha$. If the alternative hypothesis that $X$ and $Y$ are dependent is true and the assumptions of Theorem 2 and (4) hold, then the permutation test of $\gamma(P_{n,X}, X, Y)$ based on $B \geq 1/\alpha - 1$ permutations is consistent, that is, $P(\hat{p}_{\hat{X}, \hat{Y}} \leq \alpha) \to 1$ as $n \to \infty$, where $\hat{p}_{\hat{X}, \hat{Y}}$ is the p-value.

The proof of Theorem 4 is given in Section S2.4 in the supplementary material. Theorem 4 shows that the proposed permutation test is also theoretically compatible with the proposed wavelet thresholding method in Step 1.

5. Tuning Parameter Selection

In this section, we discuss the selection of tuning parameters involved in the two-step procedure proposed in Section 3. These include $\kappa$, $\tau_X$, $\tau_Y$, and $\delta$ in Step 1 and $\beta_X$ and $\beta_Y$ in Step 2, where $Z = X$ or $Y$.

First, to guarantee $\tau_X > 1$ and $\tau_Y > e$, we suggest $\tau_X = 1.0001$ and $\tau_Y = 1.0001e$ which are slightly larger than their respective lower bounds, unless domain knowledge is available.

Second, for $\kappa$ which captures noise amplification and $\delta$ which reflects the noise level, we adopt crude estimates for them based on the top two levels of the wavelet coefficients (Johnstone and Silverman 1997). Explicitly, we obtain $\hat{\kappa} = \log_2(\delta_{J1}/\delta_{J-1})$ and $\hat{\delta} = \delta_{J}/2^{J+1}$, where $\delta_{J}$ is median \( \sqrt{\textrm{median}(P_{z_{jk} k} : k = 0, \ldots, 2^{J-1})} / \textrm{median}(W) \) for $j = J - 1, J$, and $W$ is a standard normal random variable.

Finally, for the smoothness parameter $\beta_X$, we will first discuss its role in dependency detection and then propose a data-adaptive selection method for it.

In Section 4, Theorem 2 seems to imply that given $\alpha_X$ and $\alpha_Y$, the best choice is $\beta_X = \beta_Y = 0$ because the corresponding denoising error attains the best rate of convergence. However, this choice of $\beta_X$ and $\beta_Y$ may result in a poor dependency detection especially when the dependency of $X$ and $Y$ originates from high frequencies.

For illustration, by Definition 1 and (2), we consider the first-order approximation (Chakraborty and Zhang 2019, theorem 5.1)
\[ \gamma(P_{XY}, X, Y) \approx \sum_{jX, jY} \gamma \left( P_{XY}, 2^{jX+Y} \kappa_X, 2^{jX+Y} \kappa_Y \right), \]
where $\kappa_{X} (z, z') = ||\theta_{z}||^2 + ||\theta_{z'}||^2 - ||\theta_{z} - \theta_{z'}||^2$ for $j_{Z} \geq -1$, with $(z, Z) = (x, X, X)$ or $(y, Y, Y)$ and Euclidean norm $||\cdot||_2$, and $c_{X} = 4 \sqrt{E}|X - X'|^2|P_X E|Y - Y'|^2|P_Y - X$ and $Y$ being the independent copies of $X$ and $Y$, respectively. Apparently, $\gamma(P_{XY}, 2^{jX+Y} \kappa_X, 2^{jX+Y} \kappa_Y)$ measures the dependency contribution to the HSIC at $j_X$ and $j_Y$ of $X$ and $Y$ respectively, which is zero if and only if $X$ and $Y$ are independent at $j_X$ and $j_Y$. If $\beta_X = \beta_Y = 0$, the scaling factors $2^{jX} = 2^{jY}$ for all $j_X \geq -1, j_Y \geq -1$ and it will be very difficult to detect the dependency between $X$ and $Y$ at high frequencies since the dependency contributions contained at high frequencies are very likely to be overwhelmed by the independent signals at low frequencies. Therefore, we aim to select $\beta_X$ and $\beta_Y$ such that the dependency contributions at high frequencies, if any, are detectable.

The idea of the proposed tuning method is to balance the dependency contributions to HSIC at all frequency scales such that they are approximately the same. To lessen the computational burden, a marginal selection algorithm is proposed in the sense that the optimal $\beta_X$ is selected only based on $X$ without reliance on $Y$. Note that, by Appendix A in Sejdinovic et al. (2013) and the properties of distance covariance, the dependency contribution at each $j_X, j_Y \geq -1$ satisfies
\[ \gamma\left(P_{XY}, 2^{jX+Y} \kappa_X, \kappa_Y \right) \leq 2^{2jX+2jY} \gamma\left(P_{XY}, \kappa_X, \kappa_Y \right), \]
where $\gamma(P_{Z, Z}) = ||P_{Z}^{(g_{z})} \otimes P_{Z}^{(g_{z})} (P_{ZZ}^{(g_{z})}(s, \cdot)) - P_{Z}^{(g_{z})}(P_{Z}^{(g_{z})}(s, \cdot))||^2_{H_{k}(\kappa_{Z})^{g_{z}}, \kappa_{Z}^{g_{z}}, JZ} \geq -1$, is essentially a distance variance (Székely, Rizzo, and Bakirov 2007) with $(z, Z) = (x, X, X)$ or $(y, Y, Y)$ (Sejdinovic et al. 2013). Thus, we propose to select $\beta_X$ by balancing $2^{2jX+2jY} \gamma\left(P_{XY}, \kappa_X, \kappa_Y \right)$ at all $j_X \geq -1$. If $2^{2jX+2jY} \gamma\left(P_{XY}, \kappa_X, \kappa_Y \right) \approx C$ where $C > 0$ is a constant, then $2\beta_X + 1/2 \log_2 2^{2jX+2jY} \gamma\left(P_{XY}, \kappa_X, \kappa_Y \right) \approx \log_2 C$, so $\beta_X$ may
be selected as the estimated slope of the linear regression on 
\(-2f_j \log_2 \gamma(P_{X_j \kappa_X^{(j)}})/2\).

