



Comparing spectra and coherences for groups of unequal size

Hemant Bokil^{a,*}, Keith Purpura^b, Jan-Mathijs Schoffelen^d, David Thomson^c, Partha Mitra^a

^a Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY 11724, USA

^b Weil Medical College of Cornell University, 1300 York Avenue, New York, NY 10021, USA

^c Department of Mathematics and Statistics, Queens University, Kingston, Ont. K7L3N6, Canada

^d F.C. Donders Center for Cognitive Neuroimaging, P.O. Box 9101, NL-6500 HB Nijmegen, The Netherlands

Received 11 April 2006; received in revised form 11 July 2006; accepted 12 July 2006

Abstract

Spectra and coherences are standard measures of association within and between time series. These measures have several advantages over their time-domain counterparts, not the least of which is the ability to derive and estimate confidence intervals. However, comparing spectra and coherences between two groups of observation is a problem that has not received much attention. This problem is important in neuroscience since it is often of great interest to determine whether the estimates differ between distinct experimental/behavioral conditions. Here we propose one approach to this problem. Based on the known distributional properties of spectral and coherence estimates, we derive a test for equality of two spectral or coherence estimates. The test is applicable to unequal sample sizes. We also derive jackknifed estimates of the variance of the proposed test statistic. We suggest that comparing the estimates obtained from the jackknife procedure with the theoretical estimates provides a robust means of determining whether the data in question shows non-Gaussian or non-stationary behavior. Finally, we present applications of the method to simulated and real data.

© 2006 Elsevier B.V. All rights reserved.

Keywords: Spectra; Coherences; Multi-group jackknife; Non-Gaussian fluctuations

1. Introduction

Measuring the autocorrelation of a single stochastic process, or the strength of association between two stochastic processes is a problem that frequently occurs in neuroscience. It is particularly important in understanding recordings from multiple electrodes. The frequency domain spectrum and the coherency are the fundamental measures used in a large majority of signal processing applications, and have been shown to be useful in the analysis of neural data. For example, local field potential spectra from the lateral intraparietal area of macaques have been shown to exhibit directional tuning in a memory guided saccade task (Pesaran et al., 2002). Spectra and coherences of neural activity from area V4 of macaques have been shown to be affected by the attentional state of the monkey (Fries et al., 2001; Womelsdorf et al., 2006). Spectral measures have also been found useful in the study of Parkinson's disease (Llinas et al., 1999) and in the study of birdsong (Tchernichovski et

al., 2004). Finally, spectra have also been used as the basis for a novel algorithm to predict saccadic eye movements in monkeys from recorded neural activity (Bokil et al., 2006a). A short review of some of these developments can be found in Bokil et al. (2006b).

Frequency domain measures have the particular advantage that they treat point processes (e.g., spike trains) and continuous processes (e.g., local field potential measurements) in the same way. Spectra can therefore be computed for spike trains (Jarvis and Mitra, 2001; Rosenberg et al., 1989), as well as for local field potentials (LFP) (Pesaran et al., 2002), and coherency estimates can be computed directly for pairs of spike trains (Jarvis and Mitra, 2001; Rosenberg et al., 1989), spike-LFP pairs (Pesaran et al., 2002) as well as LFP-LFP combinations. Also, the magnitudes of the complex coherency, namely the coherence, is a well normalised quantity that can be pooled across recordings. Finally, working in the frequency domain has the advantage that confidence intervals on the estimated quantities are relatively easy to construct (Brillinger, 1975).

One technical problem that has not been satisfactorily treated so far is the comparison of spectra and coherences from

* Corresponding author. Tel.: +1 516 422 4155; fax: +1 516 367 8389.
E-mail address: bokil@cshl.edu (H. Bokil).

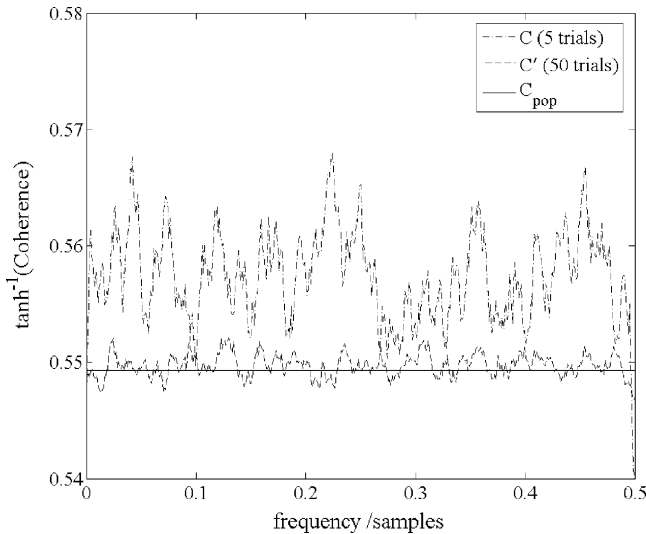


Fig. 1. Coherence for two groups of simulated Gaussian time series, showing that the bias of a coherence estimate depends on the number of trials. Fewer trials lead to larger bias. The coherences were computed using the multi-taper method with a time-bandwidth product $NW = 5$ and $K = 2NW - 1 = 9$ tapers. This corresponds to 90 and 900 degrees of freedom (see Section 2.1) for details.

two groups with unequal number of trials. Such a case may arise, for example, in testing whether the underlying population showed significant change with a change in a stimulus parameter or a behavioral state variable (such as attentional state). Since the number of trials in the two conditions may not always be the same, it is necessary to compare estimated quantities from unequally sized groups. However, spectral and coherence estimates are biased and the bias depends on size of the group. Therefore, the comparison of these quantities between the two groups is a somewhat nontrivial problem. As an example of this unequal bias, Fig. 1 shows the estimated coherence for two groups with 5 and 50 trials, respectively.

The coherences were computed by averaging over 1000 realizations of 5 and 50 pairs of Gaussian time series, respectively, each with population coherence 0.5. In contrast to the coherence estimated from 50 trials, the estimated coherence for the group with 5 trials shows substantial deviation from the population coherence.

In this paper, we propose a statistical test for the equality of two spectral or coherence estimates. In contrast to previous studies that addressed this question (Amjad et al., 1997; Brillinger, 1975, Chapter 8 in), our work explicitly addresses the issue of unequal bias in the two estimates. For Gaussian data, our test statistics are shown to be distributed as a unit normal when the two population spectra or coherences are equal. In addition, we derive jackknifed estimates of variance of this statistic based on the multi-group jackknife procedure of Arvesen (1969). The jackknife provides a robust estimate of the variance, free of distributional assumptions. Therefore, inconsistency between the jackknifed estimates and the unit normal distribution can be used as a diagnostic of non-Gaussian behavior. The utility of the method is illustrated by applications to simulated and neurobiological time series data.

