Statistical analysis of coupled time series with Kernel Cross-Spectral Density operators.

Michel Besserve  
MPI for Intelligent Systems, Tübingen  
michel.besserve@tuebingen.mpg.de

Nikos K. Logothetis  
MPI for Biological Cybernetics, Tübingen  
nikos.logothetis@tuebingen.mpg.de

Bernhard Schölkopf  
MPI for Intelligent Systems, Tübingen  
bs@tuebingen.mpg.de

Abstract

Many applications require the analysis of complex interactions between time series. These interactions can be non-linear and involve vector valued as well as complex data structures such as graphs or strings. Here we provide a general framework for the statistical analysis of these dependencies when random variables are sampled from stationary time-series of arbitrary objects. To achieve this goal, we study the properties of the Kernel Cross-Spectral Density (KCSD) operator induced by positive definite kernels on arbitrary input domains. This framework enables us to develop an independence test between time series, as well as a similarity measure to compare different types of coupling. The performance of our test is compared to the HSIC test using i.i.d. assumptions, showing strong improvements in terms of detection errors, as well as the suitability of this approach for testing dependency in complex dynamical systems. This similarity measure enables us to identify different types of interactions in electrophysiological neural time series.

1 Introduction

Complex dynamical systems can often be observed by monitoring time series of one or more variables. Finding and characterizing dependencies between several of these time series is key to understanding the underlying mechanisms of these systems. Whereas this problem can be addressed easily in linear systems [4], this is a more complex issue in non-linear systems. Whereas higher order statistics can provide helpful tools in specific contexts [15], and have been extensively used in system identification, causal inference and blind source separation (see for example [10, 13, 5]); it is difficult to derive a general approach with solid theoretical results accounting for a broad range of interactions. Especially, studying the relationships between time series of arbitrary objects such as texts or graphs within a general framework is largely unaddressed.

On the other hand, the dependency between independent identically distributed (i.i.d.) samples of arbitrary objects can be studied elegantly in the framework of positive definite kernels [19]. It relies on defining cross-covariance operators between variables mapped implicitly to Reproducing Kernel Hilbert Spaces (RKHS) [7]. It has been shown that when using a characteristic kernel for the mapping [9], the properties of RKHS operators are related to statistical independence between input variables and allow testing for it in a principled way with the Hilbert-Schmidt Independence Criterion (HSIC) test [11]. However, the suitability of this test relies heavily and the assumption that i.i.d. samples of random variables are used. This assumption is obviously violated in any non-trivial setting involving time series, and as a consequence trying to use HSIC in this context can
lead to incorrect conclusions. Zhang et al. established a framework in the context of Markov chains [22], showing that a structured HSIC test still provides good asymptotic properties for absolutely regular processes. However, this methodology has not been assessed extensively in empirical time series. Moreover, beyond the detection of interactions, it is important to be able to characterize the nature of the coupling between time series. It was recently suggested that generalizing the concept of cross-spectral density to Reproducible Kernel Hilbert Spaces (RKHS) could help formulate non-linear dependency measures for time series [2]. However, no statistical assessment of this measure has been established. In this paper, after recalling the concept of kernel spectral density operator, we characterize its statistical properties. In particular, we define independence tests based on this concept as well as a similarity measure to compare different types of couplings. We use these tests in section 3 to compute the statistical dependencies between simulated time series of various types of objects, as well as recordings of neural activity in the visual cortex of non-human primates. We show that our technique reliably detects complex interactions and provides a characterization of these interactions in the frequency domain.

2 Background and notations

2.1 Random variables in Reproducing Kernel Hilbert Spaces

Let $\mathcal{X}_1$ and $\mathcal{X}_2$ be two (possibly non vectorial) input domains. Let $k_1(,,) : \mathcal{X}_1 \times \mathcal{X}_1 \to \mathbb{C}$ and $k_2(,,) : \mathcal{X}_2 \times \mathcal{X}_2 \to \mathbb{C}$ be two positive definite kernels, associated to two separable Hilbert spaces of functions, $\mathcal{H}_1$ and $\mathcal{H}_2$ respectively. For $i \in \{1, 2\}$, they define a canonical mapping from $x \in \mathcal{X}_i$ to $\mathbf{x} = k_i(,,) \in \mathcal{H}_i$, such that $\forall f \in \mathcal{H}_i$, $f(x) = \langle f, \mathbf{x} \rangle_{\mathcal{H}_i}$ (see [19] for more details). In the same way, this mapping can be extended to random variables, so that the random variable $X_i \in \mathcal{X}_i$ is mapped to the random element $X_i \in \mathcal{H}_i$. Statistical objects extending the classical mean and covariance to random variables in the RKHS are defined as follows:

- the Mean Element (see [3][1]): $\mu_i = E [X_i]$,
- the Cross-covariance (see [6]): $C_{ij} = \text{Cov} [X_i, X_j] = E [X_i \otimes X_j^*] - \mu_i \otimes \mu_j^*$,

where we use the tensor product operator (see [6]) on $\mathcal{X}_i \times \mathcal{X}_j \to \mathcal{H}_i \otimes \mathcal{H}_j$. Interestingly, the link between $C_{ij}$ and covariance in the input domains is given by the Hilbert-Schmidt scalar product

$$\langle C_{ij}, f \otimes f^* \rangle_{HS} = \text{Cov} [f_i(X_i), f_j(X_j)], \quad \forall (f_i, f_j) \in \mathcal{H}_i \otimes \mathcal{H}_j$$

Moreover, the Hilbert-Schmidt norm of the operator in this space has been proved to be a measure of independence between two random variables, whenever kernels are characteristic [11]. Extension of this result has been provided in [22] for Markov chains. If the time series are assumed to be k-order Markovian, then results of the classical HSIC can be generalized for a structured HSIC using universal kernels based on the state vectors $(x_1(t),...,x_1(t+k),x_2(t),...,x_2(t+k))$. The statistical performance of this methodology has not been studied extensively, in particular its sensitivity to the dimension of the state vector. The following sections propose an alternative methodology.

2.2 Kernel Cross-Spectral Density operator

Consider a bivariate discrete time random process on $\mathcal{X}_1 \times \mathcal{X}_2 : \{X(t)\} = \{(X_1(t),X_2(t))\}_{t=-\infty,0,1,\ldots}$. We assume stationarity of the process and thus use the following translation invariant notations for the mean elements and cross-covariance operators:

$$E X_{i}(t) = \mu_{i}, \quad \text{Cov} [X_{i}(t+\tau), X_{j}(t)] = C_{ij}(\tau)$$

The cross-spectral density operator was introduced for stationary signals in [2] based on second order cumulants. Under mild assumptions, it is a Hilbert-Schmidt operator defined for all normalized frequencies $\nu \in [0:1]$ as:

$$S_{12}(\nu) = \sum_{k \in \mathbb{Z}} C_{12}(k) \exp(-k2\pi\nu) = \sum_{k \in \mathbb{Z}} C_{12}(k)z^{-k}, \quad \text{for } z = e^{2\pi i \nu}.$$
This object summarizes all the cross-spectral properties between the families of processes \( \{f(X_1)\}_{f \in \mathcal{H}_1} \) and \( \{g(X_2)\}_{g \in \mathcal{H}_2} \) in the sense that the cross-spectrum between \( f(X_1) \) and \( g(X_2) \) is given by \( S_{12}^f(g)(\nu) = \langle f, S_{12}g \rangle \). We therefore refer to this object as the Kernel Cross-Spectral Density operator (KCSD).

3 Statistical properties of KCSD

3.1 Measuring independence with the KCSD

One interesting characteristic of the KCSD is given by the following theorem [2]:

**Theorem 1.** Assume the kernels \( k_1 \) and \( k_2 \) are characteristic [9]. The processes \( X_1 \) and \( X_2 \) are pairwise independent (i.e. for all integers \( t \) and \( t' \), \( X_1(t) \) and \( X_2(t') \) are independent), if and only if \( \| S_{12}(\nu) \| = 0 \), \( \forall \nu \in [0, 1] \).

While this theorem states that KCSD can be used to test pairwise independence between time series, it does not imply independence between arbitrary sets of random variables taken from each time series in general. However, if the joint probability distribution of the time series is encoded by a Directed Acyclic Graph (DAG), the following Theorem shows that independence in this broader sense is achieved under mild assumptions.

**Proposition 2.** If the joint probability distribution is encoded by a DAG, under the Markov property and faithfulness assumption, pairwise independence between time series implies the mutual independence relationship \( \{ X_1(t) \}_{t \in \mathbb{Z}} \perp \perp \{ X_2(t) \}_{t \in \mathbb{Z}} \).