In practice, we could estimate \(\gamma(P_{X_j \kappa_X^{(j)}})\) by \(\gamma(P_{n,\hat{X}\kappa_X^{(j)}})\)
for each \(j\geq -1\), but its accuracy is poor for very high
frequencies due to noise contamination. Thus, we only consider
\(j = \max_{j\geq 1} \gamma(P_{n,\hat{X}\kappa_X^{(j)}}) \geq \gamma(P_{n,\hat{\epsilon}\kappa_X^{(j)}})\) where
\(\hat{\epsilon} = \hat{X} - \bar{X}\) is the residual, such that the distance variances
of all \(j\leq \hat{j}\) are not smaller than that of the residual. If a
known frequency band is of interest in the context of a study,
for example, the alpha band of brain signals, then one may
alternatively select \(\beta_X\) by balancing \(2^{2f_j \hat{j}} \sqrt{\gamma(P_{X_j \kappa_X^{(j)}})}\) over
that frequency band. Last, we remark that the computational
benefit of the proposed marginal approach for tuning parameter
selection is substantial when many tests have to be performed,
such as in the functional connectivity analysis (Section 7).

6. Simulation

In this section we evaluate the numerical performance of our
proposed wavelet-based HSIC method wavHSIC in both con-
trolling the Type I error probability and statistical power. We
also compare it with a few representative existing methods,
including

(a) Pearson correlation (Pearson). It is a one-sample \(t\)-test
based on Fisher-Z transformed correlation coefficients
of all subjects. The correlation coefficient for each subject
is obtained by applying the Pearson correlation formula to the
bivariate time series of the subject, without adjusting for any
possible dependence within the time series. It is a popular
functional connectivity measure in neuroscience (e.g., He
et al. 2012). (b) Dynamical correlation (dnm, Dubin and Müller 2005). It
is defined as the expectation of the cosine of the \(L^2\)
angle between the standardized versions of two random func-
tions. (c) Global temporal correlation (gtmp, Zhou, Lin, and Wang
2018). It is the integral of the Pearson correlation obtained at
each time point. (d) Bias-corrected distance covariance (dCov-c, Székely
and Rizzo 2013). It is a \(t\)-test designed to correct the bias of
distance covariance for high-dimensional multivariate data.
We apply it by treating the discrete measurements of two
random functions as multivariate data. If the bias is not
corrected, it is equivalent to wavHSIC with \(\beta_X = \beta_Y = 0\).
(e) Functional principle component analysis (FPCA)-based
distance covariance (FPCA, Kosorok 2009). The distance
covariance (Székely, Rizzo, and Bakirov 2007) is applied to
top functional principle component (FPC) scores which
cumulatively account for 95% of the variation of each
random function. When all FPC scores are used, it is
equivalent to wavHSIC when \(\beta_X = \beta_Y = 0\).
(f) Functional linearity test (KMSZ, Kokoszka et al. 2008).
It is an approximate chi-squared test for the nullity of
the coefficient function by assuming a functional linear
model between the two random functions. The model
fitting requires a satisfactory approximation of each random
function by its top FPC scores and we select those which
cumulatively account for 95% of variation of each random
function.

(g) Permutation-based functional linearity test (KMSZ-p). It
is the same as KMSZ except that the \(p\)-value is obtained
by permutation. Such a modification can be regarded as a
finite-sample correction of KMSZ.

(h) Projection-based mean independence test (PSS, Patilea,
Sánchez-Sellero, and Saumard 2016). For a functional
response \(Y\) and a functional predictor \(X\), PSS aims to
test the conditional mean independence of \(Y\) given \(X\),
that is, \(E(Y|X) = E(Y)\), a.s. PSS is a model-free test that
does not specify a model for \(E(Y|X)\). It requires a finite-
dimensional projection of \(X\) and uses wild bootstrap to find
critical values. To implement PSS, we used the R package
fdaHsic\(^1\), which is publicly available at http://webspersois.
usc.es/persoais/cesar.sanchez/.

The first five (a)–(e) in comparison are model-free methods.
KMSZ is one of the most popular model-based methods in the
FDA literature, but it can only test for linearity. PSS and FMD
handle nonlinear effects of the functional predictor, but only
on the mean of the functional response, so they can only test a
weaker notion of independence. Hereafter, for bivariate random
functions \((X, Y)\), PSS\((Y \sim X)\) denotes testing \(E(Y|X) = E(Y)\),
a.s. using PSS. Moreover, PSS(Omnibus) denotes the omnibus
test which takes the smaller \(p\)-value between those obtained by
PSS\((Y \sim X)\) and PSS\((X \sim Y)\) respectively. FMD\((X \sim Y)\),
FMD\((Y \sim X)\), and FMD(Omnibus) are similarly defined.
To obtain \(p\)-values, 1999 permutations were used for wavHSIC,
dnm, gtmp, FPCA, and KMSZ-p; while 1999 bootstrap samples
were used for PSS and FMD. We declare statistical significance
in each simulated data based on the level of significance 0.05.

