

## Non-parametric rank statistics for spectral power and coherence

**Bahman Nasserolelami, Stefan Dukic, Teresa Buxo, Amina Coffey, Roisin McMackin, Muthuraman Muthuraman, Orla Hardiman, Madeleine M. Lowery, Edmund C. Lalor**

### Angaben zur Veröffentlichung / Publication details:

Nasserolelami, Bahman, Stefan Dukic, Teresa Buxo, Amina Coffey, Roisin McMackin, Muthuraman Muthuraman, Orla Hardiman, Madeleine M. Lowery, and Edmund C. Lalor. 2019. "Non-parametric rank statistics for spectral power and coherence." bioRxiv. Laurel Hollow, NY: Cold Spring Harbor Laboratory.  
<https://doi.org/10.1101/818906>.

# Non-Parametric Rank Statistics for Spectral Power and Coherence

Bahman Nasserolelami\*<sup>1</sup>, Stefan Dukic<sup>1</sup>, Teresa Buxo<sup>1</sup>, Amina Coffey<sup>1</sup>, Roisin McMackin<sup>1</sup>, Muthuraman Muthuraman<sup>2</sup>, Orla Hardiman<sup>1,3,4</sup>, Madeleine M. Lowery<sup>†5</sup>, and Edmund C. Lalor<sup>†3,6,7</sup>

<sup>1</sup>Academic Unit of Neurology, Trinity College Dublin, the University of Dublin, Dublin, Ireland.

<sup>2</sup>Biomedical Statistics and Multimodal Signal Processing Unit, Department of Neurology, University Medical Center of the Johannes Gutenberg-University, Mainz, Germany

<sup>3</sup>Trinity College Institute of Neuroscience, Trinity College Dublin, the University of Dublin, Dublin, Ireland.

<sup>4</sup>Beaumont Hospital, Dublin, Ireland

<sup>5</sup>School of Electrical and Electronic Engineering, University College Dublin, Dublin, Ireland

<sup>6</sup>Trinity Centre for Bioengineering, Trinity College Dublin, the University of Dublin, Dublin, Ireland.

<sup>7</sup>Departments of Biomedical engineering and Neuroscience, University of Rochester, Rochester, NY, USA.

June 11, 2018

## Abstract

Despite advances in multivariate spectral analysis of neural signals, the statistical inference of measures such as spectral power and coherence in practical and real-life scenarios remains a challenge. The non-normal distribution of the neural signals and presence of artefactual components make it difficult to use the parametric methods for robust estimation of measures or to infer the presence of specific spectral components above the chance level. Furthermore, the bias of the coherence measures and their complex statistical distributions are impediments in robust statistical comparisons between 2 different levels of coherence. Non-parametric methods based on the median of auto-/cross-spectra have shown promise for robust estimation of spectral power and coherence estimates. However, the statistical inference based on these non-parametric estimates remain to be formulated and tested. In this report a set of methods based on non-parametric rank statistics for 1-sample and 2-sample testing of spectral power and coherence is provided. The proposed methods were demonstrated and tested using simulated neural signals in different conditions. The results show that non-parametric methods provide robustness against artefactual components. Moreover, they provide new possibilities for robust 1-sample and 2-sample testing of the complex coherency function, including both the magnitude and phase, where existing methods fall short of functionality. The utility of the methods were further demonstrated by examples on experimental neural data. The proposed approach provides a new framework for non-parametric spectral analysis of digital signals. These methods are especially suited to neuroscience and neural engineering applications, given the attractive properties such as minimal assumption on distributions, statistical robustness, and the diverse testing scenarios afforded.

## 1 Introduction

The frequency domain analysis of time series, including the analysis of spectral power and the co-

herence between two (or multiple) signals (Brillinger, 2001) is pivotal to the analysis of neural signals. The spectral analyses of neuro-electric signals such as electroencephalographic (EEG) and electromyographic (EMG) signals have been specifically of interest (Rosenberg, Halliday, Breeze, & Conway, 1998; D. Halliday et al., 1995; D. M. Halliday & Rosenberg, 1999), as they reflect the intensity of neural oscillations in

\* Correspondence: Room 5.43, Trinity Biomedical Sciences Institute, 152-160 Pearse Street, Dublin D02 R590, Ireland. E-Mail: [nasserob@tcd.ie](mailto:nasserob@tcd.ie)

† Joint Last Author.

the nervous system, as well as the short and long range communications in the neural circuits through the oscillation synchrony or coherence (Siegel, Donner, & Engel, 2012). The discovery and quantification of fundamental neurophysiological phenomena such as cortico-cortical coherence (Walter, 1963; Koles & Flor-Henry, 1987), intermuscular coherence (Person & Kudina, 1968; Farmer, Bremner, Halliday, Rosenberg, & Stephens, 1993) and cortico-muscular coherence (Conway *et al.*, 1995; D. Halliday *et al.*, 1995) and subsequent contributions to the study of motor system neurophysiology (Boonstra, 2013; Wijk, M, Beek, & Daffertshofer, 2012) and pathophysiology (K. M. Fisher, Zaaimi, Williams, Baker, & Baker, 2012; Nasserolelami *et al.*, 2017) have been underpinned by spectral power and coherence analysis of the neuroelectric signals.

Despite advances in the application of spectral power and coherence analysis of neural signals in (motor) neuroscience (Nasserolelami, Lakany, & Conway, 2014; Waldert *et al.*, 2009), neural and rehabilitation engineering (Aleksandra Vuckovic *et al.*, 2014; Xu *et al.*, 2014; A. Vuckovic *et al.*, 2015) and neuro-diagnostics (Iyer *et al.*, 2015; Nasserolelami *et al.*, 2017), the statistical inference of these measures in practical and real-life scenarios remains a challenge. The high dimensional nature of EEG and EMG signals that span a few to hundreds of (spatial) channels, different frequency-bands and different time windows is a challenge that can be addressed by using contracted/abstract measures (e.g. band-specific averages, number of detections and maximum/peak value and locations), or more systematically by false discovery rate corrections (Benjamini & Hochberg, 1995; Benjamini, Krieger, & Yekutieli, 2006) and empirical Bayesian inference (Efron, Tibshirani, Storey, & Tusher, 2001; Efron, 2004, 2007). There are, however, more fundamental challenges in statistical inference of spectral measures of neuro-electric signals. First, the statistical distribution of the neural signals can deviate from normal distribution, due to the complex underlying neurophysiological phenomena (Nasserolelami *et al.*, 2014; Mehrkanoon, Breakspear, & Boonstra, 2014), non-stationarity (A. Anwar *et al.*, 2014; van Ede, Quinn, Woolrich, & Nobre, 2018), and other potentially unknown factors. Moreover, presence of various artefactual components (Tatum, Dworetzky, & Schomer, 2011) lead to extreme outliers (Nolan, Whelan, & Reilly, 2010) that lead to non-normal distributions and dramatical bias of the parametric measures of centrality. Artefact rejection methods (Mohr, Nasserolelami, Iyer, Hardiman, & Lalor, 2017; Nolan *et al.*, 2010) can reduce some (but not all) of these unwanted effects due to artefacts, but not those originating from the underlying neurophysiological complexity and the inherent non-stationarity or non-linearities. These issues are major obstacles against using the existing parametric methods (Brillinger, 2001; Bloomfield, 2004; Priestley, 1982) for inferring the presence of specific spectral

components above the chance level (1-sample testing) or presence of a significant differences between conditions or sets of signals (2-sample testing).

We have recently shown that using the median rather than mean for estimating the auto-spectra, cross-spectrum, and coherence can attenuate the effect of artefactual components in EEG connectivity analysis (Dukic *et al.*, 2017). However, the analysis of the statistical significance of non-parametric spectra and coherence measures can no longer be assessed using the traditional estimates for the expected values of spectra that rely on the mean operator (Rosenberg *et al.*, 1998; D. Halliday *et al.*, 1995; D. M. Halliday & Rosenberg, 1999). Non-parametric rank statistics (e.g. Sign Tests, Wilcoxon's Signe dRank Test, Mann-Whitney U Test, and Kruskal-Wallis Analysis of Variance) have provided promising approaches to statistical inference (Akritas, Lahiri, & Politis, 2014; Puri & Sen, 1971). However, these methods have not been translated to spectral signal analysis. In this report a new non-parametric approach for estimating the spectral power and coherence measures is explained and new methods based on non-parametric rank statistics are developed in order to assess the significance of the power and coherence, as well the differences between the levels of two power or coherence measures. We will demonstrate and test the utility of these measures through simulated signals and provide an example on the application of these measures in the context of research on neuro-electric signals.

## 2 Original Formulation for Spectra and Coherence

Consider  $x(t)$  and  $y(t)$  to be time domain signals (with  $t$  representing time), and  $x_i(t)$ ,  $y_i(t)$  representing one of the  $L$  epochs of the signals. The frequency domain representation of the signal, i.e. the (Discrete) Fourier Transform of each epoch is then represented by the complex-valued  $X_i(f)$  and  $Y_i(f)$  ( $f$  is the frequency). The auto-spectral densities  $F_{xx}$  and  $F_{yy}$ , as well as the cross-spectral density  $F_{xy}$  are defined as (Bloomfield, 2004; Brillinger, 2001):

$$F_{xx}(f) = \mathcal{E}\{X_i(f)X_i(f)^*\} \quad (1)$$

$$F_{yy}(f) = \mathcal{E}\{Y_i(f)Y_i(f)^*\} \quad (2)$$

$$F_{xy}(f) = \mathcal{E}\{X_i(f)Y_i(f)^*\} \quad (3)$$

where  $*$  denotes the complex conjugate and  $\mathcal{E}\{\cdot\}$  denotes the expected value, which is typically taken as the arithmetic mean of the values across the  $L$  epochs, i.e.

$$\bar{F}_{xx}(f) = \frac{1}{L} \sum_{i=1}^L X_i(f)X_i(f)^* \quad (4)$$

$$\bar{F}_{yy}(f) = \frac{1}{L} \sum_{i=1}^L Y_i(f)Y_i(f)^* \quad (5)$$

$$\bar{F}_{xy}(f) = \frac{1}{L} \sum_{i=1}^L X_i(f)Y_i(f)^* \quad (6)$$

The bar symbol ( $\bar{\cdot}$ ) here indicates that the expected value was found using the mean operator.