2. Method

We begin our discussion with the multi-taper spectral estimation method which is our method of choice for estimating spectra and coherences. Following this, we discuss the proposed test statistics in Section 2.2 and the jackknifed estimates of the variance of the proposed statistics in Section 2.3. Finally, Section 2.4 details the procedure to test the null hypothesis (H_0) of equal spectra or coherences.

2.1. Multi-taper spectral estimation

There are a number of methods of spectral estimation (Percival and Walden, 1993), the most principled of which is the multi-taper spectral estimation method (Thomson, 1982). Given a time series $X_n, n = 1, 2, \dots, N$, the conventional estimate of the spectrum is

$$S(f) = \left| \frac{1}{N} \sum_{n=1}^N \exp(2\pi i f n) h_n X_n \right|^2 = |\tilde{X}(f)|^2, \quad (1)$$

where the sequence $h_n, n = 1, 2, \dots, N$ (called a data taper) is typically taken to be a smooth function that falls off toward the edges of the observation window, i.e. around $n = 1$ and $n = N$. Note that $\tilde{X}(f)$ is the tapered Fourier transform of the sequence X_n . In contrast to the above estimate, the multi-taper estimate uses an orthogonal family of tapers called Slepian sequences. These sequences have the property that for a given data length N , and a frequency bandwidth W (chosen by the user), the first $K = 2NW - 1$ sequences are optimally concentrated in the frequency range $[-W, W]$. Given these Slepian sequences, the simplest multi-taper estimate of the spectrum is given by

$$S^{\text{MT}}(f) = \frac{1}{K} \sum_{k=1}^K \left| \frac{1}{N} \sum_{n=1}^N \exp(2\pi i f n) u_n^k X_n \right|^2 = \frac{1}{K} \sum_{k=1}^K |\tilde{X}_k(f)|^2, \quad (2)$$

where $u_n^k, n = 1, 2, \dots, N$ is the k th Slepian sequence. Similarly, given another time series Y_n , the multi-taper estimate of the coherence between X and Y is given by

$$C_{\text{MT}}(f) = \frac{\sum_{k=1}^K \tilde{X}_k^*(f) \tilde{Y}_k(f)}{\sqrt{\left| \sum_{k=1}^K \tilde{X}_k(f) \right|^2 \left| \sum_{k=1}^K \tilde{Y}_k(f) \right|^2}}, \quad (3)$$

when dealing with multiple trials one Fourier transform is computed for each trial and each taper and Eqs. (2) and (3) are modified by replacing the averages over tapers by averages over tapers and trials. Since Slepian sequences are mutually orthogonal and the trials are interchangeable, estimates computed with the different tapers and trials are statistically independent and averaging over them reduces the variance of the estimate. Furthermore, since the Slepian sequences are optimally concentrated in frequency, multi-taper estimates also have reduced bias. Thus, the multi-taper method provides estimates with good bias-variance characteristics.

The discussion of the previous paragraph implicitly assumed that the observed data are stationary, i.e. the correlation between

observations at time t and t' depend only on the absolute time difference $|t - t'|$. Neurobiological time series are not stationary over long timescales since they reflect changes in the environment and behavior. However, there is evidence from a number of studies that stationarity holds over timescales of a few hundred milliseconds (see Bokil et al., 2006b, and references therein). The standard approach in this case is to compute spectra (and related quantities) over using data segments of a short duration. This leads to the time dependent spectrum $S(t, f)$ which is the spectrum computed from a short segment of data centered at time t , and the results of this paper could be used to determine significance of observed differences between such time dependent quantities.

Spectrum and coherence estimates, like other statistical estimates, are characterised by an appropriate degrees of freedom (DOF), which approximately count the number of independent real variables used in the estimate. For a multi-taper estimate involving K tapers and N_T trials, the DOF is given approximately by $2N_T K$. Note that the DOF are reduced near zero and the Nyquist frequency; at $f=0$ one gets half the DOF given above. The DOF is tabulated in Jarvis and Mitra (2001) for the multi-taper and other spectral estimates (such as lag window estimates), along with corrections for the DOF suitable for point process calculations. We note that while the discussion of the previous paragraph was cast in terms of sampled data, a similar approach also applies to point process data (Jarvis and Mitra, 2001; Rosenberg et al., 1989). We also note that the method discussed in this paper is applicable to any spectral and coherence estimates, not just the multi-taper ones. In the subsequent discussion we use $2m$ to denote the degrees of freedom.

2.2. Test statistics

If $S(f)$ and $C(f)$ denote a spectrum and a coherence estimate, respectively, computed with $2m$ DOF, then $\log(S(f))$ and $\tanh^{-1}(C(f))$ are approximately Gaussian with expectation and variance given by Thomson and Chave (1991).

$$\begin{aligned} E[\log(S(f))] &= \log(S_{\text{pop}}(f)) + \psi(m) - \ln(m), \\ V[\log(S(f))] &= \psi'(m), \\ E[\tanh^{-1}(C(f))] &= \tanh^{-1}(C_{\text{pop}}(f)) + \frac{1}{2m-2}, \\ V[\tanh^{-1}(C(f))] &= \frac{1}{2m-2}. \end{aligned} \quad (4)$$

Here S_{pop} denotes the population spectrum, and C_{pop} denotes the population coherence. The variances given by the above formulae are independent of the population quantities, a consequence of the fact that the log and \tanh^{-1} (Fisher) transformation are variance stabilising transformations for the spectrum and coherence for large DOF. Note that although the formulae are strictly valid for large DOF, asymptotic behavior sets in fairly quickly (DOF ≥ 10).

We now consider a second set of estimates $S'(f)$ and $C'(f)$ with $2m'$ DOF. The two sets of estimates may, for example, be the spectra of spiking activity and the coherence between the

spikes and the local field potentials in two different experimental conditions. Since the bias for the two estimates is different, it is natural to define bias corrected values.

$$\begin{aligned} x_{2m}(f) &= \log(S(f)) - \psi(m) + \ln(m), \\ y_{2m}(f) &= \tanh^{-1}(C(f)) - \frac{1}{2m-2}, \\ x_{2m'}(f) &= \log(S'(f)) - \psi(m') + \ln(m'), \\ y_{2m'}(f) &= \tanh^{-1}(C'(f)) - \frac{1}{2m'-2}. \end{aligned} \quad (5)$$

Since the log spectra and the Fisher transformed coherences are Gaussian, the quantities

$$\begin{aligned} \Delta x(f) &= \frac{x_{2m}(f) - x_{2m'}(f)}{\sqrt{\psi'(m) + \psi'(m')}}, \\ \Delta y(f) &= \frac{y_{2m}(f) - y_{2m'}(f)}{\sqrt{1/(2m-2) + 1/(2m'-2)}}, \end{aligned} \quad (6)$$

will be distributed as a unit normal viz. $\Delta x(f) \sim N(0, 1)$ and $\Delta y(f) \sim N(0, 1)$, when the two population spectra or coherences are equal. We therefore suggest that $\Delta x(f)$ and $\Delta y(f)$ be used as test statistics to test the null hypothesis (H_0) of equal population spectra or coherences, respectively. Since the proposed distributional properties of the test statistics depend on the data being Gaussian, an assumption that may not be valid in practice, we provide jackknifed estimates of the variance of the test statistic in the next section.