We generated 199 simulated datasets, where the number
199 is chosen to prevent empirical Type I and Type II error
probabilities from coinciding with the level of significance 0.05.
In each simulated dataset \(n = 50\) or 200 independent sub-
jects with bivariate functions \((X_i(t), Y_i(t)) : t \in [0,1], i = 1,\ldots,n\) were generated where for the \(i\)th subject, \(X_i(t)=\sum_{k=1}^{16} \xi_k \phi_k(t)\) and \(Y_i(t)=\sum_{k=1}^{16} \xi_k \phi_k(t+0.2)\) with \(\phi_{2k-1}(t) = \sqrt{2} \cos(2\pi k t), \phi_{2k}(t) = \sqrt{2} \sin(2\pi k t)\) for \(k = 1,\ldots,8\). We
considered three settings with different dependency structures
of the bivariate functional data which are controlled by the FPC
scores \((t_{ijk}, \zeta_{ik}) : k = 1,\ldots,16; i = 1,\ldots,n\).

\(^1The package is only for Windows platform. For the user-chosen parameters
required by this package, we followed the recommendation in Section 4.1
of Patilea, Sánchez-Sellero, and Saumard (2016) and set the bandwidth
\(h = n^{-2/5}\), penalty coefficient \(\omega = 2\), grid size \(n_q = 50\) and number of
FPCs which cumulatively account for 95% of the variation of the functional
predictor.
• Setting 1. We generated $\eta_{ik} \sim N(0, k^{-1.05})$, $k = 1, \ldots, 16$ and $\zeta_{ik} \sim N(0, k^{-1.2})$, $k = 1, \ldots, 16$ independently.
• Setting 2. With $\rho = 0$ for $k = 1, \ldots, 8$ and $\rho = 0.6$ for $k = 9, \ldots, 16$, we generated

$$\begin{pmatrix} \eta_{ik} \\ \zeta_{ik} \end{pmatrix} \sim N\left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} k^{-1.05} & \rho k^{-1.125} \\ \rho k^{-1.125} & k^{-1.2} \end{bmatrix} \right).$$

• Setting 3. For $k = 1, \ldots, 8$, $\eta_{ik} \sim N(0, k^{-1.05})$ was generated independently of $\zeta_{ik} \sim N(0, k^{-1.2})$. For $k = 9, \ldots, 16$, $\eta_{ik} \sim N(0, k^{-1.05})$ and $\zeta_{ik} = \eta_{ik}^2 - \eta_{ik}^2$.

Apparently, $X$ and $Y$ are independent in Setting 1 and dependent in Settings 2 and 3. In Setting 2, the FPC scores of $X$ and $Y$ are linearly correlated but only at high spectral frequencies; while in Setting 3 they are linearly uncorrelated but dependent only at high spectral frequencies, so it is more difficult to detect dependency for all methods in Setting 3 than Setting 2.

Both functions are measured at $m = 64$ or 256 equidistant points on the time domain [0, 1]. We added Gaussian noise to all measurements with signal-to-noise ratio (SNR) = 4 or 8, which is the variance of all measurements over the noise variance. The noise terms were generated independently across subjects. Within each subject, we experimented with both independent (white noise) and dependent (correlated noise) settings. For the dependent setting, the Gaussian noise was generated by differencing the fractional Brownian motion with Hurst exponent 0.7.

Since all methods in comparison require noiseless functions, we used the same denoising procedure in Step 1 for all of them for fairness. We chose the CDJV wavelet basis functions with vanishing moment $D = 10$ for both $X$ and $Y$, which leads to $\alpha_X = \alpha_Y \approx 2.902$ (Daubechies 1992). The tuning parameters $\beta_X$ and $\beta_Y$ were selected by the method in Section 5. The results are given in Tables 1–6.

Tables 1 and 4 show that all methods are almost always able to control Type I error probabilities except for PSS($X \sim Y$) and the two omnibus tests when the two random functions are truly independent. Relatively, KMSZ is very conservative in many cases and KMSZ-p corrects its $p$-values to some extent. However, KMSZ-p seems more likely to detect spurious dependency when $(n, m) = (50, 64)$, so does dCov-c when $(n, m) = (200, 256)$.

Tables 2, 3, 5, and 6 show that the statistical powers of all methods typically improve when one of $n$, $m$, and SNR increases under Setting 2, but unnecessarily under Setting 3 except for KMSZ, KMSZ-p, and wavHSIC. This demonstrates the difficulty of setting 3 in detecting dependency to some extent. Except wavHSIC, all model-free methods have very low powers in all scenarios under either Setting 2 or 3, which indicates their poor performances in detecting linear dependency in high frequencies or nonlinear dependency. The performance of KMSZ is satisfactory for $n = 200$ under Setting 2 when the relationship between $X$ and $Y$ is truly linear. KMSZ-p improves the statistical power of KMSZ further for $n = 50$ under Setting 2 by permutation. However, both KMSZ and KMSZ-p are poor at testing nonlinear dependency in Setting 3. The performances of PSS and FMDD, which can detect nonlinear mean dependency, are comparable with those of dCov-c and FPCA in Settings 2 and 3, but worse than those of KMSZ and KMSZ-p in Setting 2 where $X$ and $Y$ are linearly dependent.