The standard deviation of auto-spectra can at frequency  $f$  be approximated using the variance  $\text{var}\{\bar{F}_{xx}(f)\} \approx (F_{xx}(f))^2/L$ . In practice, the more useful relationship is that of the signal's asymptotic variance. Therefore,  $\log_e(F_{xx}(f))$  will have the standard deviation of  $\pm L^{-\frac{1}{2}}$  (Bloomfield, 2004; D. Halliday et al., 1995). This standard deviation (or variance) value can be used to identify statistically significant oscillations. Specific spectral components are significantly different from this signal value as tested by a 1-sample t-test (or a z-test). The log-transformed parametric variances are, however, may not be stable and suitable for 2-sample testing of the levels of spectral power in 2 different conditions.

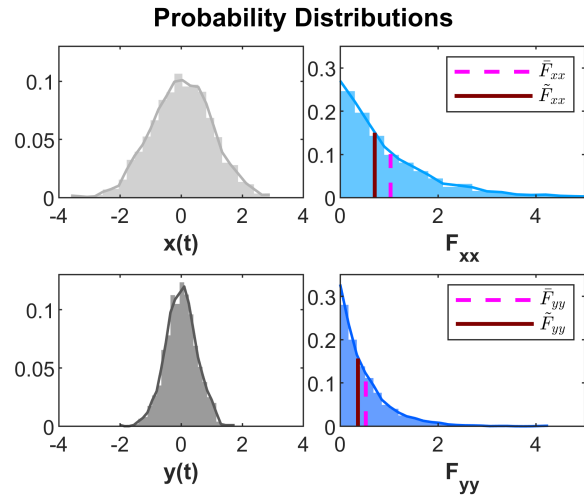
The spectral coherency function is defined as a function of the auto-spectra and cross-spectrum, and therefore given by:

$$C_{xy}(f) = \frac{F_{xy}(f)}{\sqrt{F_{xx}(f)F_{yy}(f)}} \quad (7)$$

The coherence, defined as  $|C_{xy}(f)|^2$  is a commonly used measure to assess the frequency domain correlation or phase synchrony between  $x(t)$  and  $y(t)$ . It can be rewritten and interpreted as a measure of phase synchrony across epochs that is weighted by the magnitude values (Cohen, 2014). Statistical analysis of coherence is complicated in the general form, as it has a hyper-geometric distribution (Priestley, 1982). However, for the null hypothesis (where the coherence is 0), the distribution can be very accurately approximated. In this case a  $\tanh^{-1}(\cdot)$  transforms the null coherence distribution to normal (Brillinger, 2001; Bloomfield, 2004; D. Halliday et al., 1995). Therefore for a given significance level  $\alpha$ , the one-tail threshold value for significant coherence would correspond to  $1 - \alpha^{1/(L-1)}$ , or:

$$p = (1 - |C_{xy}(f)|^2)^{(L-1)} \quad (8)$$

Equation (8) can be used for hypothesis testing against a significance threshold  $\alpha$  such as 0.05 or 0.001. This estimate is very accurate under the normal distribution assumption of the time domain signals, and is a powerful statistical test. However, the



**Figure 1. Statistical Distribution of the Time-Domain Signals, the Raw Auto-Spectra and the Estimation of Auto-Spectra using the Mean ( $\bar{F}$ ) and Median ( $\tilde{F}$ ).** The simulated data are 2000 epochs of 1-s long random normal data (white noise) at 1000Hz for  $x(t)$ .  $y(t)$  was taken as linear combination of similar random data with weight 0.2 and signal  $x$  with the weight  $0.8.e^{j\pi/4}$ . The data used in histograms were taken from 10<sup>th</sup> time points and frequency values at 10Hz..

2-sample testing of 2 different coherence values, possibly with the ability to detect changes in both amplitude and phase is not adequately established.

### 3 Non-Parametric Estimation of Spectra and Coherence

We have recently demonstrated (Dukic et al., 2017) that in order to reduce the effects of artefactual components and account for non-normal distribution of signals, we may choose to use the Median operator as an estimator for the auto- and cross-spectra:

$$\tilde{F}_{xx}(f) = \text{Median}_i[\{X_i(f)X_i(f)^*\}] \quad (9)$$

$$\tilde{F}_{yy}(f) = \text{Median}_i[\{Y_i(f)Y_i(f)^*\}] \quad (10)$$

$$\tilde{F}_{xy}(f) = \text{Median}_i[\{X_i(f)Y_i(f)^*\}] \quad (11)$$

The tilde symbol ( $\tilde{\cdot}$ ) here indicates that the expected value was found using the median operator.

Figure 1 shows the statistical distribution of 2 exemplary raw auto-spectra, as well as the auto-spectra values estimated using the traditional mean operator,  $\bar{F}_{xx}(f)$  or the new median operator,  $\tilde{F}_{xx}(f)$ . As the autospectra  $F_{xx}(f)$  and  $F_{yy}(f)$  are real-valued 1D variables, the definition on median, as the value of the inverse cumulative distribution function at 0.5,  $\text{iCDF}(0.5)$ , is straightforward.

Figure 2 shows the statistical distribution of an exemplary raw cross-spectra. The 2D distribution of the complex-valued raw cross-spectra, as well as the

marginal distributions (projected to the real and imaginary axes, respectively) are shown. The 2D cross-spectrum values estimated using the traditional mean operator,  $\bar{F}_{xy}(f)$  or the new median operator,  $\tilde{F}_{xy}(f)$  are shown on the 2D and marginal distributions.

The definition of the 2D median can be based on several (slightly) different criteria (Pranab K. Sen, 2005; Niinimaa & Oja, 2006). The simplest estimate of 2D median is the marginal median, whose components are calculated as 1D median in each dimension of the multivariate data (Niinimaa & Oja, 2006). Therefore, the complex (marginal) median is found by combining the median of the real and the median of the imaginary values (Dukic *et al.*, 2017):

$$\tilde{F}_{xy}(f) = \text{Median}_i[\{\text{real}(X_i(f)Y_i(f)^*)\}] + j \cdot \text{Median}_i[\{\text{imag}(X_i(f)Y_i(f)^*)\}] \quad (12)$$

The other approach is to use the **Spatial Median**, which is based on the minimum Euclidean distance from data points (Niinimaa & Oja, 2006).

$$\tilde{F}_{xy}(f) = \arg \min_{\Theta} \left( \sum_{i=1}^L \|X_i(f)Y_i(f)^* - \Theta\| \right) \quad (13)$$

This is a more favourable choice, given its more useful properties in 1-sample and 2-sample tests.

Following the calculation of the non-parametric auto-spectra and cross-spectrum, the coherence may be estimated similar to the parametric case:

$$\tilde{C}_{xy}(f) = \frac{\tilde{F}_{xy}(f)}{\sqrt{\tilde{F}_{xx}(f)\tilde{F}_{yy}(f)}} \quad (14)$$

This non-parametric coherency, usually described in terms of the magnitude  $|\tilde{C}_{xy}|$ , can be interpreted similar to the parametric definition  $|\bar{C}_{xy}|$ . However, both measures are biased and the values that correspond to specific significance levels depend on the number of epochs  $L$  used for calculation, as well as other potential spectral smoothing procedure. Therefore, to account for these concerns and to elevate the numerical instability of  $|\tilde{C}_{xy}|$  found by small number of epochs  $L$ , the spectral power and coherence measures are best described in terms of their p-values. This will be established in the following section.

## 4 Non-parametric Rank Statistics for Spectral Power

The statistical distribution of the raw auto-spectra  $\{|X_i(f)|^2\}$  can be used to define specific percentiles

of the distributions using the inverse cumulative distribution functions, iCDF. For example, the 2-tail 95% confidence interval with  $\alpha = 0.05$  would be:

$$[iCDF_{\{|X_i(f)|^2\}}(\alpha/2), iCDF_{\{|X_i(f)|^2\}}(1 - (\alpha/2))] \quad (15)$$

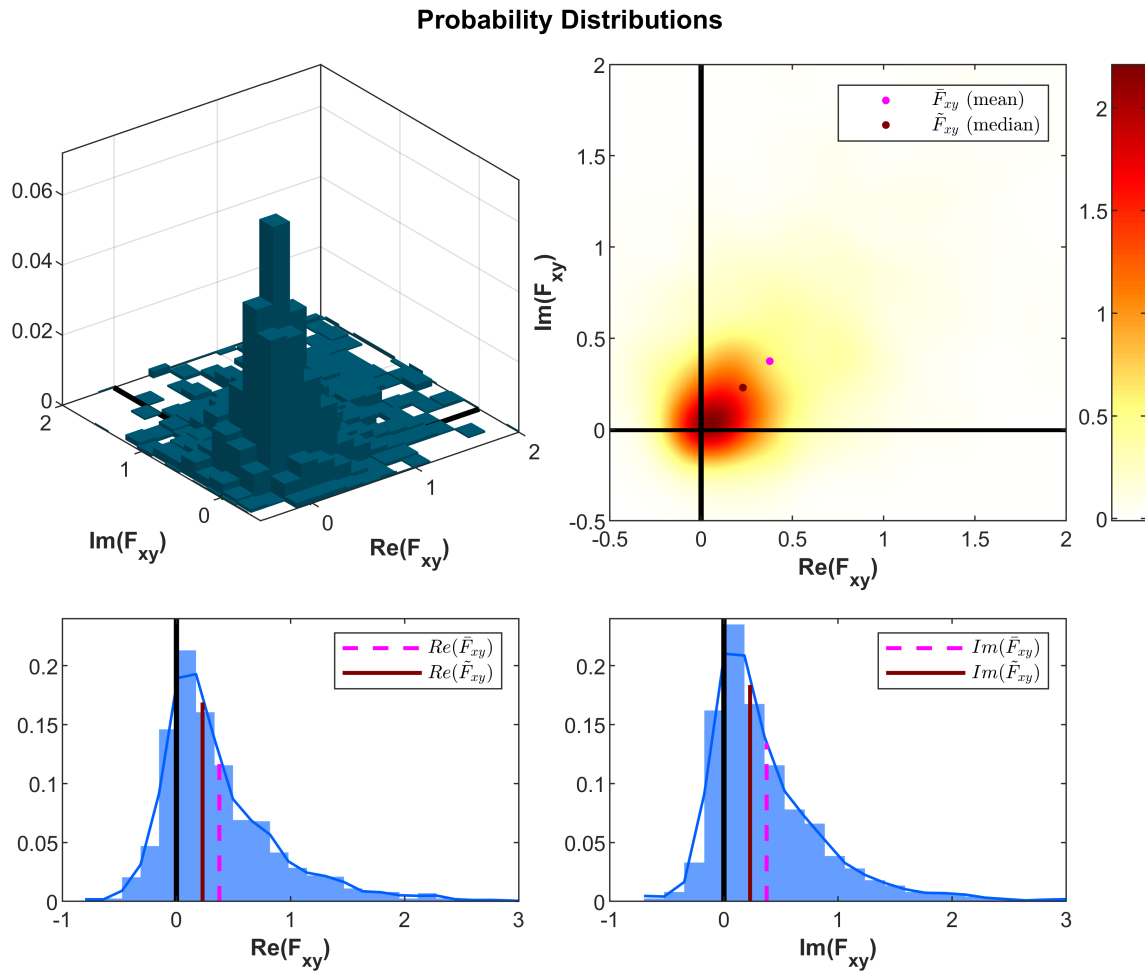
which is defined non-parametrically on the signal values at the desired frequency or band  $f$ . In case the sampling distribution (commonly needed for statistical testing) of the measures  $\bar{F}_{xx}(f)$  or  $\tilde{F}_{xx}(f)$  are required, they can be found by bootstrapping; however, this is not primarily of interest, as the statistics will be handled using non-parametric rank-based methods.