2.3. Multi-group jackknife

The jackknife technique was invented by Quenouille (1949) as a means to reduce bias in statistical estimates. Tukey (1958) pointed out that in addition to reducing bias, the jackknife also provides approximate confidence intervals on the estimated statistics. There are many situations where exact distributions of the statistical estimates are unknown and difficult to determine. The jackknife has therefore found wide application in many estimation problems (Miller, 1964). The basic idea behind the technique is as follows: consider independent, identically distributed observations $\{x_1, x_2, \dots, x_n\}$, and a parameter θ that is to be estimated. Let $\theta_n^0 \equiv \theta_n^0(x_1, x_2, \dots, x_n)$ be an estimate of θ using all the observations, and let $\theta_{n-1}^i \equiv \theta_{n-1}^i(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$ be an estimate formed by dropping the i th observation. Then, defining pseudovalues $\hat{\theta}_i = n\theta_n^0 - (n-1)\theta_{n-1}^i$, the jackknifed estimate of θ is then given by $\hat{\theta} = n^{-1} \sum_{i=1}^n \hat{\theta}_i$, and the jackknifed estimate of the variance of $\hat{\theta}$ is given by $(n(n-1))^{-1} \sum_{i=1}^n (\hat{\theta}_i - \hat{\theta})^2$. Note that if θ is the mean, the jackknifed estimates are identical to the sample estimates.

The jackknife does not work well in certain situations; for example, it does not work well when θ is the median (see Miller, 1964, for this and other examples). However, it has been shown to work well in many cases where the statistic is locally linear in the observations. In particular, it works for variances (Miller, 1968) and correlations (Hinkley, 1978). As discussed in the previous

section, multi-taper estimates of the spectrum with N_T trials and K tapers are based on computing $m = N_T K$ Fourier transforms, each corresponding to one trial and one taper. Similarly, the coherence is based on computing $m = N_T K$ Fourier transform pairs. Since the trials can be assumed to be interchangeable, the jackknifed estimates can be obtained by leaving out one taper of one trial in turn (Thomson and Chave, 1991). This procedure has been shown to be useful in analysis of a wide variety of time series (Thomson and Chave, 1991), including neurobiological time series (Pesaran et al., 2002).

In the two-group case, the jackknife procedure is ambiguous. We implemented (1969, Arvesen's procedure) and found that it worked well in simulations. In the subsequent discussion individual Fourier transforms or Fourier transform pairs are referred to as one observation. Given an estimate of the statistic Δz computed with all available observations, the method is based on defining single-group leave-one-out estimates Δz_{i0} and Δz_{0j} , where Δz_{i0} denotes estimates where the i th observation from the first group is left out, while the whole second group is used, and vice versa. Then, defining pseudovalues $\Delta \hat{z}_{i0}$ and $\Delta \hat{z}_{0j}$ by

$$\begin{aligned} \Delta \hat{z}_{i0} &= m \Delta z - (m - 1) \Delta z_{i0}, \\ \Delta \hat{z}_{0j} &= m' \Delta z - (m' - 1) \Delta z_{0j}, \end{aligned} \quad (7)$$

the jackknifed estimate of Δz is given by

$$\Delta \hat{z}_J = \frac{1}{m + m'} \left[\sum_{i=1}^m \Delta \hat{z}_{i0} + \sum_{j=1}^{m'} \Delta \hat{z}_{0j} \right]. \quad (8)$$

Finally, defining the means of the pseudovalues by $\overline{\Delta \hat{z}_{i0}} = m^{-1} \sum_{i=1}^m \Delta \hat{z}_{i0}$, and $\overline{\Delta \hat{z}_{0j}} = m'^{-1} \sum_{j=1}^{m'} \Delta \hat{z}_{0j}$, Arvesen (1969, Theorem 16) proved that conditions similar to those required for the corresponding results in the one sample case (Miller, 1964) lead to the following expression for the jackknifed variance of Δz .

$$\begin{aligned} \sigma_J^2 &= \frac{1}{m(m-1)} \sum_{i=1}^m [\Delta \hat{z}_{i0} - \overline{\Delta \hat{z}_{i0}}(f)]^2 \\ &+ \frac{1}{m'(m'-1)} \sum_{j=1}^{m'} [\Delta \hat{z}_{0j} - \overline{\Delta \hat{z}_{0j}}]^2. \end{aligned} \quad (9)$$

Note that one might think of replacing the statistic Δz with its jackknifed estimate $\Delta \hat{z}_J$ (Eq. (8)). However, fluctuations of jackknifed estimates of test statistics have been shown to be substantially larger than the corresponding jackknifed estimates of the variance for moderate sample sizes (Hinkley, 1978). Therefore, we adopt the approach of taking the jackknifed variance (Eq. (9)) as an estimate of the variance of the statistic Δz , and verify this approach in simulations. This approach is also adopted in applications of the single-group jackknife to spectrum and coherence computations (Thomson and Chave, 1991).

We note in passing that the estimate in Eq. (8) was criticized by Ahmad (1981), who pointed out that it exhibits a first order bias, and proposed an alternate estimate of the statistic and its variance. Ahmad's estimate for the statistic was rediscovered

more recently by Schectman and Suojin (2004). However, the issue of bias reduction is not relevant to the application discussed here, since our test statistics of Eq. (6) are explicitly unbiased. In addition, we found that the estimate of the variance given by Ahmad (1981) is systematically above 1 for simulated Gaussian data. In contrast, the variance estimated from Eq. (9) is quite close to the expected value of 1.

The discussion above has been restricted to the leave-one-out jackknifed estimates. It can however be implemented in a more general setting where, instead of leaving out one observation, the m_i ($i = 1, 2$) observations in the two samples are divided into g_i subgroups, each subgroup containing m_i/g_i observations. Then, leave-one-subgroup-out estimates, the corresponding pseudovalues, and the jackknifed estimators can be defined, using the above equations with appropriate changes in the summation limits. This procedure may be used in cases where computational efficiency is a concern.