Tables 2, 3, 5, and 6 also demonstrate the appealing performance of wavHSIC. It is always the most powerful method, and substantially better than the other methods. Only the powers of KMSZ and KMSZ-p are comparable with those of wavHSIC when the sample size $n = 200$ is large and the linearity assumption is valid under Setting 2. For fixed $(n, m, \text{SNR})$, the medians of the selected parameters $\beta_X$ and $\beta_Y$ for wavHSIC are always similar between Settings 2 and 3 since they were tuned marginally regardless of the dependency structure. On average, both $\beta_X$ and $\beta_Y$ were considerably away from zero, which confirms the need and benefit of choosing them properly to enhance the detection sensitivity of wavHSIC.

We also performed an additional simulation study described in Section S3.2 of the supplementary material, which follows the same settings in Section 1.2 of the supplementary material of Lee, Zhang, and Shao (2020). The results also demonstrate the superiority of wavHSIC.

### Table 1. Empirical Type I error probabilities for all methods under Setting 1 with white noise.

| Setting 1 with white noise | $n = 50$ | $n = 200$ |
|---------------------------|----------|----------|
|                           | $m = 64$ | $m = 256$ | $m = 64$ | $m = 256$ |
|                           | SNR=4    | SNR=8    | SNR=4    | SNR=8    | SNR=4    | SNR=8    |
| Pearson                   | 0.0452   | 0.0352   | 0.0503   | 0.0452   | 0.0704   | 0.0704   | 0.0553   | 0.0553   |
| dnm                       | 0.0452   | 0.0452   | 0.0653   | 0.0503   | 0.0553   | 0.0553   | 0.0653   | 0.0603   |
| gtemp                     | 0.0503   | 0.0603   | 0.0653   | 0.0603   | 0.0653   | 0.0653   | 0.0603   | 0.0603   |
| dCov-c                    | 0.0653   | 0.0603   | 0.0603   | 0.0603   | 0.0503   | 0.0603   | 0.0704   | 0.0754   |
| FPCA                      | 0.0503   | 0.0452   | 0.0503   | 0.0452   | 0.0452   | 0.0503   | 0.0503   | 0.0452   |
| KMSZ                      | 0.0201   | 0.0101   | 0.0101   | 0.0151   | 0.0201   | 0.0151   | 0.0402   | 0.0251   |
| KMSZ-p                    | 0.0804   | 0.0905   | 0.0553   | 0.0402   | 0.0302   | 0.0352   | 0.0402   | 0.0352   |
| PSS($X \sim Y$)           | 0.0804   | 0.0754   | 0.1005   | 0.0804   | 0.0653   | 0.0905   | 0.0402   | 0.0955   |
| PSS($Y \sim X$)           | 0.0402   | 0.0754   | 0.0704   | 0.0452   | 0.0302   | 0.0754   | 0.0352   | 0.0305   |
| PSS(Omnibus)              | 0.1156   | 0.1307   | 0.1558   | 0.1106   | 0.0953   | 0.1407   | 0.0754   | 0.1407   |
| FMDD($X \sim Y$)          | 0.0553   | 0.0552   | 0.0704   | 0.0653   | 0.0503   | 0.0603   | 0.0553   | 0.0603   |
| FMDD($Y \sim X$)          | 0.0553   | 0.0552   | 0.0704   | 0.0653   | 0.0503   | 0.0603   | 0.0503   | 0.0503   |
| FMDD(Omnibus)             | 0.0653   | 0.0603   | 0.0704   | 0.0704   | 0.0653   | 0.0704   | 0.0704   | 0.0704   |
| wavHSIC                   | 0.0452   | 0.0352   | 0.0503   | 0.0653   | 0.0402   | 0.0402   | 0.0251   | 0.0251   |
| median($\beta_X$)         | 0.948    | 0.989    | 0.983    | 0.991    | 0.959    | 1.000    | 0.990    | 1.001    |
| median($\beta_Y$)         | 0.671    | 0.724    | 0.733    | 0.741    | 0.696    | 0.745    | 0.747    | 0.761    |

**NOTE:** The last two rows provide the medians of the selected $\beta_X$ and $\beta_Y$ for wavHSIC.
Table 2. Empirical powers for all methods under Setting 2 with white noise.