### 4.1 1-Sample Power: Testing for Significant Spectral Activity

A hypothesis that is more commonly needed to be tested is the presence of a neural oscillations (spectral power) at a specific frequency (band), beyond that of other frequencies in the null situation (e.g. a white noise). The above-mentioned confidence intervals may prove useful. However, in practical applications such as in neural signal analysis it is usually desired to establish the significant presence of activity in specific frequencies with respect to the others. The following procedure explains a method for such a statistical testing.

If  $|X_i(f_j)|^2$  shows the power pertaining to the epoch  $i$  (of total  $L$ ), and the frequency value or frequency band  $j$  (of total  $M$  selected frequency bands for analysis, excluding 0 or DC values), we may find for each epoch  $i$ , the tie-adjusted rank  $R_{xx,i}(f_j)$  among the  $M$  values which will give values in the range  $[1, M - 1]$ . We can then use the centred version of the rank  $R_{xx,i}(f_j) - (M/2)$ , and consequently apply a non-parametric test such as Wilcoxon's Signed Rank test on the  $L$  centred rank values  $R_{xx,i}(f_j) - (M/2)$ ,  $i = 1 \dots L$  which tests for higher or lower frequency content at the tested frequency  $f_j$ . This procedure detects a significant presence (or absence) of spectral or oscillatory activity under the null assumption of the white noise distribution. The test can be modified for testing against different null hypotheses (e.g. a noise with  $1/f$  distribution), by subtracting the profile of the null auto-spectrum from the  $\{|X_i(f_j)|^2\}$  data before forming the rank values. Figure 3 provides a pictorial explanation of this testing procedure for detecting a sinusoidal component mixed with a white noise. It shows a median-based estimate of auto-spectrum  $\tilde{F}_{xx}(f)$ , as well as the p-value of the spectrum associated with the oscillatory activities in the signal.

Notice that if the intention is to detect only the increase (or decrease) in spectral components a right-tail or left-tail test may be used.



**Figure 2. Statistical Distribution of the Raw Cross-Spectra in 2D and Marginal Plots and the Estimation of Cross-Spectrum using the Mean ( $\bar{F}_{xy}$ ) and Median ( $\tilde{F}_{xy}$ ).** The 2D plots (top) show the probability distribution functions as a function of the real and imaginary components and the 1D plots (bottom) show the marginal distribution of the real and imaginary components in the presence of a non-zero coherence. The frequency values correspond to the 10Hz of the  $x$  and  $y$  signal as in Figure 1.

#### 4.2 2-Sample Power: Testing for Significant Change of Power of Between Groups

Testing for between-group comparisons is straightforward as the data  $\{|X_i(f)|^2\}$  and  $\{|Y_i(f)|^2\}$  can be directly tested with Mann-Whitney U test. As the non-parametric two-sample location test can be applied on 1D data with non-normal distributions, the power values in individual epochs can be directly compared. Notice that the number of epochs for this comparison can be dissimilar.

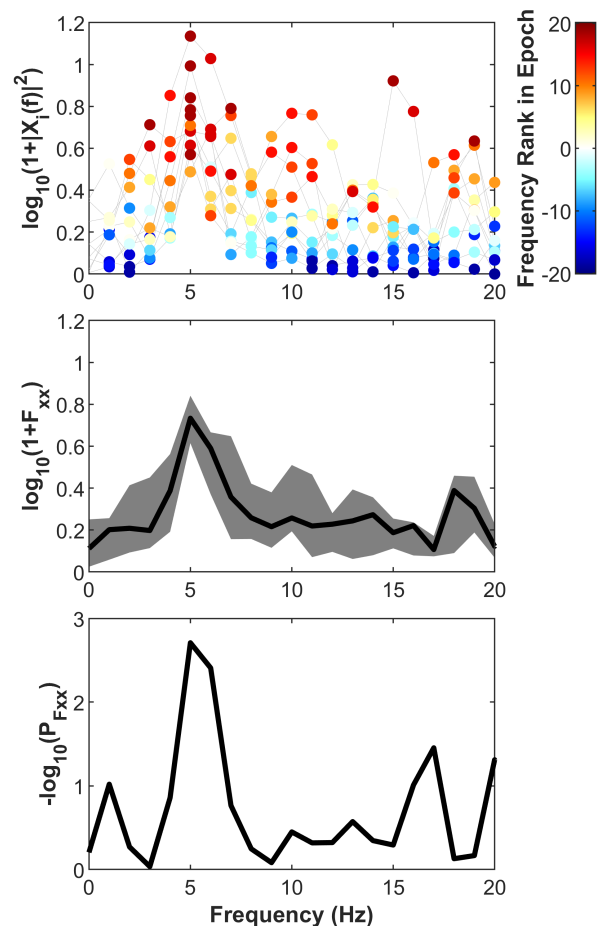
### 5 Non-parametric Rank Statistics for Coherence

The cross-spectrum  $\tilde{F}_{xy}$  or the coherence  $\tilde{C}_{xy} = \tilde{F}_{xy} / \sqrt{\tilde{F}_{xx}\tilde{F}_{yy}}$  defined based on the medians of the auto- and cross-spectra can be tested against the null hypothesis. The test for presence of significant co-

herence is a 1-sample 2D location problem where the cross-spectrum is compared against  $0 + j0$ . The test for comparing the coherence between 2 pairs of possibly coherent signals is a 2-sample 2D location problem. Each of these scenarios will be discussed as follows.

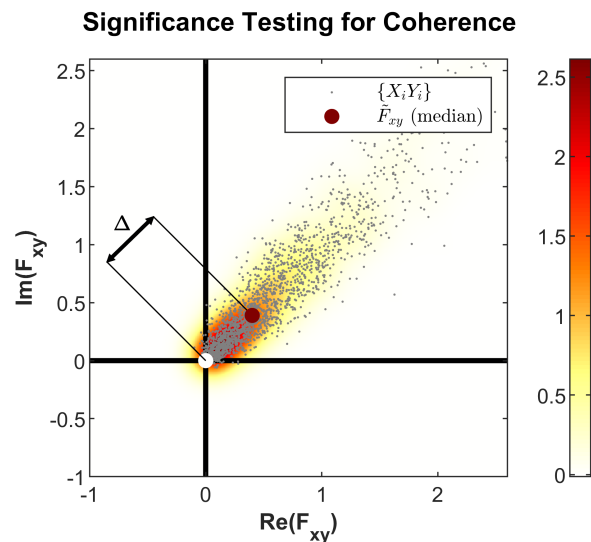
#### 5.1 1-sample Coherence: Testing for Significant Coherence

The  $L$  complex values  $\{X_i(f)Y_i(f)^*\}$  in the 2-Dimensional complex plane need to be compared to the point  $(0, 0)$ . This is a 2-dimensional 1-sample location problem (Figure 4). A parametric version of such a test is the 1-sample Hotelling's  $T^2$  test (Hotelling, 1931). However, in addition to the parametric nature of the test, the 2D distribution of cross-spectrum data is not normal and is not straightforward to convert to normal by either functional (e.g.  $\log(\cdot)$ ) (Box & Cox, 1964) or inverse normal transformations (Beasley, Erickson, & Allison, 2009). Therefore, in order to



**Figure 3. Finding the Significance of Oscillatory Power by Rank Statistics.** **Top:** Rank ordering is first applied across frequencies at each epoch. The centred ranks are subsequently tested at each frequency using Wilcoxon’s Signed Rank test. **Middle:** Median-based estimate of auto-spectrum  $\log_{10}(1 + \bar{F}_{xx})$  and the interquartile range of the raw auto-spectra **Bottom:** p-value of the spectral power against the null hypothesis of no significant presence of specific oscillatory components (white noise). The data for simulation was taken as 10 epochs of 40 sampled at 40Hz, a linear combination of white noise with the weight of 1 and a sinusoidal waveform at 5Hz with the weight of 0.85. .

follow the non-parametric route,  $\{X_i(f)Y_i(f)^*\}$  can be tested against zero using non-parametric 2-dimensional sign or rank statistics. The statistical tests based on marginal medians (Pranab Kumar Sen & Puri, 1967; Puri & Sen, 1971) are simple to use. However, they are not affine-invariant; therefore, in the context of spectral analyses they would produce variable results for the same level of coherence but with different phase differences (corresponding to a rotation transformation in the complex plane). Therefore, only affine-invariant tests are suitable for this 2D location problem. An example of these tests are the tests based on spatial median and **spatial (signed) ranks** (Hannu Oja & Randles, 2004; Hannu Oja, 2010; Nordhausen & Oja, 2011). In R (R Core



**Figure 4. Non-Parametric 1-Sample Testing for Significant Coherence.** The dots show the complex-values raw cross-spectra data points and the colour shows the 2D probability density function for the same data. The raw cross-spectra (i.e. the median-based estimate of coherence  $\tilde{C}_{xy}$ ) are compared against the zero, i.e. the origin (0, 0). This corresponds to testing for the difference between the coherence magnitude  $|\tilde{C}_{xy}|$  (or  $\Delta = |\tilde{C}_{xy}| - 0$ ) against zero. Notice that as the the division by  $\sqrt{\tilde{F}_{xx}\tilde{F}_{yy}}$  is only a scalar scaling factor in the definition of coherence in Equation (14), the comparison against zero for cross-spectrum and coherence are equivalent. The simulated data are similar to the description in Figure 1, taken at 10Hz

Team, 2016), the R-package MNM (Nordhausen & Oja, 2011; Nordhausen, Mottonen, & Oja, 2018), this test has been implemented as “*mv.1sample.test*”, a signed-rank based test with inner standardization. Notice that this 1-sample test in fact corresponds to the presence of significant magnitude. A separate non-parametric test to compare the phase of the coherence against a hypothesised value  $\phi_0$  can be achieved by applying the Wilcoxon’s Signed Rank test on the data  $\{\angle(X_i(f)Y_i(f)^*) - \phi_0\}$  (represented in the range  $[-\pi, \pi]$ ).