2.4. Testing the null hypothesis

Having obtained the test statistic and the jackknifed estimates of the variance, a final issue that needs discussion is the multiple comparisons nature of this problem. Since the test statistic is a function of frequency, testing for equality of spectra or coherences involves multiple tests. Thus, there is a possibility that a certain number of the tests would be rejected simply by chance. To address this issue, we note that differences in spectra (coherences) occurring by chance are likely to be at isolated frequencies, while neurophysiological differences are likely to occur in bands of contiguous frequencies. This suggests that the null hypothesis should be rejected only when such rejection can be carried out for bands of contiguous frequencies. Now, noting that spectrum (coherence) estimates at two different frequencies are correlated when the frequencies differ by less than the bandwidth and approximately uncorrelated otherwise (for the multi-taper method the bandwidth is $2W$), we suggest the following testing procedure:

- Compute $\Delta x(f)(\Delta y(f))$ from Eq. (6) and the corresponding jackknifed estimates of the variance, $\sigma_J^2(f)$ from Eq. (9) discussed in the previous section.
- If $\sigma_J^2(f)$ is close to 1 at all frequencies, compute a $N(0, 1)$ based two-sided, $100 \times (1 - p)\%$ confidence interval $[-R(f), R(f)]$ around 0 at a p -value, p , of choice. Note that in this case, $R(f)$ is independent of frequency.
- If $\sigma_J^2(f)$ is moderately different from 1 at some frequencies, compute a $N(0, \sigma_J^2)$ based two-sided, $100 \times (1 - p)\%$ confidence band $[-R(f), R(f)]$ around 0 at a p -value, p , of choice.
- Declare the frequencies at which $|\Delta x(f)| > R(f)$ ($|\Delta y(f)| > R(f)$) as candidate frequencies for rejecting the null hypothesis.
- Reject the null hypothesis at those candidate frequencies which constitute contiguous bands whose width is larger than the bandwidth $2W$.
- If $\sigma_J^2(f)$ is substantially different from 1 at certain frequencies, this is indicative of strong non-Gaussian behavior that needs to be studied further. The test can still be used based on $N(0, \sigma_J^2)$, but its results should be used with caution.

3. Analysis of simulated data

If X is a white, Gaussian time series with unit variance, $Y = aX + \sqrt{1 - a^2}\eta$ where a is between 0 and 1, and η is Gaussian random noise with unit variance, independent of X , then the population spectrum of X and Y are both given by $S_{\text{pop}} = 1$, and population coherence between X and Y is given by $C_{\text{pop}} = a$. Thus, a measures the population coherence between X and Y .

We first verify that Δx and Δy are indeed distributed as $N(0, 1)$ when the null hypothesis is known to be true for all frequencies. We generated 1000 realizations of the pairs of time series X and Y with $N_T = 8$ trials. In addition we generated the same number of independent realizations of pairs of time series X' and Y' with varying number of trials $N'_T = N_T, 2N_T, 3N_T, 4N_T, 5N_T$. Each pair of time series was constructed according to the equations at the beginning of this paragraph with $a=0.5$. Therefore, the population spectra $S_X = S'_X = 1$ and the population coherences $C_{XY}(f) = C_{X'Y'}(f) = 0.5$ and the null hypothesis of equal population spectra and coherences are true.

For each frequency we computed the test statistics $\Delta x(f)$ and $\Delta y(f)$ for each of the 1000 realizations, and performed a Kolmogorov–Smirnov (KS) test ($p=0.05$) to determine whether the observed values of these statistics were consistent with $N(0, 1)$. Fig. 2 shows the fraction of rejected and accepted KS tests. As expected the frequencies at which the KS tests which were rejected was close to $p\%$ of the total number of frequencies (Fig. 2), and this fraction did not depend systematically on the

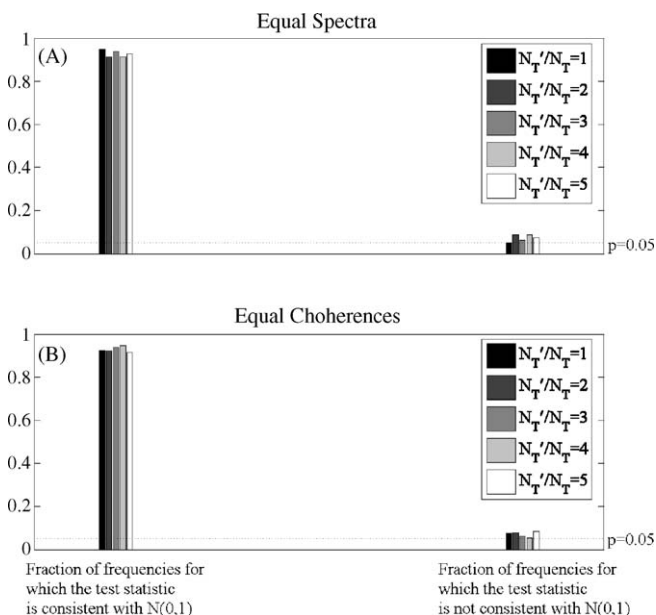


Fig. 2. Kolmogorov–Smirnov (KS) test to check whether the proposed statistics are consistent with $N(0, 1)$ when the null hypothesis is known to be true (simulated Gaussian data). (A) The fraction of frequencies at which the KS test rejected (and did not reject) the (true) null hypothesis of equal population spectrum. (B) The fraction of frequencies at which the KS test rejected (and did not reject) the (true) null hypothesis of equal population coherence. Less than 8% of the KS tests are rejected, slightly above the chance level ($p=0.05$, horizontal line). Note that there is no systematic dependence on the skewness. All quantities were computed using the multi-taper method with a time-bandwidth product $NW=5$.

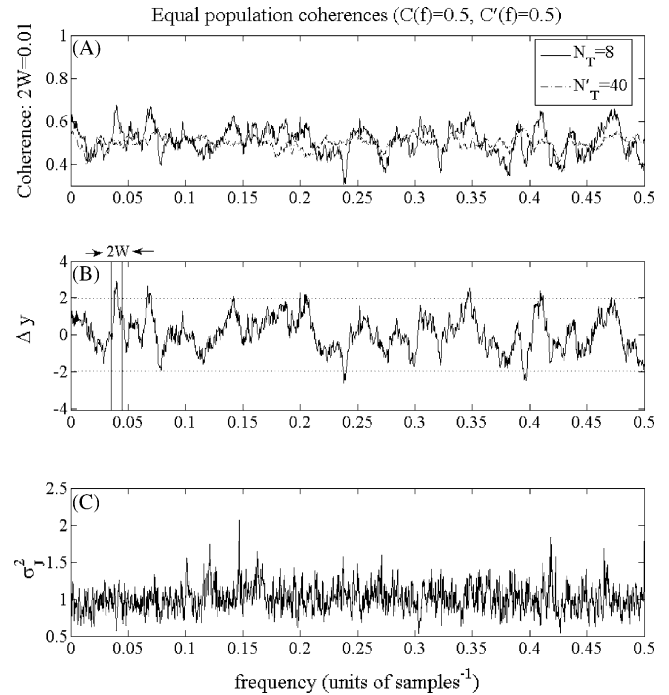


Fig. 3. Comparison of the coherence from two pairs of Gaussian time series with equal population coherence. (A) Coherence for $N_T=8$ and $N'_T=40$ trials. (B) The test statistic Δy , the $p=0.05$ confidence band (dotted lines) around zero based on $N(0, 1)$. Since the number of frequencies at which $\Delta y(f)$ is outside the confidence band is less than 5% of the total number of frequencies, we do not reject the null hypothesis anywhere. (C) The jackknifed variance is close to 1 as expected for Gaussian data.

skewness of the number of trials in the two groups. For these and other computations in this section we used the multi-taper method (Percival and Walden, 1993) with a time-bandwidth product $NW=5$. Thus, when the population spectra and coherences are known to be equal the test statistic is indeed distributed as $N(0, 1)$, with the errors occurring at the chance level.