| Setting 2 with | m = 64 | m = 256 | m = 64 | m = 256 |
|---------------|--------|---------|--------|---------|
| white noise   | n = 50 | n = 200 | n = 50 | n = 200 |
| Power         | SNR=4  | SNR=8   | SNR=4  | SNR=8   |
| Pearson       |        |         |        |         |
| dnm           | 0.0854 | 0.0804  | 0.0804 | 0.0804  |
| gtemp         | 0.0653 | 0.0653  | 0.0804 | 0.0754  |
| dCov-c        | 0.0955 | 0.1055  | 0.0804 | 0.0905  |
| FPCA          | 0.1558 | 0.1608  | 0.1307 | 0.1508  |
| KMSZ          | 0.4221 | 0.4925  | 0.5025 | 0.5075  |
| KMSZ-p        | 0.7035 | 0.7889  | 0.7688 | 0.7990  |
| PSS(X ~ Y)    | 0.0955 | 0.1055  | 0.0804 | 0.0905  |
| PSS(Y ~ X)    | 0.0653 | 0.0653  | 0.0553 | 0.0653  |
| PSS(Omnibus)  | 0.1558 | 0.1608  | 0.1307 | 0.1508  |
| FMDD(X ~ Y)   | 0.0854 | 0.0905  | 0.0905 | 0.0854  |
| FMDD(Y ~ X)   | 0.0754 | 0.0804  | 0.0704 | 0.0653  |
| FMDD(Omnibus) | 0.0955 | 0.1005  | 0.0905 | 0.0854  |
| wavHSIC       | 0.9548 | 0.9849  | 0.9849 | 0.9899  |
| PSS               |        |         |        |         |
| KMSZ          | 0.0101 | 0.0101  | 0.0201 | 0.0251  |
| KMSZ-p        | 0.1106 | 0.0854  | 0.1307 | 0.1357  |
| PSS(X ~ Y)    | 0.0754 | 0.0854  | 0.0905 | 0.1055  |
| PSS(Y ~ X)    | 0.0653 | 0.0553  | 0.0754 | 0.0804  |
| PSS(Omnibus)  | 0.1357 | 0.1307  | 0.1508 | 0.1658  |
| FMDD(X ~ Y)   | 0.0804 | 0.0804  | 0.0804 | 0.0804  |
| FMDD(Y ~ X)   | 0.0955 | 0.1005  | 0.0955 | 0.1005  |
| FMDD(Omnibus) | 0.1005 | 0.1005  | 0.0955 | 0.1005  |
| wavHSIC       | 0.2613 | 0.3618  | 0.3367 | 0.407   |
| median(β_X)   | 0.942  | 0.587   | 0.975  | 0.983   |
| median(β_Y)   | 0.674  | 0.720   | 0.741  | 0.752   |

NOTE: The last two rows provide the medians of the selected β_X and β_Y for wavHSIC.

Table 3. Empirical powers for all methods under Setting 3 with white noise.

| Setting 3 with | m = 64 | m = 256 | m = 64 | m = 256 |
|---------------|--------|---------|--------|---------|
| white noise   | n = 50 | n = 200 | n = 50 | n = 200 |
| Power         | SNR=4  | SNR=8   | SNR=4  | SNR=8   |
| Pearson       |        |         |        |         |
| dnm           | 0.0452 | 0.0402  | 0.0603 | 0.0603  |
| gtemp         | 0.0804 | 0.0804  | 0.0754 | 0.0704  |
| dCov-c        | 0.0754 | 0.0804  | 0.0754 | 0.0704  |
| FPCA          | 0.1558 | 0.1608  | 0.1307 | 0.1508  |
| KMSZ          | 0.0101 | 0.0101  | 0.0201 | 0.0251  |
| KMSZ-p        | 0.1106 | 0.0854  | 0.1307 | 0.1357  |
| PSS(X ~ Y)    | 0.0754 | 0.0854  | 0.0905 | 0.1055  |
| PSS(Y ~ X)    | 0.0653 | 0.0553  | 0.0754 | 0.0804  |
| PSS(Omnibus)  | 0.1357 | 0.1307  | 0.1508 | 0.1658  |
| FMDD(X ~ Y)   | 0.0804 | 0.0804  | 0.0804 | 0.0804  |
| FMDD(Y ~ X)   | 0.0955 | 0.1005  | 0.0955 | 0.1005  |
| FMDD(Omnibus) | 0.1005 | 0.1005  | 0.0955 | 0.1005  |
| wavHSIC       | 0.2613 | 0.3618  | 0.3367 | 0.407   |
| median(β_X)   | 0.942  | 0.587   | 0.975  | 0.983   |
| median(β_Y)   | 0.724  | 0.771   | 0.773  | 0.790   |

NOTE: The last two rows provide the medians of the selected β_X and β_Y for wavHSIC.

Table 4. Empirical Type I error probabilities for all methods under Setting 1 with correlated noise.

| Setting 1 with | m = 64 | m = 256 | m = 64 | m = 256 |
|---------------|--------|---------|--------|---------|
| white noise   | n = 50 | n = 200 | n = 50 | n = 200 |
| Type I error rate | SNR=4  | SNR=8   | SNR=4  | SNR=8   |
| Pearson       |        |         |        |         |
| dnm           | 0.0352 | 0.0352  | 0.0452 | 0.0452  |
| gtemp         | 0.0402 | 0.0452  | 0.0603 | 0.0603  |
| dCov-c        | 0.0603 | 0.0553  | 0.0553 | 0.0653  |
| FPCA          | 0.0452 | 0.0452  | 0.0452 | 0.0402  |
| KMSZ          | 0.0151 | 0.0000  | 0.0151 | 0.0151  |
| KMSZ-p        | 0.0804 | 0.0754  | 0.0553 | 0.0452  |
| PSS(X ~ Y)    | 0.0452 | 0.0553  | 0.0603 | 0.0955  |
| PSS(Y ~ X)    | 0.0553 | 0.0704  | 0.0553 | 0.0754  |
| PSS(Omnibus)  | 0.1005 | 0.1156  | 0.1106 | 0.1508  |
| FMDD(X ~ Y)   | 0.0553 | 0.0603  | 0.0653 | 0.0653  |
| FMDD(Y ~ X)   | 0.0452 | 0.0452  | 0.0503 | 0.0503  |
| FMDD(Omnibus) | 0.0553 | 0.0653  | 0.0704 | 0.0704  |
| wavHSIC       | 0.0402 | 0.0402  | 0.0553 | 0.0653  |
| median(β_X)   | 1.014  | 1.020   | 0.996  | 0.997   |
| median(β_Y)   | 0.752  | 0.765   | 0.750  | 0.754   |