## 5.2 2-Sample Coherence: Testing for Significant Difference of Coherence Between Groups

In 2-sample problems, it may be desired to compare the coherence between the 2 signals  $x(t)$  and  $y(t)$  in an experimental condition 1 with  $L_1$  epochs against their coherence in experimental condition 2 with  $L_2$  (or to compare the synchrony of 2 different pairs of signals). The existing methods for comparing 2 levels of coherence values in 2 groups has limitations. One limitation is the inability to simultaneously test for a difference in magnitude and phase of coherence, with no

solid method to test for phase difference. Importantly, comparisons at individual subject levels are not established.

### 5.2.1 Comparison of Phase

The comparison of phase between the spectral data  $\{X_i(f)Y_i(f)^* | i = 1 \dots L_1\}$  (with coherence  $\tilde{C}_1(f)$ ) and  $\{X_k(f)Y_k(f)^* | k = 1 \dots L_2\}$  (with coherence  $\tilde{C}_2(f)$ ) can be achieved using the circular statistical tests. The parametric and non-parametric circular tests for comparing  $\{\angle(X_i(f)Y_i(f)^*)\}$  and  $\{\angle(X_k(f)Y_k(f)^*)\}$  would be the Watson-Williams (Watson & Williams, 1956) and Fisher's (N. I. Fisher, 1995) tests. Such tests provide valid results only under certain assumption (e.g. on the presence of a minimum circular clustering in data), therefore they may not provide reliable estimates and a uniform p-value distribution under the null condition. An alternative simple approach is to find the circular mean of the data in both samples and then find the (minimum) circular distance of the data in each group to this common circular mean. The distances are consequently compared using the 2 sample t-test (as the parametric case) with unequal variances assumption. Given that the circular data are not actually normally distributed (nor necessarily have a Von-Mises-like circular distribution), the non-parametric option to compare the phase values would be to apply the Mann-Whitney U test on the previously found distances in each group.

However, importantly, the results of such tests for 1D phase comparisons are limited, as they are not weighted by the magnitude of the raw cross-spectra and their validity is conceptually questioned when one or both of the coherence values are very low and non-significant.

### 5.2.2 Comparison of Magnitude

The comparison of the magnitude of coherence between the spectral data  $\{X_i(f)Y_i(f)^*\}$  and  $\{X_k(f)Y_k(f)^*\}$  is not straight-forward, due to the effects of phase in shaping the resultant coherence magnitudes. Consequently, only the resultant coherence values  $\tilde{C}_1(f)$  and  $\tilde{C}_2(f)$  are available for testing, but not the accompanying standard deviations or individual data points. Nevertheless, under the assumption of equality of the phases between two coherent conditions (or its lack of importance), a statistical trick may be used for a quick test of difference between the strength of coherences. First both coherence values are compared against the null condition, i.e. two 1-sample tests are performed separately in each group and the resulting p-values  $p_1$  and  $p_2$  are found by either a parametric or non-parametric test as described by Equation (8) or in Section 5.1. The p-values are then transformed to z-scores  $z_1$  and  $z_2$  by  $iCDF_{\mathcal{N}(0,1)}(\cdot)$ , subtracted from each other and then

difference Z-score is significance tested according to the resultant variance of the 2 z distributions, i.e.  $\sqrt{2}$ . In other words:

$$z_1 = iCDF_{\mathcal{N}(0,1)}(p_1) \quad (16)$$

$$z_2 = iCDF_{\mathcal{N}(0,1)}(p_2) \quad (17)$$

$$P_{diff,L} = CDF_{\mathcal{N}(0,1)}\left(\frac{z_2 - z_1}{\sqrt{2}}\right) \quad (18)$$

$$P_{diff,R} = 1 - P_{diff,L} \quad (19)$$

$$P_{diff} = 2 \cdot \min(P_{diff,L}, P_{diff,R}) \quad (20)$$

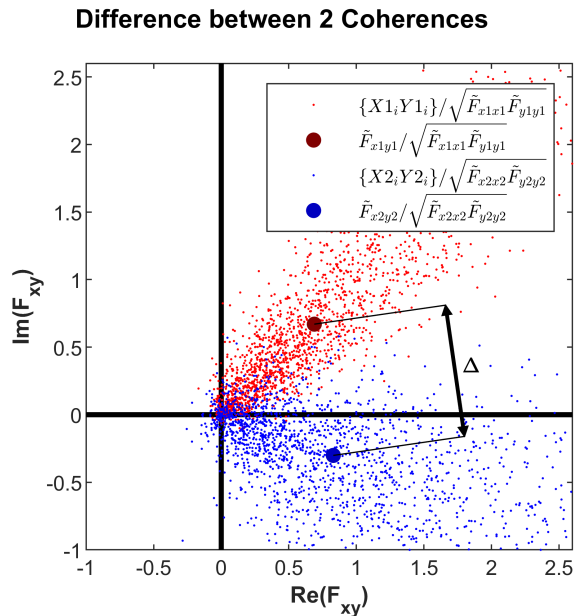
where  $\mathcal{N}(0, 1)$  denotes the normal distribution with mean 0 and standard deviation 1; (i)CDF stands for (inverse) cumulative distribution function. The  $P_{diff,L}$  and  $P_{diff,R}$  are the left- and right-tail difference p-values and  $P_{diff}$  the two tail p-value of difference between the 2 coherence levels. This procedure is similar to the Stouffer's method (Stouffer, 1977; Westfall, 2014) for combining 2 or more p-values, except that the two z-scores are subtracted (rather than summed).

### 5.2.3 Simultaneous Comparison of Magnitude and Phase

Using an approach similar to the 1-sample problem (See section 5.1), we may use 2-dimensional 2-sample location problems by considering the spectral data  $\{X_i(f)Y_i(f)^*\}$  and  $\{X_k(f)Y_k(f)^*\}$  as data points in the 2-dimensional real-complex plane (Figure 5). The difference can be tested parametrically using the 2-sample Hotelling's  $T^2$  test. This test, however, may not be preferred, given the sensitivity of test to the numerically calculated covariance matrices, which will be accentuated by the non-normal distribution of data and presence of artefacts. Similar to the 1-sample case, the 2-dimensional non-parametric tests can be used for this comparison. A simple and computationally efficient test is the Jurečková-Kalina (JK) test (Jurečková & Kalina, 2012) or two-sample tests for marginal medians (Puri & Sen, 1971). However, in order to achieve affine invariance property (as for the 1-sample case), tests based on spatial median and **spatial ranks** (Hannu Oja & Randles, 2004; Hannu Oja, 2010; Nordhausen & Oja, 2011) are preferred. The MMN R-package (Nordhausen & Oja, 2011; Nordhausen et al., 2018), has implemented this test as "*mv.Csample.test*", a rank based test with inner standardization. Notice that this 2-/multi-sample test can in fact quantify the location difference between the 2 coherence levels, which may originate from differences in the magnitude, phase or both.

## 6 Numerical Examples

Several demonstrative examples are provided using both simulated data, as well as real-life examples using experimental EEG and EMG data.



**Figure 5. Non-Parametric 2-Sample Testing for Significant Difference between 2 Coherence Values.** The dots show the complex-valued raw normalised cross-spectra data points for 2 pairs of signals  $(x_1, y_1)$  (red) and  $(x_2, y_2)$  (blue). The difference of distant between the raw normalised cross-spectra (i.e. the median-based estimates of coherence values  $\hat{C}_{x_1y_1}$  and  $\hat{C}_{x_2y_2}$ ) is compared against the zero. This corresponds to testing for the magnitude of difference between the coherence values  $\Delta = |\hat{C}_{x_2y_2} - \hat{C}_{x_1y_1}|$  against zero. The simulated data are similar to the description in Figure 1, taken at 10Hz.  $x_1$  and  $x_2$  are constructed as white noise signals including 2000 epochs of 1s at 1000Hz.  $y_1$  is built by a linear combination of white noise and  $x_1$  with a factor  $0.8 \cdot e^{j\pi/4}$ .  $y_2$  is built by a linear combination of white noise and  $x_2$  with a factor  $0.7 \cdot e^{-j\pi/8}$ .

## 6.1 Simulated Data

Before generating the demonstrating examples below, the sanity of all of the tests were verified by ensuring that under the null condition the probability density function of the p-values are uniform.

In all the simulations, epochs of normally distributed random numbers (white noise) were considered as  $x(t)$  and  $y(t)$  signals. The presence of coherence was simulated by adding a sinusoidal waveform to the data in all the epochs, which constituted a ratio  $r$  of the combined total signal. A second condition for non-normal distributions was considered, where contamination by artefactual components yield non-normal distributions (referred to here as the condition with artefacts). This latter condition was emulated by adding large white noises with non-zero baselines to 10% of the epochs. The p-values, expressed as  $-\log_{10}(p)$  were compared across the normal signal condition and the non-normal contaminated signal, when tested by both the commonly-used parametric tests, as well as the newly proposed use of non-

parametric tests. Each condition was simulated several times to afford stable estimates of the average values for  $-\log_{10}(p)$ .