Having verified that the test statistics are indeed distributed as expected, we now consider the case of a single realization with relatively few trials that is more akin to experimental situations. Fig. 3 shows the coherences $C_{XY}(f)$, $C_{X'Y'}(f)$, the test statistic $\Delta y(f)$, and the jackknifed estimate of the variance $\sigma_j^2(f)$ for the case of simulated data consisting of two groups, each with a small number of trials ($N_T = 8$, $N'_T = 40$) and with equal population coherence. As seen in panel B, the test statistic is outside the confidence band (based on $N(0, 1)$) only at isolated frequencies. We therefore do not reject the null hypotheses at any frequency. Note that as expected with simulated Gaussian data, the jackknifed estimates of the variance computed from Eq. (9) are close to 1 (panel C).

Finally, we consider the case where the population coherences are unequal (Fig. 4). We picked the two pairs, XY and $X'Y'$ to have unequal population coherences ($C_{XY} = 0.5$, $C_{X'Y'} = 0.7$). In this case, the test statistic is outside the confidence band based on $N(0, 1)$ almost everywhere. We therefore reject null hypotheses at all such frequencies. As in panel C of Fig. 3 the jackknifed estimates of the variance computed from Eq. (9) are close to 1 (Fig. 4C). Note also that the variances for this case,

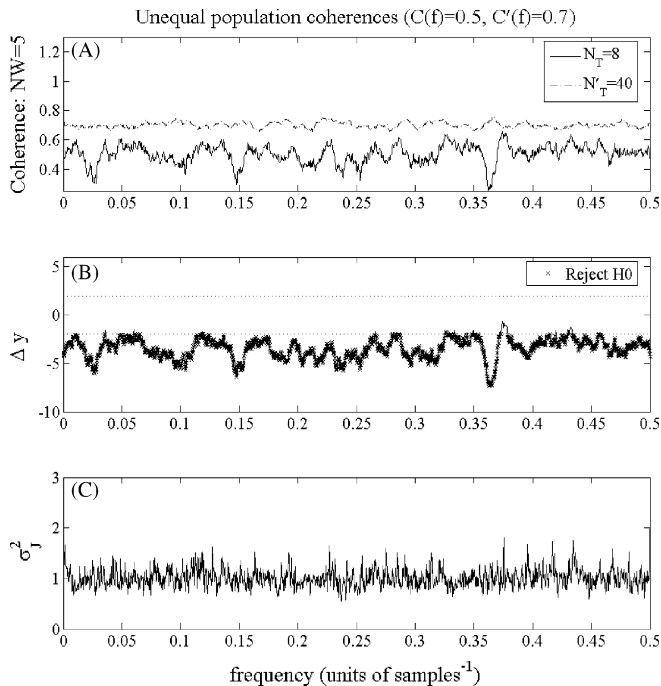


Fig. 4. Comparison of the coherence from two pairs of Gaussian time series with unequal population coherences. (A) Coherence for $N_T=5$ and $N'_T=50$ trials. (B) The test statistic Δy and the $p=0.05$ confidence band (dotted lines) around zero based on $N(0, 1)$. The null hypothesis of equal population coherences at all frequencies for which $\Delta y(f)$ is outside the confidence bands (marked with crosses) since the fraction of such frequencies is substantially larger than 0.05. (C) The jackknifed variance is close to 1 as expected for Gaussian data.

as well as the previous one, do show deviations from 1 reaching values as high as 2 at certain frequencies. However, this occurs only at isolated frequencies, and indicates the normal variability in the jackknifed estimate of the variance.

4. Analysis of experimental data

We now apply the method discussed here to two neurophysiological datasets. The first set consists of simultaneous Magnetoencephalography (MEG, acquired over the left motor cortex) and bipolar surface Electromyographic (EMG) recordings of a human subject who periodically extended his right wrist for intermittent periods of 8 s. The subject's behavior can therefore be categorized into two conditions: (i) relaxation condition, when the subject's wrist was relaxed, and (ii) isometric contraction, when the subject's wrist was extended. The second dataset consists of simultaneous local field potential recordings from the frontal lobe in and around the arcuate sulcus and in regions near the fundus of the intraparietal sulcus in the parietal lobe in an adult rhesus monkey. The monkey's eye movements were monitored using an infrared eye tracker and it was determined that there were certain epochs of the recording during which the monkey's eyes were closed. The periods of eye closure utilized here occurred during the normal course of a recording session in which the monkey was sitting comfortably in a primate chair and being rewarded for correct performance on a number of visuomotor tasks. These periods were probably coincident with transitions between drowsiness and early stage

1 sleep. At their longest, these periods were 3 min in length, but typically they lasted well under 1 min. The monkey never exhibited REM during these sessions. Thus, the two behavioral conditions of interest here are: (i) eyes closed and (ii) eyes open.

4.1. MEG-EMG recordings

To analyse the MEG-EMG recordings, we segmented the data from the contraction and relaxation conditions into epochs of 1 s duration leading to $N_T=131$ segments in the contraction condition and $N'_T=171$ segments in the relaxation condition. To verify that the method works when the null hypothesis is known to be true, we divided the contraction segments into two groups with 52 segments in one group and 79 segments in the other group, and computed the coherence between the bipolar surface-EMG, and a gradiometer overlying the left motor cortex for each group of segments in the contraction condition. These and other computations in this section were carried out using the multi-taper method with a frequency bandwidth $W=\pm 10$ Hz. The results are shown in Fig. 5. Fig. 5(B) shows that the jackknifed estimate of the variance is very close to 1, and verifies that the data is Gaussian. We therefore use $N(0, 1)$ as the basis for the test. The statistic $\Delta y(f)$ is outside the $100 \times (1-p)\%$ confidence band around $\Delta y(f)=0$ between 45 and 50 Hz ($p=0.05$). Since the bandwidth is 12 Hz, we cannot reject the null hypothesis at these frequencies.