NOTE: The last two rows provide the medians of the selected β_X and β_Y for wavHSIC.
Remark 3. It is worth noting that the development of the asymptotic distribution of wavHSIC as in Theorem 3 requires the number of measurements per curve \( m \) to be large compared to the sample size \( n \) (see (4)), but the simulation results here show that the finite sample performance of wavHSIC is still satisfactory, even when \( m \) is small relatively to \( n \). However, this is not entirely surprising. First, under the null hypothesis that \( X \) and \( Y \) are independent, a poor pre-smoothing due to a relatively small \( m \) does not inflate the empirical Type I error probability since the remaining noise does not enhance the dependency between \( X \) and \( Y \) and the critical value is obtained by permutation. Second, under the alternative hypothesis that \( X \) and \( Y \) are dependent, as long as \( m \) is sufficiently large such that the dependency signals can be captured by the wavelet coefficients, wavHSIC can still detect dependency, but its power may be worse if (4) is not satisfied.

### 7. Real Data Application

We applied our proposed method to study human brain functional connectivity using the MEG dataset collected by the HCP. MEG measures magnetic fields generated by human neuronal activities with a high temporal resolution. Before source reconstruction, the signals from all MEG sensors outside head were preprocessed following the HCP MEG pipeline reference (www.humanconnectome.org/software/hcp-meg-pipelines) and the preprocessed data are publicly accessible from the HCP website. To obtain the electric activity signals from cortex regions, we applied the source reconstruction procedure of MEG signals to the cerebral cortex atlas provided by Glasser et al. (2016) using the linearly constrained minimum variance beamforming method in the MATLAB package FieldTrip.

### Table 5. Empirical powers for all methods under Setting 2 with correlated noise.

| Setting 2 with | \( m = 64 \) | \( m = 256 \) | \( n = 50 \) | \( n = 200 \) |
|---------------|-------------|-------------|-------------|-------------|
| white noise   |             |             |             |             |
| Power         | SNR=4       | SNR=8       | SNR=4       | SNR=8       |
| Pearson       | 0.0854      | 0.0854      | 0.0804      | 0.0804      |
| dnn           | 0.0653      | 0.0653      | 0.0754      | 0.0704      |
| gtemp         | 0.0653      | 0.0653      | 0.0905      | 0.0905      |
| dCov-c        | 0.1005      | 0.0955      | 0.0854      | 0.0854      |
| FPCA          | 0.0854      | 0.0804      | 0.0754      | 0.0804      |
| KMSZ          | 0.5427      | 0.5628      | 0.5126      | 0.5327      |
| KMSZ-p        | 0.8241      | 0.8141      | 0.8191      | 0.8291      |
| PSS(\(X \sim Y\)) | 0.0955  | 0.0955      | 0.0905      | 0.1005      |
| PSS(\(Y \sim X\)) | 0.0553  | 0.0653      | 0.0704      | 0.0603      |
| PSS(Omnibus)  | 0.1407      | 0.1156      | 0.1558      | 0.1508      |
| FMDD(\(X \sim Y\)) | 0.1055  | 0.0905      | 0.0854      | 0.0854      |
| FMDD(\(Y \sim X\)) | 0.0804  | 0.0905      | 0.0704      | 0.0704      |
| FMDD(Omnibus) | 0.1106      | 0.1055      | 0.0905      | 0.0905      |
| wavHSIC       | 0.9950      | 0.9950      | 0.9899      | 0.9899      |
| median(\(\beta_X\)) | 1.011  | 1.022       | 0.990       | 0.995       |
| median(\(\beta_Y\)) | 0.749  | 0.765       | 0.757       | 0.760       |

### Table 6. Empirical powers for all methods under Setting 3 with correlated noise.

| Setting 3 with | \( m = 64 \) | \( m = 256 \) | \( n = 50 \) | \( n = 200 \) |
|---------------|-------------|-------------|-------------|-------------|
| white noise   |             |             |             |             |
| Power         | SNR=4       | SNR=8       | SNR=4       | SNR=8       |
| Pearson       | 0.0402      | 0.0402      | 0.0603      | 0.0603      |
| dnn           | 0.0704      | 0.0653      | 0.0704      | 0.0704      |
| gtemp         | 0.0854      | 0.0804      | 0.0704      | 0.0603      |
| dCov-c        | 0.1055      | 0.1106      | 0.1005      | 0.1005      |
| FPCA          | 0.0854      | 0.0905      | 0.1005      | 0.1005      |
| KMSZ-p        | 0.0905      | 0.0854      | 0.1256      | 0.1256      |
| PSS(\(X \sim Y\)) | 0.0905  | 0.0905      | 0.0754      | 0.0754      |
| PSS(\(Y \sim X\)) | 0.0704  | 0.0653      | 0.0754      | 0.0854      |
| PSS(Omnibus)  | 0.1558      | 0.1558      | 0.1457      | 0.1457      |
| FMDD(\(X \sim Y\)) | 0.0804  | 0.0804      | 0.0905      | 0.0905      |
| FMDD(\(Y \sim X\)) | 0.1005  | 0.1005      | 0.0955      | 0.0955      |
| FMDD(Omnibus) | 0.1106      | 0.1055      | 0.1005      | 0.1005      |
| wavHSIC       | 0.4221      | 0.4472      | 0.4422      | 0.4422      |
| median(\(\beta_X\)) | 1.022  | 1.030       | 0.986       | 0.987       |
| median(\(\beta_Y\)) | 0.799  | 0.810       | 0.800       | 0.802       |