### 6.1.1 1-Sample Spectral Power

Figure 6(top) shows the simulation results of a sinusoidal waveform (signal), mixed with a background white noise (noise). As the Signal-Ratio  $r$ , defined by the percentage of the amplitudes, increases from 0 to higher values, the detection of significant oscillatory component, quantified by the p-values of the parametric and non-parametric tests are shown. The  $-\log_{10}(p)$  increases as function of  $r$  for both the parametric and non-parametric tests. While the parametric test is more sensitive at higher  $r$  values, it is dramatically affected by artefacts. The non-parametric test showed a reasonable level of robustness against artefacts.

### 6.1.2 2-Sample Spectral Power

Figure 6(bottom) shows the simulation results of 2 sinusoidal waveform (signals), mixed with background white noise (noise). As the effect size  $d$ , defined by the difference of the two signal ratios  $r_1 = 0.3$  and  $r_2$ , increases from 0 to higher values, the detection of significant differences in spectral power are shown. The  $-\log_{10}(p)$  increases as function of  $d$  for both the parametric and non-parametric tests. The parametric and non-parametric tests result in very similar p-values as a function of the decrease and increase in the effect size. However, the non-parametric tests afford a considerably higher level of robustness against artefacts.

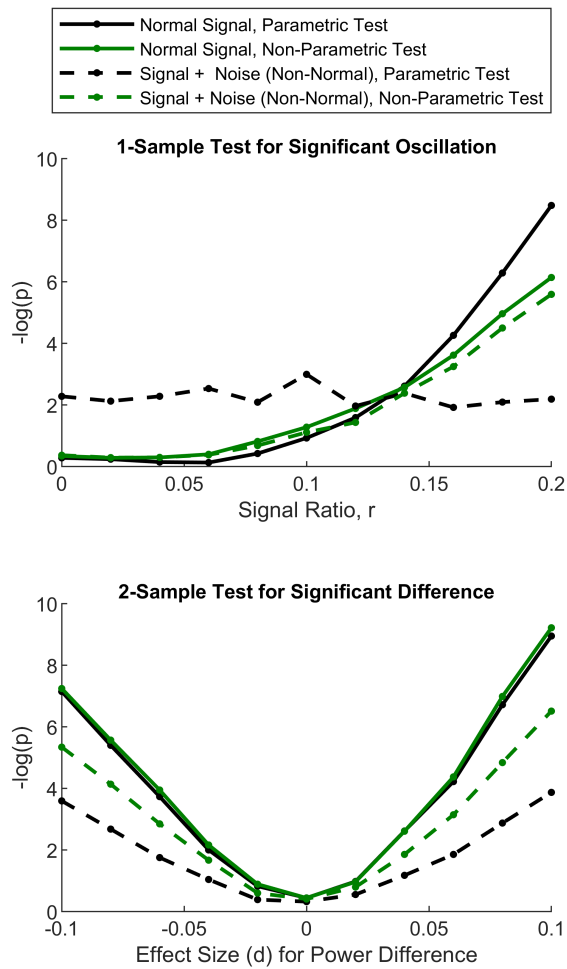
### 6.1.3 1-Sample Coherence

Figure 7 shows the simulation results of 2 signals with  $r$ -percent shared synchronous sinusoidal oscillations. As  $r$  increases from 0 to 0.5, the detection of significant coherence is shown. The  $-\log_{10}(p)$  increases as function of  $r$  for both the parametric and non-parametric tests. While the non-parametric test is more sensitive at higher  $r$  values, it is dramatically affected by artefacts. The non-parametric test showed a reasonable level of robustness against artefacts.

### 6.1.4 2-Sample Coherence

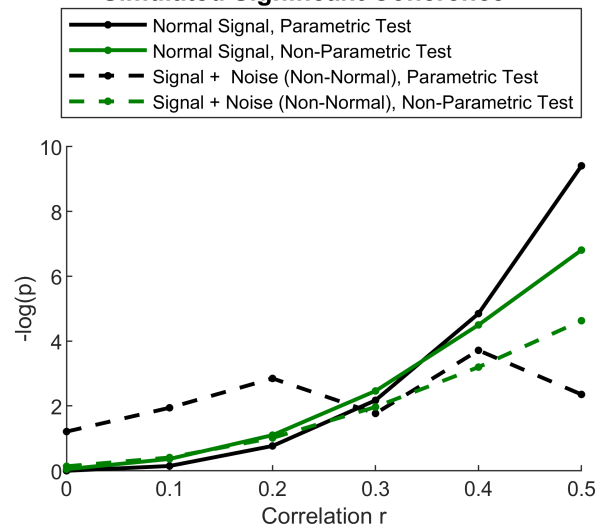
To inspect the efficient detection of differences in 2 coherence levels, including both the magnitude and phase, 2 series of simulations were carried out using 2D tests. Each set of simulations assessed the effect of differences in either magnitude or the phase of coherence. In addition to comparisons between the parametric and non-parametric tests on normal and contaminated data, the 1D tests for comparing the magnitude or phase are provided as a reference.

### Robustness of Non-Parametric Tests for Simulated Spectral Power



**Figure 6. Simulation: Non-Parametric Statistical Tests of Spectral Power are Robust Against Artefacts in Simulations of 1-Sample and 2-Sample Comparisons.** The simulated data include 50 epochs of random standard normal data (white noise) with 1s duration at 40Hz, combined linearly with a sinusoidal waveform at 10Hz with the relative signal weight  $r$ . For 1-sample test (top) the presence of the oscillatory component was detected in the signal combined with noise at different values of signal ratio  $r$ . For 2-sample tests (bottom), different levels of effect size  $d$  were tested by choosing  $r_1 = 0.3$  for  $x_1$  and varying  $r_2$  values for  $x_2$ , so that  $d = r_2 - r_1 = r_2 - 0.3$ . Artefactual components were added to 10% of epochs (randomly chosen) which were a combination of random white noise with standard deviations of 10, and a random shift between -5 and +5. All the tests were repeated 100 times and the average values are shown. The tests were performed on the values that correspond to 10Hz. See section 4.1 and Figure 3 for the details of 1-Sample non-parametric tests. For the 2-sample non-parametric tests, Mann-Whitney's U test was used. For parametric tests, the z-scores, 1-sample and 2-sample t-tests replaced the centred ranks, Wilcoxon's Signed Rank, and Mann-Whitney U tests, respectively.

### Robustness of Non-Parametric Tests for Simulated Significant Coherence



**Figure 7. Simulation: The Non-Parametric 1-Sample Statistical Test, Spatial Signed Rank, for Presence of Significant Coherence is Robust Against Artefacts.** The simulated data for  $x_1$  include 50 epochs of random standard normal data (white noise) with 1s duration at 40Hz.  $y_1$  is built by a linear combination of white noise and  $x_1$  with a factor  $r \cdot e^{j\phi}$ . Artefactual components were added to 10% of epochs (randomly chosen) which were a combination of random white noise with standard deviations of 2, and a random shift between -2 and +2. All the tests were repeated 10 times and the average values are shown. The tests were performed on the values that correspond to 10Hz. The results shown correspond to phase difference  $\phi = 0$ ; the simulations with other phase difference values yielded the same results. See section 5.1 for details on the on-parametric Spatial Signed Rank Test and equation (8) for the description of the parametric test.

**Magnitude Difference.** Figure 8 shows the simulation results of 2 pairs of signals with correlation values of  $r$  and 0.5 and the same phase values of coherence. As the difference between the correlation values changes from -0.3 to 0.3, the detection of significant difference between coherence levels is shown. The  $-\log_{10}(p)$  increases as function of difference for both the parametric and non-parametric tests. While the 2D parametric test is more sensitive than the 2D non-parametric test at higher  $r$  values, it is dramatically affected by artefacts. The 2D non-parametric test showed a reasonable level of robustness against artefacts. Similarly, the 1D parametric test is more sensitive than the 1D non-parametric test (and even the 2D parametric test) at higher  $r$  values. It is, however, considerably affected by artefacts. The 1D non-parametric test showed a reasonable level of robustness against artefacts, but does not show a major advantage over the 2D non-parametric test except at lower coherence/difference values.

**Phase Difference.** Figure 9 shows the simulation results of 2 pairs of signals with similar correlation values of 0.5, but with coherence phase values changing systematically between  $[0, \pi]$ . As the difference between the phase values changes from 0 to  $\pi$ , the detection of significant difference between coherence values is shown. The  $-\log_{10}(p)$  increases as function of difference for both the parametric and non-parametric tests. While the 2D parametric test is more sensitive than the 2D non-parametric test at higher  $r$  values, it is dramatically affected by artefacts. The 2D non-parametric test showed a reasonable level of robustness against artefacts. Both the parametric and non-parametric 1-D tests for phase differences, performed poorly against the 2D tests.

## 6.2 Experimental Data

The statistical tests described above were applied on real-life experimental data to further demonstrate their utility in practice. The experimental task was a sustained isometric pincher grip at 10% Maximum Voluntary Contraction (MVC) level. This was performed in 30 trials each lasting 5s. Experimental EEG and EMG data, including the Ear-Lobe-Referenced (ELR) EEG at electrodes  $C_3$  and  $C_4$ , as well as bipolar EMG recorded from First Dorsal Interosseous (FDI) muscle were collected during the motor task. EEG and EMG signals were recorded at 2048Hz, filtered between 0 (DC) and 410Hz and stored for analysis. Data from a healthy control subject (age=?, gender=?) was used for analysis. A total of 102 epochs each lasting 1s that were correctly recorded were checked for artefacts (Dukic *et al.*, 2017) and 99 acceptable epochs were used for analysis. Given the rejection of extreme artefacts, it is expected that both parametric and non-parametric provide comparable results.