**Proof.** The proof uses the fact that the faithfulness and Markov property assumptions provide an equivalence between the independence of two sets of random variables and the d-separation of the corresponding sets of nodes in the DAG (see [17]).

For arbitrary times \( t \) and \( t' \), assume the DAG contains an arrow linking the nodes \( X_1(t) \) and \( X_2(t') \). This is an unblocked path linking this two nodes; thus they are not d-separated. As a consequence of faithfulness, \( X_1(t) \) and \( X_2(t') \) are independent. Since this contradicts our initial assumptions, there cannot exist any arrow between \( X_1(t) \) and \( X_2(t') \).

Since this holds for all \( t \) and \( t' \), there is no path linking the nodes of each time series and we have \( \{ X_1(t) \}_{t \in \mathbb{Z}} \perp \perp \{ X_2(t) \}_{t \in \mathbb{Z}} \) according to the Markov property (any joint probability distribution on the nodes will factorize in two terms, one for each time series).

As a consequence, the use of KCSD to test for independence is justified under the widely used faithfulness and Markov assumptions of graphical models. As a comparison, the structured HSIC proposed in [22] is theoretically able to capture all dependencies within the range of k samples by only assuming k-order Markovian time series.

3.2 Fourth order kernel cumulant operator

Statistical properties of KCSD require assumptions regarding the higher order statistics of the time series. Analogously to covariance, higher order statistics can be generalized as operators in (tensor products of) RKHSs. An important example in our setting is the joint quadricumulant (4th order cumulant) (see [4]). We skip the general expression of this cumulant to focus on its simplified form for four centered scalar random variables:

\[
\kappa(X_1, X_2, X_3, X_4) = \mathbb{E}[X_1X_2X_3X_4] - \mathbb{E}[X_1X_2]\mathbb{E}[X_3X_4] - \mathbb{E}[X_1X_3]\mathbb{E}[X_2X_4] - \mathbb{E}[X_1X_4]\mathbb{E}[X_2X_3]
\]

This object can be generalized for the case random variables are mapped to random elements living in two RKHSs. The quadricumulant operator \( \mathcal{K}_{1234} \) is a linear operator in the Hilbert space \( L(\mathcal{H}_1 \otimes \mathcal{H}_2, \mathcal{H}_1 \otimes \mathcal{H}_2) \), such that \( \kappa(f_1(X_1), f_2(X_2), f_3(X_3), f_4(X_4)) = \langle f_1 \otimes f_2, \mathcal{K}_{1234} f_3 \otimes f_4 \rangle \), for arbitrary elements \( f_i \). The properties of this operator will be useful in the following sections due to the following lemma.

**Lemma 3.** [Property of the tensor quadricumulant] Let \( X_1^c, X_2^c, X_3^c, X_4^c \) be centered random elements in the Hilbert space \( \mathcal{H}_1 \) and \( X_2^c, X_3^c, X_4^c \) centered random elements in \( \mathcal{H}_2 \) (the centered random element is defined by \( X_i^c = X_i - \mu_i \), then

\[
\mathbb{E}\left[\langle X_1^c, X_2^c \rangle_{\mathcal{H}_1}, \langle X_3^c, X_4^c \rangle_{\mathcal{H}_2}\right] = \text{Tr} \mathcal{K}_{1234} + \langle C_{1,2,3,4} \rangle + \text{Tr} C_{1,3} \text{Tr} C_{2,4} + \langle C_{1,4}, C_{3,2} \rangle
\]
In the case of two jointly stationary time series, we define the translation invariant quadricumulant between the two stationary time series as:

\[ \mathcal{K}_{12}(\tau_1, \tau_2, \tau_3) = \mathcal{K}_{1234}(X_1(t + \tau_1), X_2(t + \tau_2), X_1(t + \tau_3), X_2(t)) \]

### 3.3 Estimation with the Kernel Periodogram

In the following, we address the problem of estimating the properties of cross-spectral density operators from finite samples. The idea for doing this analytically is to select samples from a time-series with a tapering window function \( w : \mathbb{R} \rightarrow \mathbb{R} \) with a support is included in \([0,1]\). By scaling this window according to \( w_T(k) = w(k/T) \), and multiplying it with the time series, \( T \) samples of the sequence can be selected. The windowed periodogram estimate of the cross-spectral density operator for \( T \) successive samples of the time series is

\[
P_T^{12}(\nu) = \frac{1}{T \|w\|^2} F_T[X_1^\nu(k)](\nu) \otimes F_T[X_2^\nu](\nu)^*,
\]

where \( F_T[X_i^\nu] = \sum_{k=1}^T w_T(k)(X_i(k))z^{-k} \), for \( z = e^{2\pi i \nu} \), is the windowed Fourier transform of the delayed time series in the RKHS. Properties of the windowed Fourier transform are related to the regularity of the tapering window. In particular, we will chose a tapering window of bounded variation. In such a case, the following lemma holds (see supplemental information for the proof):