In contrast to the two contraction condition groups discussed in the previous paragraph, the coherence for the contraction and relaxation conditions might differ. Fig. 6 shows the results of

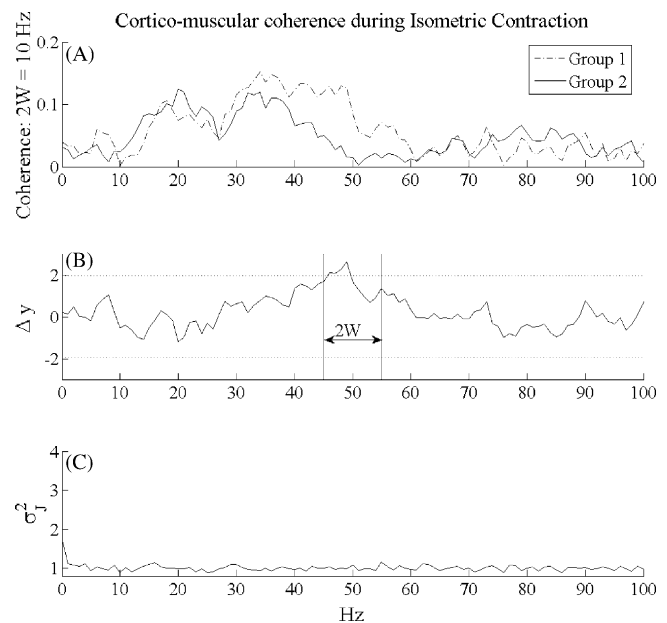


Fig. 5. Cortico-muscular coherence for two subgroups in the contraction condition. (A) Estimated coherences. (B) The jackknifed variance is close to 1 indicating Gaussian behavior, validating the use of $N(0, 1)$ to carry out the test. (C) The test statistic Δy and the $p=0.05$ confidence band (dashed lines) around zero based on $N(0, 1)$. The excursion of $\Delta y(f)$ outside the confidence band occurs for frequencies (near 50 Hz) that are within $2W$ of each other. The null hypothesis cannot be rejected anywhere.

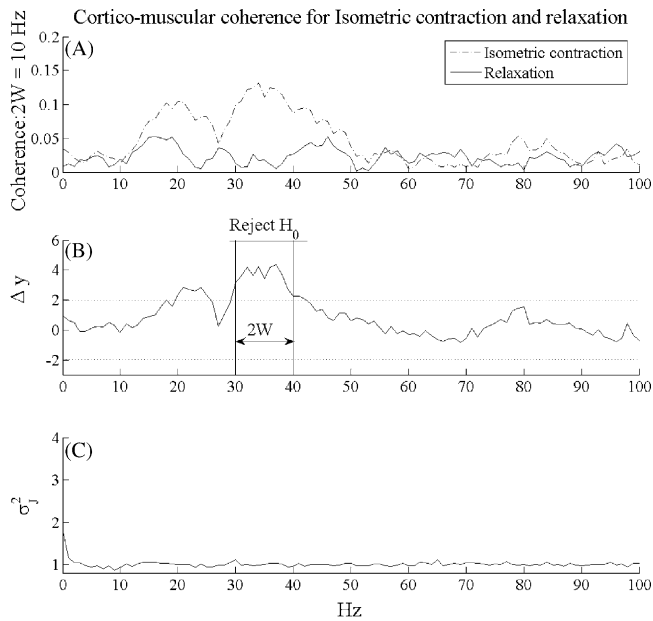


Fig. 6. Cortico-muscular coherence for contraction and relaxation conditions. (A) Estimated coherences. (B) The jackknifed variance is close to 1 indicating Gaussian behavior. (C) The test statistic Δy , the $p=0.05$ confidence band (dashed lines) around zero based on $N(0, 1)$ and the frequencies at which the null hypothesis of equal coherence was rejected (thick horizontal line).

comparing the coherence in these two conditions. As in Fig. 5, the jackknifed variance is close to 1 (Fig. 6(B)) and we used $N(0, 1)$ to carry out the test. In contrast to Fig. 5, the test statistic is now outside the confidence band for two groups of frequencies. Since the bandwidth is the same as before (12 Hz), the first group of frequencies, 20–25 Hz, is ignored. However, the null hypothesis is rejected for the second group of frequencies, between 29 and 42 Hz since the range of these frequencies is larger than the bandwidth. Note that these two bands differ only by a 3 Hz. It is therefore possible that an estimate with greater bandwidth would lead to rejection of the null hypothesis for the whole 25–40 Hz range. However, we did not investigate this issue further since the point of this exercise is simply to show the efficacy of the method for real data.

4.2. Frontal-parietal electrode recordings

For the electrophysiological data set, we segmented data acquired during the eyes closed and eyes open conditions into segments of 1 s duration, and obtained $N_T = 350$ and $N'_T = 400$ segments for the eyes closed and eyes open conditions, respectively. We then computed the spectra of the activity in each lobe and the coherence of the activity between the two lobes for the two conditions, using the multi-taper method with bandwidth $W = \pm 6$ Hz.

Fig. 7 shows the spectra from the frontal lobe for the two conditions. The spectra for the two conditions appear to differ at low (<35 Hz), with higher power in the eyes closed condition. In panel B, we show the jackknifed estimate of the variance. The jackknifed variance estimate shows moderate deviations from 1 between the 0 and 35 Hz range. Correspondingly, the confidence

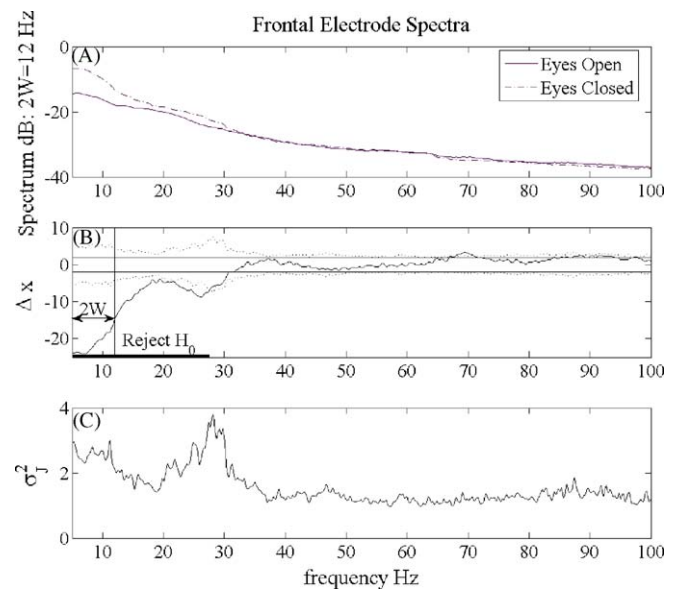


Fig. 7. Comparison of the power spectra recorded in the frontal eye fields for the eyes open and eyes closed conditions. (A) Power spectra for the eyes open (awake) and eyes closed (asleep) states. (B) The jackknifed variance differs moderately 1, indicating the presence of mild non-Gaussian behavior. (C) The statistic $\Delta x(f)$, the $p=0.05$ confidence band around zero based on $N(0, 1)$ (thin horizontal lines), and the $p=0.05$ confidence band based on $N(0, \sigma_J^2)$ (dashed lines). The statistic is outside the $N(0, \sigma_J^2)$ based confidence band from 0 to 28 Hz and the null hypothesis is rejected at these frequencies (thick horizontal line).