### 6.2.1 Spectral Power

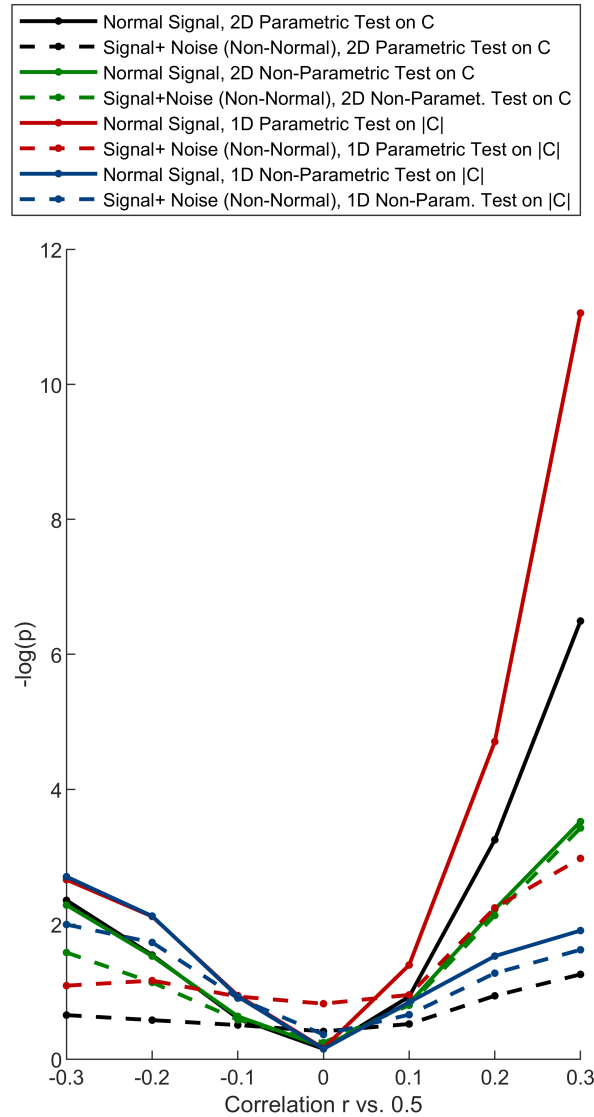
Figure 10 shows the distribution of the raw auto-spectra for an EEG signal at different confidence intervals, the autospectrum calculated using the mean operator, as well as the auto-spectrum calculated using the median operator. Notice that even after the log-transform the spectrum has a non-normal distribution, as evidenced by the difference between the mean and median, as well as the poor prediction of the 68% CI by  $\pm\sigma$  range. The p-values have similar overall behaviour, given that EEG signal is being tested against white noise as the null hypothesis. The parametric test is more effective and powerful at lower dominant frequencies, whereas the  $-\log_{10}(p)$  for non-parametric test is limited by the finite sample size and is saturated. On the other hand, non-parametric test performs better at frequencies about 400Hz, where the physiological content of the biosignals degrade, recording equipment filters part of the signal components, and the signal to noise ratio is very low.

Figure 11 compared the spectral EEG power between the 2 electrodes  $C_3$  and  $C_4$  above the contralateral and ipsilateral motor cortices of the active ongoing motor task. Both the parametric and non-parametric tests based on the raw distribution of auto-spectra prove useful in detecting low-alpha, low and high beta and low gamma changes of spectral power. The minor difference between the non-parametric and parametric tests (including stronger detections by the non-parametric test) can be due to minor difference in stationarity and the components from multiple physiological processes.

### 6.2.2 Coherence

Figure 12 shows the p-values corresponding to the traditional 1D statistical analysis of coherence using Equation (8), as well as the new 2D non-parametric test (Spatial Signed Rank). Both tests provide very comparable results. The non-parametric test provides stronger detection of synchronies in the beta and gamma band. The non-parametric detection of alpha band coherence is weaker. This can be due to the large transient alpha oscillations that are not dominant in the full distribution. The absolute value or the magnitude of coherence are both comparable with mean and median estimates. However, as the relationship between the coherence and p value is mediated by the number of epochs, there is always a bias term. This relationship is known for the parametric case, Equation (8), but is not available for the non-parametric case. The p-value spectrum therefore provides the same function and resolves these issues. The phase values corresponding to significant coherences show a range of non-trivial values using both the mean and median-based estimation, including a specific positive slope in alpha band and zero slope at high beta band.

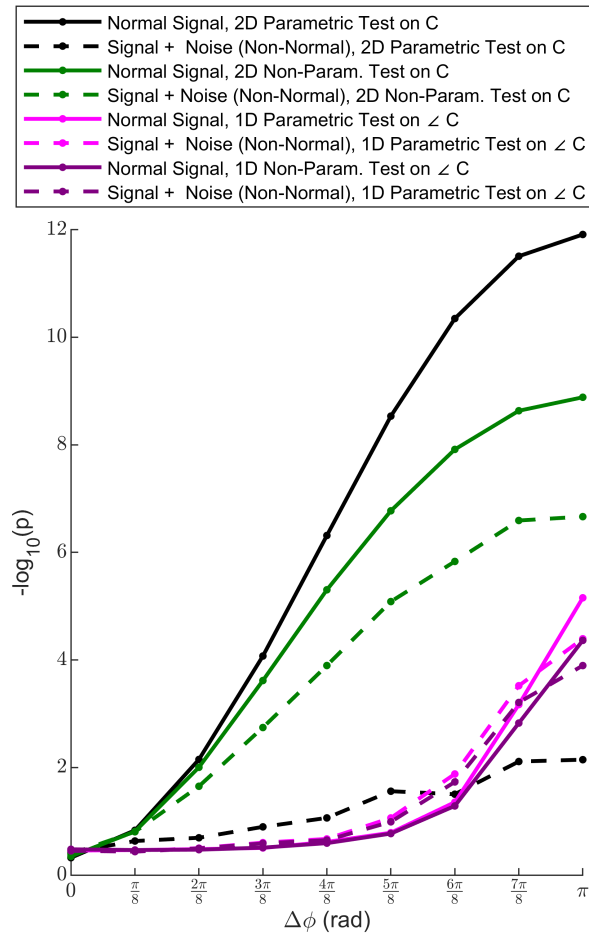
### Robustness of Non-Parametric Tests for Simulated Difference in Coherence Magnitude



**Figure 8. Simulation: The Non-Parametric 2-Sample Statistical Tests based on Spatial Rank and Spatial Signed Rank for Comparison of the Coherence Magnitude between 2 Conditions are Robust Against Artefacts.**

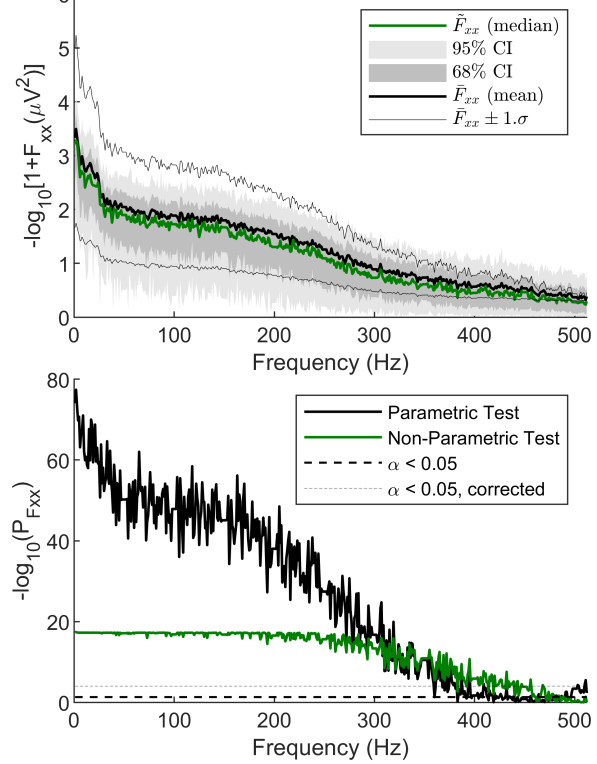
The 1D tests of difference use the statistical trick in section 5.2.2 for finding the difference based on 2 p-values (calculated parametrically from equation (8) or non-parametrically from 1-sample Spatial Signed Rank test in section 5.1). The 2D tests are 2-sample Hotelling's  $T^2$  (parametric) and the 2-sample Spatial Rank test (non-parametric). Notice that the non-parametric tests are always more robust than parametric tests, where the 2D 2-sample test performs better at higher coherence values/differences and the 1D statistical trick on 1-sample non-parametric test (Section 5.1 and Figure 7) performs better on lower coherence values/differences. The simulated data for  $x_1$  include 50 epochs of random standard normal data (white noise) with 1s duration at 500Hz with weight 0.5, combined linearly with a sine waveform at 30Hz with weight 0.05.  $y_1$  is built similarly by a linear combination of white noise and sinusoidal waveform.  $x_2$  and  $y_2$  are similarly constructed, but with the weightings of  $1 - r$  and  $0.1r$  between the white noise and sinusoidal components. Artefactual components were added to 10% of epochs (randomly chosen) which were a combination of random white noise with standard deviations of 2, and a random shift between -2 and +2. All the tests were repeated 100 times and the average values are shown. The tests were performed on the values that correspond to 30Hz. The results shown correspond to phase difference  $\phi = 0$ ; the simulations with other phase difference values yielded the same results.

### Robustness of Non-Parametric Tests for Simulated Difference in Coherence Phase

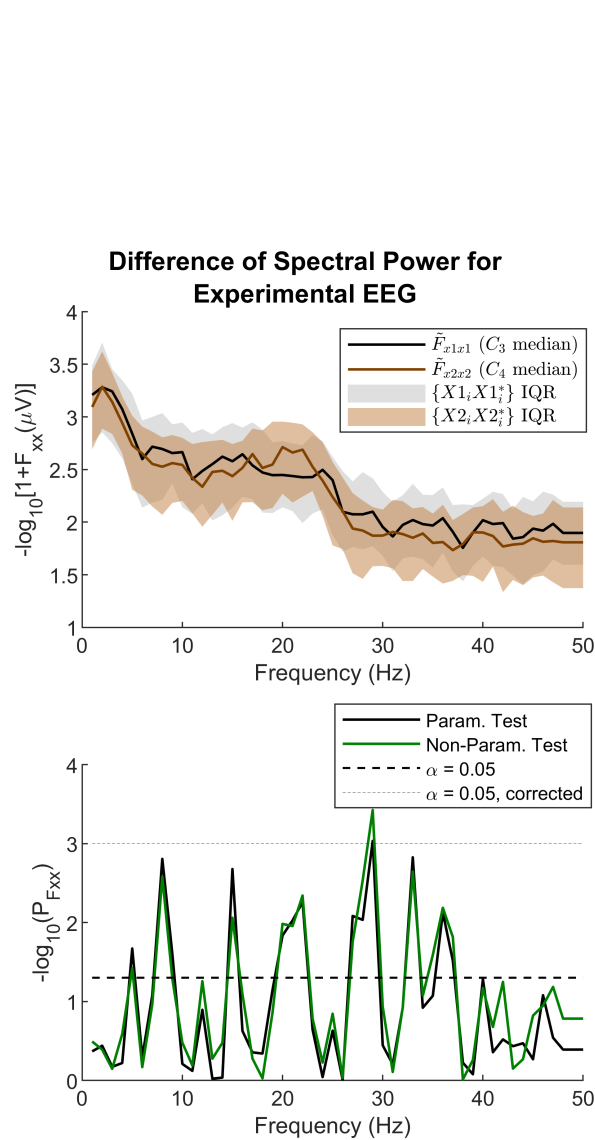


**Figure 9. Simulation: The Non-Parametric 2-Sample Statistical Tests based on Spatial Rank and Spatial Signed Rank for Comparison of the Coherence Phase between 2 Conditions are Robust Against Artefacts.** The 1D tests of difference use the circular 2 sample t-test (parametric) and circular Mann-Whitney U test (non-parametric) as described in section 5.2.1). The 2D tests are 2-sample Hotelling's  $T^2$  (parametric) and the 2-sample Spatial Rank test (non-parametric). Notice that the 2D non-parametric test performs better than 1D tests and 2D non-parametric test and is robust. The simulated data is the same as data in Figure 8, with  $r_1 = r_2 = 0.5$ , the 0 phase between the sinusoidal components of  $x_1$  and  $x_2$  and phase difference of  $\Delta\phi$  between  $x_2$  and  $y_2$ .