**Lemma 4.** [A property of bounded variation functions] Let \( w \) be a bounded function of bounded variation then for all \( k \), \( \sum_{t=-\infty}^{+\infty} w_T(t+k)w(t) - \sum_{t=-\infty}^{+\infty} w_T(t)^2 \leq C|k| \)

Using this assumption, the above periodogram estimate is asymptotically unbiased as shown in the following theorem

**Theorem 5.** Let \( w \) be a bounded function of bounded variation, if \( \sum_{k \in \mathbb{Z}} |k| \| \mathcal{C}_{12}(k) \|_{HS} < +\infty \), \( \sum_{k \in \mathbb{Z}} |k| \text{Tr}(\mathcal{C}_{ii}(k)) < +\infty \) and \( \sum_{(k,i,j) \in \mathbb{Z}^3} |\text{Tr}(\mathcal{K}_{12}(k,i,j))| < +\infty \),

then \( \lim_{T \to +\infty} \mathbb{E} P_T^{12}(\nu) = S_{12}(\nu), \nu \neq 0 (mod 1/2) \)

**Proof.** By definition,

\[
P_T^{12}(z) = \frac{1}{T \|w\|^2} \left( \sum_{k \in \mathbb{Z}} w_T(k)X_1^\nu(k)z^{-k} \right) \otimes \left( \sum_{n \in \mathbb{Z}} w_T(n)X_2^\nu(n)z^{-n} \right)^*
\]

\[
= \frac{1}{T \|w\|^2} \sum_{k \in \mathbb{Z}} \sum_{n \in \mathbb{Z}} z^{n-k}w_T(k)w_T(n)X_1^\nu(k) \otimes X_2^\nu(n)^*
\]

\[
= \frac{1}{T \|w\|^2} \sum_{\delta \in \mathbb{Z}} z^{-\delta} \sum_{n \in \mathbb{Z}} w_T(n + \delta)w_T(n)X_1^\nu(n + \delta) \otimes X_2^\nu(n)^*, \text{ using } \delta = k - n.
\]

Thus using Lemma 4,

\[
\mathbb{E}P_T^{12}(z) = \frac{1}{T \|w\|^2} \sum_{\delta \in \mathbb{Z}} z^{-\delta} \left( \sum_{n \in \mathbb{Z}} w_T(n)^2 + O(|\delta|) \right) \mathcal{C}_{12}(\delta)
\]

\[
= \frac{1}{T \|w\|^2} \left( \sum_{n \in \mathbb{Z}} \frac{\|w(n)^2\|}{\|w\|^2} \right) \sum_{\delta \in \mathbb{Z}} z^{-\delta} \mathcal{C}_{12}(\delta) + \frac{1}{T} O(\sum_{\delta \in \mathbb{Z}} |\delta| \| \mathcal{C}_{12}(\delta) \|_{HS}) \to_{T \to +\infty} S_{12}. \quad \square
\]

However, the squared Hilbert-Schmidt norm of \( P_T^{12}(\nu) \) is an asymptotically biased estimator of the population KCSD squared norm according to the following:

**Theorem 6.** Under the assumptions of Theorem 5 for \( \nu \neq 0 (mod 1/2) \)

\[
\lim_{T \to +\infty} \mathbb{E} \|P_T^{12}(\nu)\|^2_{HS} = \|S_{12}(\nu)\|^2_{HS} + \text{Tr}(S_{11}(\nu)) \text{Tr}(S_{22}(\nu))
\]

The proof of Theorem 5 is based on the decomposition in Lemma 4 and is provided in supplementary information.