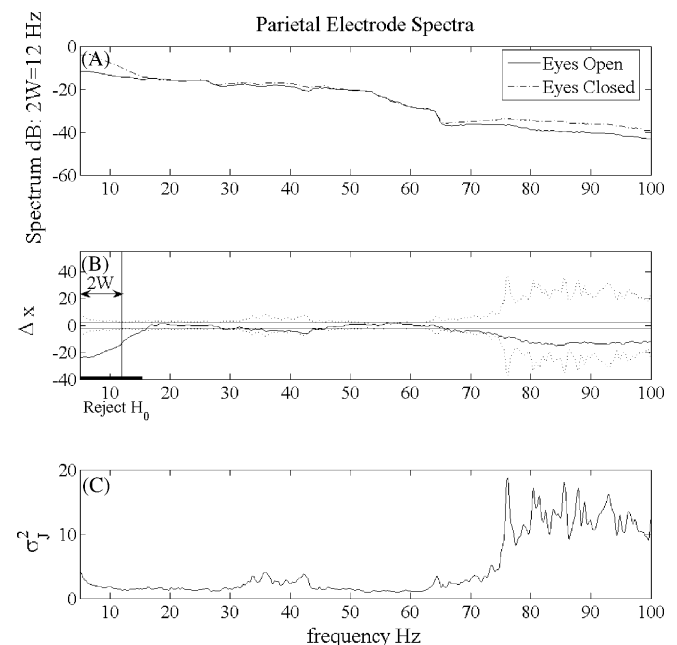


Fig. 8. Comparison of the power spectra recorded in the parietal lobe for the eyes open and eyes closed conditions. (A) Power spectra for the eyes open (awake) and eyes closed (asleep) states. (B) The jackknifed variance is substantially different from 1 for 70–100 Hz range, indicating that non-Gaussian is present to a greater extent in the parietal lobe than in the frontal lobe. (C) The statistic $\Delta x(f)$, the $p=0.05$ confidence band around zero based on $N(0, 1)$ (thin horizontal lines), and the $p=0.05$ confidence band based on $N(0, \sigma_J^2)$ (dashed lines). The statistic is outside the $N(0, \sigma_J^2)$ based confidence band from 0 to 15 Hz and the null hypothesis is rejected at these frequencies (thick horizontal line).

bands based on $N(0, 1)$ and $N(0, \sigma_j^2)$ differ at these frequencies (panel C).

As discussed previously, $N(0, \sigma_j^2)$ then provides a more conservative basis for the test of the null hypothesis of equal spectra. We therefore based the test of the null hypothesis of equal spectra on $N(0, \sigma_j^2)$. The test statistic is outside the $N(0, \sigma_j^2)$ based confidence bands in (approximately) the 0–28 Hz range (panel C). Since this range is substantially larger than the bandwidth $2W = 12$ Hz, we reject the null hypothesis of equal spectra for this range of frequencies. Thus, the spectra in the frontal lobe are different for the eyes open and eyes closed condition from 0 to 28 Hz.

Fig. 8 shows results of a similar analysis for the parietal lobe recordings. In contrast to the situation with the frontal lobe, the parietal lobe data (Fig. 8C) shows that there are large deviations from Gaussian behavior in the high gamma (>75 Hz range) along with moderate deviations at lower frequencies. As suggested previously, the test should be used with considerable caution for the higher frequencies. This is borne out in Fig. 8B, which shows that a test based on $N(0, \sigma_j^2)$ gives results that are substantially different from those based on $N(0, 1)$ —in fact, for frequencies greater than 65 Hz, the test statistic is outside the $N(0, 1)$ confidence bands, but within the $N(0, \sigma_j^2)$ confidence bands. The same is true in the 30–45 Hz range. However, for lower frequencies (<15 Hz), we can use the test based on $N(0, \sigma_j^2)$. Here, the test statistic is outside the confidence bands based on $N(0, \sigma_j^2)$

and the hypothesis of equal spectra is rejected. Thus, there are significant deviations in the parietal lobe spectra between the two conditions in the 0–15 Hz range.

Finally, Fig. 9 shows the coherence between the frontal and parietal lobes in the two conditions. The coherences appear to be different for most frequencies in the 0–65 Hz range and the jackknifed variance appears to differ moderately from 1. This contrasts with large values of the jackknifed variance for the parietal power spectra in Fig. 8 and is probably a consequence of the fact that the coherence is a normalised quantity. Since the jackknifed variance differs from 1, albeit moderately, we again use the confidence band based on $N(0, \sigma_j^2)$ as the basis for the test. The test statistic is outside this confidence band for 0–15 Hz. Since this range is greater than the bandwidth, 12 Hz, the hypothesis of equal coherences is rejected in this range.

5. Conclusion

To conclude, we have developed a method for comparing spectra and coherences from two groups with unequal number of trials. The method provides a statistical test for equality of estimated quantities in different experimental conditions, based on the assumption that the observed neural data is Gaussian. In addition, we also provide jackknifed estimates of the variance of the proposed statistic. Since the jackknife is robust, distribution-free method for estimating the variance, deviation of the jackknifed variance from the theoretical variance provides a simple test for non-Gaussian behavior. Finally, we verified that the method works for simulated Gaussian data and applied the method to two neurobiological data sets. For the first dataset, we found that the coherence between the MEG and EMG recordings in a human subject differ significantly between in the relaxation and isometric contraction conditions in the 28–43 Hz range. For the second data set, acquired using electrodes in the frontal and parietal lobes of a rhesus monkey, we found significant differences in the frontal lobe spectra (0–28 Hz), the parietal lobe spectra (0–15 Hz) and the frontal-parietal coherence (0–15 Hz) for the eyes open and eyes closed conditions.

Acknowledgement

This work is supported by the NIH (R01-MH71744 and R01-MH62528).

Appendix A

All computations in this paper were performed using programs written in MATLAB[®] (Mathworks Inc., Natick, MA, USA). These programs are now available as part of Chronux, a software package for the analysis of neuronal time series data. Chronux allows users to compute multi-taper estimates and confidence intervals of spectra and related quantities. It also allows users to perform local regression and local likelihood based density estimation. Chronux is available free for download at <http://www.chronux.org>.