### Estimation of Spectral Power for Experimental EEG

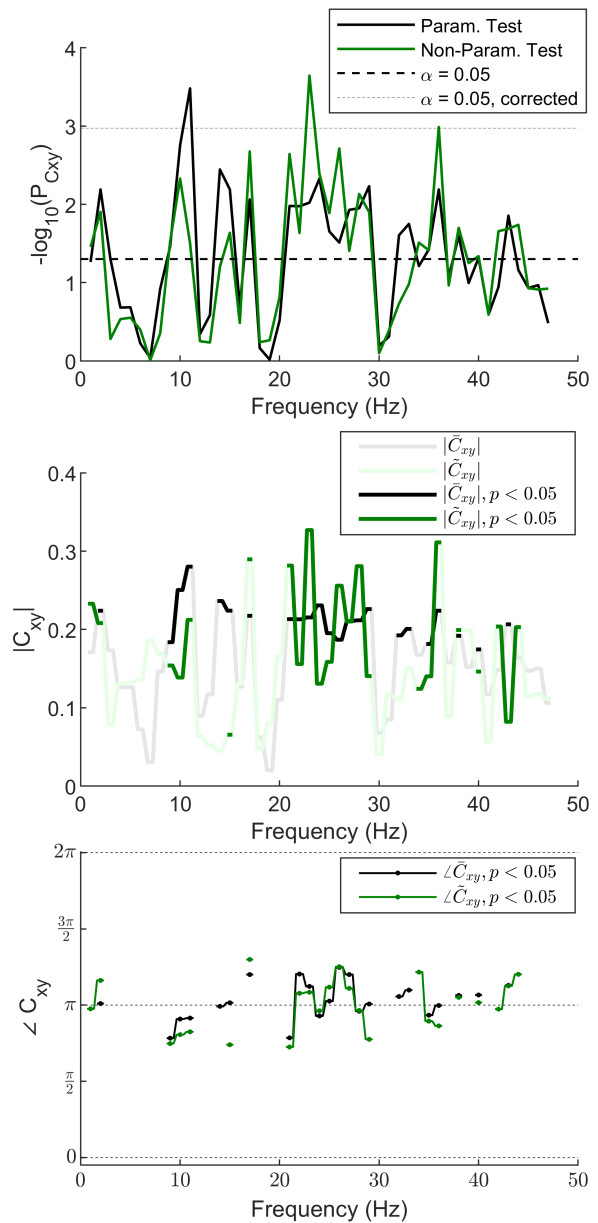


**Figure 10. Experimental Data: Both Parametric and Non-parametric Tests Detect Significant Components in Power Spectrum.** Notice that the non-normal distribution of the raw auto-spectra (top) leads to difference between the mean and median based estimates of auto-spectrum ( $\bar{F}_{xx}$  and  $\hat{F}_{xx}$ ). Both the parametric and non-parametric tests detect the significant bio-signal components (bottom) up to about 400-500Hz (the amplifier filter's cut-off frequency) with the non-parametric test saturating due to finite sample size at lower frequencies with stronger effects, while performing better at detecting weaker high-frequency components compared to its parametric counterpart. Corrected  $\alpha$  corresponds to Bonferroni correction with the number of frequencies. See section 4.1 for details of the tests and section 6.2 for details of the data used.



**Figure 11. Experimental Data: Both Parametric and Non-parametric Tests Detect Significant Differences Between 2 Power Spectra.** The 2 power spectra (top) correspond to EEG channels  $C_3$  and  $C_4$  over the left (contra-lateral) and right (ipsilateral) primary motor area of hand during an isometric pincher grip task. The p-values of both the parametric and non-parametric tests (bottom) detect the significant differences between the oscillation intensity in the 2 hemispheres. Corrected  $\alpha$  corresponds to Bonferroni correction with the number of frequencies. See section 4.2 for details of the tests and section 6.2 for details of the data used.

### Significance of Experimental Cortico-Muscular Coherence



**Figure 12. Experimental Data: Both Parametric and Non-parametric Tests Detect Significant Presence of Cortico-Muscular Coherence Between an EEG ( $C_3$ ) and an EMG (FDI) signals.** **Top:** The p-values from the 1D parametric test on  $|\bar{C}_{xy}|$  in Equation (8) and the 2D non-parametric Spatial Signed Rank test (section 5.1) show comparable detection of significant coherence. Corrected  $\alpha$  corresponds to Bonferroni correction with the number of frequencies. **Middle:** The magnitude of coherence estimated using the mean ( $|\bar{C}_{xy}|$ ) and median ( $|\tilde{C}_{xy}|$ ), follow the p-values. **Bottom:** The phase of coherence estimated using the mean ( $\angle\bar{C}_{xy}$ ) and median ( $\angle\tilde{C}_{xy}$ ) show close correspondence. See section 5.1 for details of the tests and section 6.2 for details of the data used.

Figure 13 show 2 different corticomuscular EEG-EMG coherence measures: between  $C_3$ -FDI and between  $C_4$ -FDI. The median based  $\tilde{C}_{xy}$  for both pairs show differences between these 2 channel pairs. The existing statistical techniques (parametric) do not provide direct comparisons of these measures at the level of individual subjects. The statistical trick in equation (20) provides a solution for an approximate inference between 2 pairs based on their individual p-values. This 1D comparison of magnitude (chosen to be based on  $p_1$  and  $p_2$  from non-parametric 1-sample tests) detects the changes in coherence strengths which is especially strong for low coherence values. The 2D non-parametric test based on spatial ranks detects changes in amplitude and especially the phase. A significant difference in the phase of synchrony is meaningful only if each of the signal pairs show a significant coherence in 1-sample tests, and, additionally, the 2-sample test yields significant differences. The certainty about the phase difference can be achieved if the 1D test of magnitude is non-significant, or alternatively by using a circular Mann-Whitney U-test (section 5.2.1) can confirm this.

## 7 Discussion

### 7.1 The New Perspective and Methods

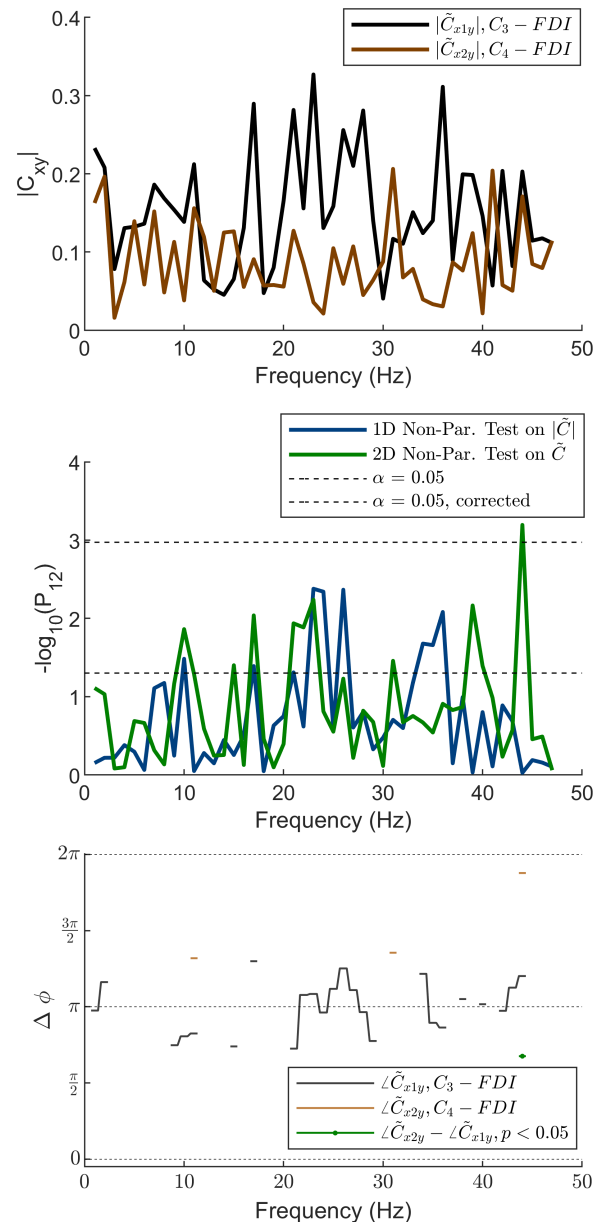
The new approach for spectral analysis takes into account the statistical distribution of the raw auto- and cross-spectra, enabling working with non-parametric estimates such as median and rank-based statistical tests. This approach which we previously introduced (Dukic *et al.*, 2017) is considerably different from the established parametric methods on spectral analyses of time series.

The approach to conduct the test on the data in 2D complex plane is a novel direction that can be achieved reasonably only by non-parametric tests (given the complex distribution of the complex data). This approach takes into account both the magnitude and phase of coherence. This enables the 1-sample testing of the coherence magnitude non-parametrically which would not be possible using existing methods. Importantly, this 2D approach enables the 2-sample comparison of two coherence levels. This would not be possible otherwise, except only for magnitude using the statistical trick in section 5.2.2. Importantly these tests are applicable in individual subjects.