This estimate of the KCSD norm requires specific correction methods to develop an independence test, we will call it the unbiased estimate of the KCSD norm. Having the KCSD defined in an Hilbert space also enables to define similarity between two KCSD operators, so that it is possible to compare quantitatively if different dynamical systems have similar couplings. The following theorem shows how periodograms enable to estimate the scalar product between two KCSD operators, which reflects their similarity.
Theorem 7. Assume assumptions of Theorem 6 hold for two independent samples of bivariate time series \( \{(X_1(t), X_2(t))\}_{t=-\ldots,-1,0,1,\ldots} \) and \( \{(X_3(t), X_3(t))\}_{t=-\ldots,-1,0,1,\ldots} \), mapped with the same couple of reproducing kernels.

\[
\lim_{T \to +\infty} \mathbb{E} \left( \left\langle P_{12}^T(\nu), P_{34}^T(\nu) \right\rangle_{HS} \right) = \left\langle S_{12}(\nu), S_{34}(\nu) \right\rangle_{HS}, \nu \neq 0 \mod 1/2
\]

The proof of Theorem 7 is similar to the one of Theorem 6 provided as supplemental information.

Interestingly, this estimate of the scalar product between KCSD operators is unbiased. This comes from the assumption that the two bivariate series are independent. This provides a new opportunity to estimate the Hilbert-Schmidt norm as well, in case two independent samples of the same bivariate series are available:

Corollary 8. Assume assumptions of Theorem 6 hold for the bivariate time series \( \{(X_1(t), X_2(t))\}_{t=-\ldots,-1,0,1,\ldots} \) and assume \( \{(\tilde{X}_1(t), \tilde{X}_2(t))\}_{t=-\ldots,-1,0,1,\ldots} \) an independent copy of the same time series, providing the periodogram estimates \( P_{12}^T(\nu) \) and \( \tilde{P}_{12}^T(\nu) \), respectively.

\[
\lim_{T \to +\infty} \mathbb{E} \left( \left\langle P_{12}^T(\nu), \tilde{P}_{12}^T(\nu) \right\rangle_{HS} \right) = \left\Vert S_{12}(\nu) \right\Vert_{HS}^2, \nu \neq 0 \mod 1/2
\]

In many experimental settings, such as in neuroscience, it is possible to measure the same time series on several datasets in independent trials. In such a case, corollary 8 states that estimating the Hilbert-Schmidt norm of the KCSD without bias is possible using two independent trials. We will call this estimate the unbiased estimate of the KCSD norm.

These estimates can be computed efficiently for T equispaced frequency samples using the fast Fourier transform of the centered kernel matrices of the two time series. In general, the choice of the kernel is a trade-off between the capacity to detect complex dependencies (a characteristic of the kernel being better in this respect), and the convergence rate of the estimate (simpler kernels related to lower order statistics usually require less samples). Related theoretical analysis can be found in [8, 12]. Unless otherwise stated, the Gaussian RBF kernel with bandwidth parameter \( \sigma \), \( k(x, y) = \exp(-\|x - y\|^2 / 2\sigma^2) \), will be used as a characteristic kernel for vector spaces. Let \( K_{ij} \) denote the kernel matrix between the i-th and j-th time series (such that \( (K_{ij})_{k,l} = k(x_i(k), x_j(l)) \)), and \( M \) be the centering matrix \( M = I - 1_T 1_T^T / T \), then we can define the centered kernel matrices \( \tilde{K}_{ij} = M K_{ij} M \). Defining the Discrete Fourier Transform matrix \( F \) such that \( (F)_{k,l} = \exp(-i2\pi kl/T) / \sqrt{T} \) the estimated scalar product is

\[
\left\langle P_{12}^T, \tilde{P}_{12}^T \right\rangle_{\nu = (0,1,\ldots,(T-1))/T} = T^{-2} \|w\|^{-4} \text{diag}(F \tilde{K}_{13} F^{-1}) \circ \text{diag}(F^{-1} \tilde{K}_{42} F)
\]

which can be efficiently computed using the Fast Fourier Transform (\( \circ \) is the element-wise product). The biased and unbiased norm estimates can be trivially retrieved from the above expression.

3.4 Shuffling independence tests

According to Theorem 1, pairwise independence between time series requires the cross-spectral density operator to be zero for all frequencies. We can thus test independence by testing whether the Hilbert-Schmidt norm of the operator vanishes for each frequency. We rely on theorem 6 and corollary 8 to compute biased and unbiased estimates of this norm. To achieve this, we generate a distribution of the Hilbert-Schmidt norm statistics under the null hypothesis by cutting the time interval in non-overlapping blocks and matching the blocks of each time series in pairs at random. Due to the central limit theorem, for a sufficiently large number of time windows, the empirical average of the statistics approaches a Gaussian distribution. We thus test whether the empirical mean differs from the one under the null distribution using a Student’s t-test. To prevent false positive resulting from multiple hypothesis testing, we control the Family-wise Error Rate (FWER) of the tests performed for each frequency. Following [16], we estimate a global maximum distribution on the family of t-test statistics across frequencies under the null hypothesis, and use the percentile of this distribution to assess the significance of the original t-tests.