For the computations of this paper, the multi-taper Fourier transforms of the relevant data were computed and these Fourier

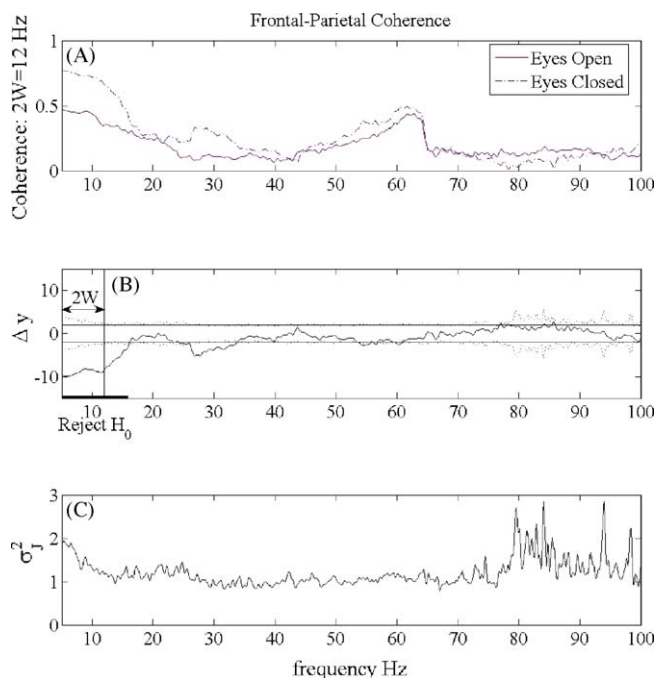


Fig. 9. Coherence between the frontal and parietal lobes for the eyes open and eyes closed conditions. (A) Coherence between the channels in the frontal and parietal lobe in the two conditions. (B) The jackknifed variance differs substantially from 1 for the lowest (<30 Hz) and highest (>80 Hz) frequencies, indicating the presence of strong non-Gaussian behavior. (C) The statistics $\Delta y(f)$, the $p = 0.05$ confidence band around zero based on $N(0, 1)$ (solid horizontal lines), and the $p = 0.05$ confidence band based on $N(0, \sigma_j^2)$ (dashed lines). The statistic is outside the $N(0, \sigma_j^2)$ based confidence band from 0 to 15 Hz and the null hypothesis is rejected at these frequencies (thick horizontal line).

transforms were used to compute the test statistics and the jackknifed variances. For example, given pre-computed Slepian sequences contained in the variable tapers,

$$J = mtfft(X, \text{tapers}, nfft, Fs) \quad (\text{A.1})$$

computes the tapered Fourier transform of a continuous valued time series X with a sampling frequency F_s . Here $nfft$ is the length of the data after padding. Then,

$$[dz, v dz] = \text{two_group_spectrum}(J_1, J_2) \quad (\text{A.2})$$

returns the test statistic $dz = \Delta x$ and the jackknifed variance, $v dz = \sigma_J^2$, where J_1 and J_2 are the multi-tapered Fourier transforms of two time series corresponding two experimental conditions. Similarly,

$$[dz, v dz] = \text{two_group_coherence}(J_{11}, J_{21}, J_{12}, J_{22}) \quad (\text{A.3})$$

computes $dz = \Delta y$ and the jackknifed variance, $v dz = \sigma_J^2$, for two pairs of time series with Fourier transforms $\{J_{11}, J_{21}\}$ and $\{J_{12}, J_{22}\}$, corresponding to two different experimental conditions.

References

- Ahmad IA. Multisample jackknife statistics. In: Proceedings of the statistical association; 1981. p. 318.
- Amjad AM, Halliday D, Rosenberg J, Conway B. An extended difference of coherence test for comparing and combining several independent coherence estimates: theory and application to the study of motor units and physiological tremor. *J Neurosci Meth* 1997;73(1):69–79.
- Arvesen J. Jackknifing u-statistics. *Ann Math Stat* 1969;40(6):2076.
- Bokil H, Pesaran B, Andersen R, Mitra P. A method for detection and classification of events in neural activity. *IEEE Trans Biomed Eng* 2006a;53(8):1678–87.
- Bokil H, Tchernichovsky O, Mitra P. Dynamic pheno-types: time series analysis techniques for characterizing neuronal and behavioral dynamics. *Neuroinformatics* 2006b;4:1.
- Brillinger DR. Time series: data analysis and theory. In: International series in decision processes. New York: Holt Rinehart and Winston; 1975.
- Fries P, Reynolds J, Rorie A, Desimone R. Modulation of oscillatory neuronal synchronization by selective visual attention. *Science* 2001;291:1560–3.
- Hinkley DV. Improving the jackknife with special reference to correlation estimation. *Biometrika* 1978;65:13–21.
- Jarvis MR, Mitra PP. Sampling properties of the spectrum and coherency of sequences of action potentials. *Neural Comput* 2001;13(4):717–49.
- Llinas RR, Ribary U, Jeanmonod D, Kronberg E, Mitra PP. Thalamocortical dysrhythmia: a neurological and neuropsychiatric syndrome characterized by magnetoencephalography. *Proc Natl Acad Sci USA* 1999;96(26):15222–7.
- Miller RGJ. A trustworthy jackknife. *The Ann Math Stat* 1964;35(4):1594–605.
- Miller RGJ. Jackknifing variances. *Ann Math Stat* 1968;39(2):567–82.
- Percival DB, Walden AT. Spectral analysis for physical applications. Cambridge: Cambridge University Press; 1993.
- Pesaran B, Pezaris JS, Sahani M, Mitra PP, Andersen RA. Temporal structure in neuronal activity during working memory in macaque parietal cortex. *Nat Neurosci* 2002;5(8):805–11.
- Quenouille M. Approximate tests of correlations in time series. *J Royal Stat Soc Ser B* 1949;11:68–84.
- Rosenberg JR, Amjad AM, Breeze P, Brillinger DR, Halliday DM. The Fourier approach to the identification of functional coupling between neuronal spike trains. *Prog Biophys Mol Biol* 1989;53(1):1–31.
- Schectman E, Suojin W. Jackknifing two-sample statistics. *J Stat Plan Inf* 2004;119:329–40.
- Tchernichovski O, Lints TJ, Deregnacourt S, Cimenser A, Mitra PP. Studying the song development process: rationale and methods. *Ann N Y Acad Sci* 2004;1016:348–63.
- Thomson DJ. Spectrum estimation and harmonic analysis. *Proc IEEE* 1982;70:1055–96.
- Thomson DJ, Chave AD. Jackknifed error estimates for spectra, coherences, and transfer functions. In: Haykin S, editor. *Advances in spectrum estimation*. Prentice Hall; 1991. p. 58–113.
- Tukey J. Bias and confidence intervals in not-quite large samples. *Ann Math Stat* 1958;29:614.
- Womelsdorf T, Fries P, Mitra P, Desimone R. Gamma-band synchronization in visual cortex predicts speed of change detection. *Nature* 2006;439:733–6.