NOTE: The last two rows provide the medians of the selected \(\beta_X\) and \(\beta_Y\) for wavHSIC.
To study the functional dependency between cortex regions under some motor activities, we focused on motor task trials where subjects moved their right hands. There were $n = 61$ subjects in the trials. For each subject, 8004 signal curves were obtained by denoising and source reconstruction procedures with around 75 repeated trials. Within each trial, the signals were recorded about every 2 ms from $-1.2$ to 1.2 sec, where the time 0 is the starting time of the motion. Since the motion in each trial usually lasts no longer than about 0.75 sec and typically a subject finished the previous movement and received a new cue between times $-0.25$ and 0 of the next trial, we considered the time domain $[-0.2525, 0.7525]$ which covers the time period of interest, with $m = 512$ sampled time points in total.

We applied the proposed method wavHSIC to perform an independence test for every pair of the MEG signals. To implement wavHSIC, we chose the CDJV wavelet basis functions with vanishing moment $D = 4$ which leads to $\alpha \approx 1.6179$. For each signal, the smoothness parameter $\beta$ was selected by the method in Section 5. For comparison, we also provided the results for the model-based test KMSZ, KMSZ-p and two model-free tests, Pearson and FPCA. KMSZ, KMSZ-p, and FPCA were based on top FPC scores which cumulatively account for 95% of the variation of each signal. The $p$-value for testing the independence between each pair of signals were obtained by 1999 permutations for wavHSIC, FPCA, and KMSZ-p. We did not include PSS and FMDD here due to their extended computing times. See Table 7 for an illustration.

The empirical cumulative distribution functions for the $p$-values of the five methods are given in Figure 1, which shows that wavHSIC is more sensitive to detecting connectivity than the other methods. To evaluate and compare the five methods at the presence of multiple testing, we set the same discovery rate at 60% to control the number of edges, or sparsity, of each brain connectivity network, which is important in evaluating the reliability of brain network metrics (e.g., Van Wijk, Stam, and Daffertshofer 2010; Tsai 2018). In this analysis, we focus on sensorimotor areas 4, 3a, 3b, 1, and 2 on the left and right hemispheres as illustrated in Figure 3 (c) which are most related to motor task trials (Glasser et al. 2016). With a controlled discovery rate, we expect an excellent connectivity detection method to identify plenty of edges within these areas.

Figures 2 and 3 (a) provide the functional connectivity networks within these sensorimotor areas obtained by the five methods. The nodes in each area were ordered from the superio-medial cortex to infero-lateral cortex following the atlas “atlas_MMP1.0_4k.mat” in FieldTrip. Compared with KMSZ, KMSZ-p, and wavHSIC, Pearson, and FPCA are substantially less sensitive to detecting functional connectivity and their corresponding networks are less structured (see Figure 2 (a) and (b)). This demonstrates the superior performances of both KMSZ, KMSZ-p, and wavHSIC in identifying connectivity patterns within these areas which are anatomically connected and functionally related to the motor task trials. Different from the overall homogeneous pattern in the network for KMSZ, several structured dark strips appear in the network obtained by KMSZ-p and wavHSIC within sensorimotor areas 4, 3a, 3b, and 1 in the right hemisphere (see Figures 2 (c)–(d) and 3 (a)). These dark strips are much clearer in Figure 3 (a) than in Figure 2 (d). This indicates that wavHSIC can more clearly identify two sub-areas in sensorimotor areas 4, 3a, 3b, and 1 in the right hemisphere, the top left (TL) and bottom right (BR) corners respectively in these corresponding colored squares as in Figure 3 (a). The signals within these four TL sub-areas or within these four BR sub-areas are strongly connected, while the connectivities between these TL and BR sub-areas are generally weak. According to Glasser et al. (2016), the four BR sub-areas in the same hemisphere correspond to face and eye portions while the four TL sub-areas correspond to upper limbs, trunk and lower limbs portions. Since the motor task involved in this dataset is raising the right hand, the connectivity patterns detected by wavHSIC are intuitively and anatomically interpretable.