4 Experiments

In the following, we validate the performance of our test, called kcsd, on several datasets in the biased and unbiased case. There is no general time series analysis tool in the literature to compare
with our approach on all these datasets. So our main source of comparison will be the HSIC test of independence (assuming data is i.i.d.). This enables us, to compare both approaches using the same kernels. For vector data, one can compare the performance of our approach with a linear dependency measure: we do this by implementing our test using a linear kernel (instead of an RBF kernel), and we call it *linear kscd*. Finally, for vector data too, we use the alternative approach of structured HSIC by cutting the time series in time windows (using the same approach as our independence test) and considering each of them as a single sample (using the RBF kernel). This will be called *block hsic*. The bandwidth of the HSIC methods is chosen proportional to the median norm of the sample points in the vector space. The p-value for all independence tests will be set to 5%.

4.1 Phase amplitude coupling

We first simulate a non-linear dependency between two time series by generating two oscillations at frequencies \( f_1 \) and \( f_2 \), and introducing a modulation of the amplitude of the second oscillation by the phase of the first one. This is achieved using the following discrete time equations:

\[
\begin{align*}
\phi_1(k+1) &= \phi_1(k) + 2\pi f_1 T_e \\
\phi_2(k+1) &= \phi_2(k) + 2\pi f_2 T_e \\
x_1(k) &= \cos(\phi_1(k)) \\
x_2(k) &= (2 + C \sin(\phi_1(k))) \cos(\phi_2(k))
\end{align*}
\]

Where the \( \epsilon_i \) are i.i.d normal. A simulation with \( f_1 = 4Hz \) and \( f_2 = 20Hz \) for a sampling frequency \( 1/T_e = 100Hz \) is plotted on Figure 1 (top left panel). For the parameters of the periodogram, we used a window length of 50 samples (0.5 s). We used an Gaussian RBF kernel to compute non-linear dependencies between the two time series. The top-middle and top-right panels of Figure 1 plot the mean and standard errors of the estimate of the squared Hilbert-Schmidt norm for this system for linear and an Gaussian RBF kernel (with \( \sigma = 1 \) and \( C = 0.1 \)) respectively (using \( f_1 = 4Hz \) and \( f_2 = 12Hz \) to ease visualization). The bias of the first estimate appears clearly in both cases at the two power picks of thesignals for the biased estimate. In the second (unbiased) estimate, the spectrum exhibits a zero mean for all but one peak (at 4Hz for the RBF kernel), which corresponds to the expected frequency of non-linear interaction between the time series. The observed negative values are also a direct consequence of the unbiased property of our estimate (Corollary 5). The coupling parameter \( C \) is further varied to test the performance of independence tests both in case the null hypothesis of independence is true (\( C=0 \)), and when it should be rejected (\( C = .4 \) for weak coupling, \( C = 2 \) for strong coupling). These two settings enable to quantify the type I and type II error of the tests, respectively. First, the influence of the bandwidth parameter of the kernel was studied in the case of weakly coupled time series. The bottom left and middle panels of Figure 1 show the influence of this parameter on the number of samples required to actually reject the null hypothesis and detect the dependency for biased and unbiased estimates respectively. It was observed that choosing an hyper-parameter close to the standard deviation of the signal (here 1.5) was an optimal strategy, and that the unbiased estimate outperformed to the biased estimate. We used these settings in our subsequent uses of the RBF kernel in all analysis. The bottom right panel of Figure 1 shows a comparison of these different errors for several independence tests.
superiority of our method for type II errors. In particular, methods based on HSIC fail to detect dependencies in the time series. We additionally tried to modulate the bandwidth parameter of the HSIC tests. Modification of the bandwidth could improve the type II error, but increased the type I error.