Next, we illustrate how to identify dependency structures between and within different frequency bands using wavHSIC. Explicitly, we first split the denoised wavelet coefficients of each brain signal into two parts, the low-frequency part (LF, $j \leq 3$) and high-frequency part (HF, $j > 3$), which approximately correspond to the Delta band ($\leq 4$Hz) and the Theta to the Ultra-Gamma bands ($> 4$Hz), respectively (e.g., Buzsaki 2006). Then for each pair of signals $(X, Y)$ as illustrated in Figure 3 (d), (e), and (f), we applied wavHSIC to (LF of $X$, LF of $Y$), (HF of $X$, HF of $Y$), and (LF of $X$, HF of $Y$), respectively. Their corresponding functional connectivity patterns are shown in Figure 3(d), (e), and (f), respectively. Note that the results for (HF of $X$, LF of $Y$) are included in Figure 3(f) by switching the roles of $X$ and $Y$. Apparently, the network in Figure 3(e) is very similar to that in Figure 3(a), which indicates that the functional dependency induced by this motor task mainly lies at high frequencies. Moreover, Figure 3(f) shows that there is essentially no dependency between the LF and HF signals. Last, Figure 3 (d) reveals that some dependency can be detected at

![Figure 1](image_url). Empirical cumulative distribution function for the $p$-values for testing the independence between every pair of the 8004 signals for each method.

### Table 7. Mean computing times (in seconds) based on one randomly selected pair of signals for the seven methods in comparison.

| Method | Pearson | FPCA | KMSZ | KMSZ-p | wavHSIC | PSS | FMDD |
|--------|---------|------|------|--------|---------|-----|------|
| Time   | 0.003 (0.002) | 0.103 (0.072) | 0.020 (0.003) | 0.158 (0.082) | 0.162 (0.017) | 43.076 (0.728) | 15.280 (0.253) |

NOTE: The values in parentheses are standard deviations.
Figure 2. Functional connectivity networks of the five sensorimotor areas in the left and right hemispheres. In the adjacency matrices in (a), (b), (c), and (d) obtained by the four methods, respectively, a bright entry indicates significant dependency between the corresponding signal pairs while a dark one indicates otherwise.

8. Discussion

In this article, we propose a model-free wavelet-based independence test for two random functions of which sample paths belong to possibly different Besov spaces. Our method is built upon HSIC endowed with characteristic kernels, which is zero if and only if the two random functions are independent. Since the Besov space with wavelet basis functions provides an effective modeling environment for sample paths with various levels of smoothness, HSIC with characteristic kernels induced by wavelet coefficients is capable of capturing the dependency at different frequencies. Therefore, the proposed method is especially powerful when the two random functions are dependent only at high frequencies, as demonstrated in Section 6. If the dependency is strong at low frequencies, then our simulation not presented here shows that the proposed method is not substantially advantageous over FPCA.

In the application to MEG functional connectivity, the proposed method by construction is only able to identify the unconditional dependency between two signal curves. Although metrics that reflect unconditional functional connectivity are still widely used in neuroscience (see, e.g., Marzetti et al. 2019, for a review), a conditional independence measure or test will be
Figure 3. Functional connectivity networks of the five sensorimotor areas in the left and right hemispheres with the same color scheme in Figure 2. The adjacency matrix (a) is obtained by wavHSIC with the smoothness parameters $\beta$, selected by the method in Section 5, illustrated in the barplot (b). The black subregion in (c) corresponds to face and eye portions and the rest of the colored area corresponds to upper limbs, trunk and lower limbs portions in the right hemisphere. The adjacency matrices in (d), (e), and (f) are obtained by applying wavHSIC to low($\leq$ 4Hz)/high($>4$Hz)-pass-filtered signals with the same $\beta$ values in (b) and the same $p$-value threshold in (a).

more convincing to identify the functional connectivity between two signal curves given all others in the brain. To address this problem, there have been some advances in functional graphical models. Most of the existing methods reply on either Gaussianity (e.g., Zhu, Strawn, and Dunson 2016; Qiao, Guo, and James 2019; Zapata, Oh, and Petersen 2019; Qiao et al. 2020; Solea and Li 2020; Zhao et al. 2021) or regression models (e.g., Lundborg, Shah, and Peters 2021), while a few exceptions assume additive structures (e.g., Li and Solea 2018; Lee et al. 2021; Solea and Dette 2021). Developing a conditional independence test with these assumptions relaxed would be an interesting future research topic.

**Supplementary Material**

The supplementary material includes background materials on distance-induced characteristic kernels and Besov spaces, technical proofs of Theorems 1–4 and additional simulations.

**Acknowledgments**

The authors thank the editor, an associate editor and two referees for their constructive comments and suggestions.

**Funding**

The research of Xiaoke Zhang is partially supported by National Science Foundation grant DMS-1832046. The research of Raymond K. W. Wong is partially supported by National Science Foundation grants DMS-1806063, DMS-1711952 and CCF-1934904.

**ORCID**

Rui Miao http://orcid.org/0000-0001-5046-6341
Xiaoke Zhang http://orcid.org/0000-0002-7272-1523
Raymond K. W. Wong http://orcid.org/0000-0001-9342-3755

**References**

Abramovich, F., and Silverman, B. W. (1998), "Wavelet Decomposition Approaches to Statistical Inverse Problems," *Biometrika*, 85, 115–129. [1880]

Antoniadis, A., and Sapatinas, T. (2007), "Estimation and Inference in Functional Mixed-Effects Models," *Computational Statistics & Data Analysis*, 51, 4793–4813. [1876]

Buzsaki, G. (2006). *Rhythms of the Brain*, Oxford: Oxford University Press. [1886]

Chakraborty, S., and Zhang, X. (2019), "A New Framework for Distance and Kernel-Based Metrics in High Dimensions," arXiv:1909.13469. [1881]

Chen, F., Jiang, Q., Feng, Z., and Zhu, L. (2020), "Model Checks for Functional Linear Regression Models Based on Projected Empirical