4.2 Time varying Markov chain

We now illustrate the use of our test in an hybrid setting. We generate a symbolic time series $x_2$ using the alphabet $S = \{1, 2, 3\}$, controlled by a continuous time series $x_1$. The coupling is achieved by modulating across time the transition probabilities of the Markov transition matrix generating the symbolic time series $x_2$ using the current value of the continuous time series $x_1$. This model is described by the following equations with $f_1 = 1$Hz: 
\[
\begin{aligned}
\phi_1(k+1) &= \phi_1(k) + \lambda \epsilon_1(k) + 2\pi f_1 T_e \\
x_1(k+1) &= \sin(\phi_1(k+1)) \\
p(x_2(k+1) = S_i | x_2(k) = S_j) &= M_{ij} + \Delta M_{ij} x_1(k)
\end{aligned}
\]

Since $x_1$ is bounded between -1 and 1, the Markov transition matrix fluctuates across time between two models represented Figure 2 (left panel). A model without these fluctuations ($\Delta M = 0$) was simulated as well to measure type I error. The time course of such an hybrid system is illustrated on the center panel of the same figure. In order to measure the dependency between these two time series, we use a k-spectrum kernel [14] for $x_2$ and a RBF kernel for $x_1$. For the k-spectrum kernel, we use $k=2$ (using $k=1$, i.e. counting occurrences of single symbols was less efficient) and we computed the kernel between words of 2 successive symbols of the time series. The RBF kernel used a bandwidth of 1 and we used time windows of 200 samples. The biased and unbiased estimates of the KCSD norm are represented at the bottom left of Figure 2 and show a clear peak at the modulating frequency (1Hz). The independence test results shown at the bottom right of Figure 2 emphasize that whereas HSIC can have a very low type II error, this can be at the expense of a high type I error. This phenomenon is a consequence of the temporal correlation between samples of the time series, resulting in a biased estimate of the HSIC. Conversely the type I error of KCSD norm test remains low, close to the chosen p-value of 5%.

4.3 Neural data: local field potentials from monkey visual cortex

We analyzed dependencies between local field potential (LFP) time series recorded in the primary visual cortex of one anesthetized monkey during visual stimulation by a commercial movie (see
Figure 3: Left: Experimental setup of LFP recordings in anesthetized monkey during visual stimulation with a movie. Right: Proportion of detected dependencies across channels (in percents) for the unbiased kcsd test for interactions between Gamma band and wide band LFP for different kernels.

Figure 3 (for a scheme of the experiment). LFP activity reflects the non-linear interplay between a large variety of underlying mechanisms. Here we investigate this interplay by extracting LFP activity in two frequency bands within the same electrode and quantify the non-linear interactions between them with our approach. LFPs were filtered into two frequency bands: 1/ a wide band ranging from 1 to 100Hz which contains a rich variety of rhythms and 2/ a high gamma band ranging from 60 to 100Hz which has been shown to play a role in the processing of visual information.

Both of these time series were sampled at 1000Hz. Using non-overlapping time windows of 1s points, we computed the Hilbert-Schmidt norm of the KCSD operator between gamma and large band time series originating from the same electrode. We performed statistical testing for all frequencies between 1 and 500Hz (using a Fourier transform on 2048 points). The results of the test averaged over all recording sites is plotted on Figure 3. We observe a highly reliable detection of interactions in the gamma band, using either a linear or non-linear kernel. This is due to the fact that the Gamma band LFP is a filtered version of the wide band LFP, making these signals highly correlated in the Gamma band. However, in addition to this obvious linear dependency, we observe significant interactions in the lowest frequencies (0.5-2Hz) which can not be explained by linear interaction (and is thus not detected by the linear kernel). This characteristic illustrates the non-linear interaction between the high frequency gamma rhythm and other lower frequencies of the brain electrical activity, which has been reported in other studies [21]. This also shows the interpretability of our approach as a test of non-linear dependency in the frequency domain.

5 Conclusion

An independence test for time series based on the concept of Kernel Cross Spectral Density estimation was introduced in this paper. It generalizes the linear approach based on the Fourier transform in several respects. First, it allows quantification of non-linear interactions for time series living in vector spaces. Moreover, it can measure dependencies between more complex objects, including sequences in an arbitrary alphabet, or graphs, as long as an appropriate positive definite kernel can be defined in the space of each time series. This paper provides asymptotic properties of the KCSD estimates, as well as an efficient approach to compute them on real data. The space of KCSD operators constitutes a very general framework to analyze dependencies in multivariate and highly structured dynamical systems. Following [13] [18], our independence test can further be combined to recent developments in kernel time series prediction techniques [20] to define general and reliable multivariate causal inference techniques.
References