### 7.2 Pros of Cons of Non-Parametric Spectral Statistics

Similar to 1D non-parametric tests based on rank statistics such as Wilcoxon's Signed Rank and Mann-Whitney's U test, the non-parametric 2D tests based on

### Significant Difference between Left and Right Experimental Cortico-Muscular Coherence



**Figure 13. Experimental Data: Both Parametric and Non-parametric Tests Detect Significant Difference between Cortico-Muscular Coherence of 2 Signal Pairs.** **Top:** The magnitude of coherence estimated using the median ( $|\tilde{C}_{xy}|$ ), show differences between the cortico-muscular coherence in the contralateral and ipsilateral EEG electrodes and FDI muscle. **Middle:** The p-values from the 1D parametric test on  $|\tilde{C}_{xy}|$  using the statistical trick for comparing 2 1-sample p-values (section 5.2.2) and the 2D non-parametric Spatial Rank test (section 5.2) both show frequency bands (especially the  $\beta$  band) with maximum differences. Corrected  $\alpha$  corresponds to Bonferroni correction with the number of frequencies. **Bottom:** The phase of coherence estimated using the median ( $\angle\tilde{C}_{xy}$ ) when the coherence is significant, is shown for  $x_1(C_3)$  and  $y(FDI)$ , as well as for  $x_2(C_4)$  and  $y(FDI)$ . The phase difference between 2 signal pairs is only plotted when both the individual pairs show significant difference, and when the 2D 2-sample non-parametric test shows significant difference. See section 5.2 for details of the tests and section 6.2 for details of the data used.

spatial signed ranks or ranks afford robustness against artefacts and outliers. They are applicable on arbitrary distributions. Theoretically, and as demonstrated by simulations they may not be as effective or powerful as parametric counterparts. Nevertheless in practical situations as shown in the real data, they provided very acceptable performance.

The non-parametric methods help to avoid computationally expensive bootstrapping methods when parametric methods are not available or are not preferred, as they do not suffer from unwanted variability in estimations.

The magnitude of non-parametric coherence  $\tilde{C}_{xy}$  is directly related to the p-values from the tests. However, unlike the parametric estimates, a closed-form solution is still unknown or does not exist. This is not a concern as the bias of the coherence measure (a function of the number of epochs  $L$ ) makes any interpretations dependent on the p-values associated to the coherence values. Non-parametric methods afford this p-value which can be plotted as  $-\log_{10}(p)$  spectrum and used as a more informative spectrum.

### 7.3 Applications

These developments and methods, while useful for broad range of neurophysiological research as well as in neuro-electro-magnetic source imaging, are especially necessary for comparison of pathological conditions against healthy individuals. Importantly, the new methods allow analyses at individual subject level which would only be practical at the level of subject/patient levels.

### 7.4 Extension to Other Measures and Analyses

Given the general nature of the methods, they can be similarly used in an extended range of methods and analyses.

#### 7.4.1 Frequency and Time-Frequency Measures

The methods can be similarly applied on most of the complex-values frequency and time-frequency method. Examples include the spectra or coherence measures calculated using the Welch's overlapping window (Terry & Griffin, 2008), spectral smoothing approach, or Multi-Taper techniques (Babadi & Brown, 2014; D. M. Halliday, Brittain, Stevenson, & Mason, 2018). Time-frequency representations based on complex values, including Short-Time Fourier Transform and Morlet wavelets (Boashash, 2003; Misiti, Misiti, Oppenheim, & Poggi, 2007) and their variations (Nasserolelami *et al.*, 2014; Mehrkanon, Breakspear, Daffertshofer, & Boonstra, 2013) can be similarly tested using the non-parametric techniques.

#### 7.4.2 Partial and Multiple Spectra

Other derivatives of the spectral measures, such as the partial spectra, partial coherence (Lindsay & Rosenberg, 2011) and multiple coherence (D. M. Halliday & Rosenberg, 1999) can be similarly tested using non-parametric methods, as they are similarly expressed in the 2-D complex plane. Care should be exercised when the null hypotheses of these measures may vary in specific applications.

#### 7.4.3 Group Analysis

Combination of time series data from several subjects or several experimental sessions can be achieved using the pooled coherence (D. M. Halliday & Rosenberg, 1999) of all the raw data points in a group. Alternatively this can be achieved using the existing methods for combining the p-values such as Stouffer's method (Stouffer, 1977; Westfall, 2014).

#### 7.4.4 High-Dimensional Statistics

Scientific inference from neural time series such as EEG and EMG signals may rely on analyses in high dimensions encompassing time, frequency and space/channels. The p-values from parametric and non-parametric analyses can be similarly corrected for multiple comparisons using the effective number of components (Nasserolelami *et al.*, 2014), random-field theory (Siegmund & Worsley, 1995; Mehrkanon *et al.*, 2014), or (adaptive) False Discovery Rate (Benjamini & Hochberg, 1995; Benjamini *et al.*, 2006). Methods such as Empirical Bayesian Inference Efron, 2007; Nasserolelami, 2018 applied to high-dimensional analyses of spectral measures (Nasserolelami *et al.*, 2017) affords the estimation of posterior probability and statistical power. Therefore, multiple comparison is straight-forward with new statistical methods.

#### 7.4.5 Source Analysis

The non-parametric statistical methods on spectral measures would greatly benefit the analysis of neural activity and connectivity in neuro-electro-magnetic source analysis (Ramírez, Wipf, & Baillet, 2010). Given the application of spectral source and connectivity analysis in healthy individuals and patient groups (Muthuraman *et al.*, 2018; A. R. Anwar *et al.*, 2016) and the crucial need to use robust unbiased measures of connectivity, the proposed methods can be of great utility.

#### 7.4.6 Future Directions

Future direction should work more on better ways to estimate the statistical power. This is automatically taken care of with some multi-variate methods such

as empirical Bayesian inference. However, currently, the straight-forward method for finding power in low-dimensional analysis is through non-null bootstrapping.

Paired 2-sample testing may be a possibility depending on the nature of analysis. This has not been adequately addressed to date. However, both parametric and non-parametric methods can be potentially used for such analyses and is expected to afford additional statistical power where applicable..

## 8 Conclusions and Recommendations

Non-parametric statistical analysis of spectral power and coherence can be used in several practical situations on 1-sample and 2-sample problems. In addition to the robustness afforded by non-parametric measures, based on rank or median, these statistical techniques are not restricted by parametric assumptions in traditional analysis and can be used in broader ranges of comparisons in neural signal analysis. In addition to rank-based 1-sample and 2-sample tests for comparing spectral power, the 2D Spatial Signed Rank and Spatial Rank Tests are the recommended tests for assessing the significance of coherence in 1-sample and 2-sample problems.

## Acknowledgement

The support of students and staff at Trinity College Dublin, the University of Dublin is highly appreciated.

This work was supported by Irish Research Council (Government of Ireland Postdoctoral Research Fellowship GOIPD/2015/213 to B.N.), Science Foundation Ireland (award SFI/16/ERC/3854), and Health Research Board of Ireland (award HRA-POR-2013-246).

## References

Akritis, M. G., Lahiri, S. N., & Politis, D. N. (Eds.). (2014). *Topics in Nonparametric Statistics*. New York, NY: Springer New York. Springer Proceedings in Mathematics & Statistics.

Anwar, A. R., Muthalib, M., Perrey, S., Galka, A., Granert, O., Wolff, S., ... Muthuraman, M. (2016, July 20). Effective Connectivity of Cortical Sensorimotor Networks During Finger Movement Tasks: A Simultaneous fNIRS, fMRI, EEG Study. *Brain Topogr*, 1–16. doi:10.1007/s10548-016-0507-1

Anwar, A., Mideska, K., Hellriegel, H., Hoogenboom, N., Krause, H., Schnitzler, A., ... Muthuraman, M. (2014, August). Multi-modal causality analysis of eyes-open and eyes-closed data from simultaneously recorded EEG and MEG. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 2825–2828). 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). doi:10.1109/EMBC.2014.6944211

Babadi, B. & Brown, E. N. (2014, May). A Review of Multitaper Spectral Analysis. *IEEE Transactions on Biomedical Engineering*, 61(5), 1555–1564. doi:10.1109/TBME.2014.2311996

Beasley, T. M., Erickson, S., & Allison, D. B. (2009, September 1). Rank-Based Inverse Normal Transformations are Increasingly Used, But are They Merited? *Behav Genet*, 39(5), 580. doi:10.1007/s10519-009-9281-0

Benjamini, Y. & Hochberg, Y. (1995, January 1). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1), 289–300.

Benjamini, Y., Krieger, A. M., & Yekutieli, D. (2006, January 9). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika*, 93(3), 491–507. doi:10.1093/biomet/93.3.491

Bloomfield, P. (2004, April 5). *Fourier Analysis of Time Series: An Introduction*. John Wiley & Sons.

Boashash, B. (Ed.). (2003). *Time-frequency signal analysis and processing: A comprehensive reference*. Elsevier.

Boonstra, T. W. (2013). The potential of corticomuscular and intermuscular coherence for research on human motor control. *Front. Hum. Neurosci*, 7, 855. doi:10.3389/fnhum.2013.00855

Box, G. E. P. & Cox, D. R. (1964, January 1). An Analysis of Transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, 26(2), 211–252. JSTOR: 2984418

Brillinger, D. R. (2001, September 1). *Time Series: Data Analysis and Theory*. SIAM.

Cohen, M. X. (2014, January 17). *Analyzing Neural Time Series Data: Theory and Practice* (1 edition). Cambridge, Massachusetts: The MIT Press.

Conway, B. A., Halliday, D. M., Farmer, S. F., Shahani, U., Maas, P., Weir, A. I., & Rosenberg, J. R. (1995, December). Synchronization between motor cortex and spinal motoneuronal pool during the performance of a maintained motor task in man. *J Physiol*, 489 ( Pt 3), 917–924. doi:10.1113/jphysiol.1995.sp021104

Dukic, S., Iyer, P. M., Mohr, K., Hardiman, O., Lalor, E. C., & Nasseroleslami, B. (2017, July). Estimation of coherence using the median is robust against EEG artefacts. In *2017 39th Annual Inter-*

- national Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 3949–3952). 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). doi:[10.1109/EMBC.2017.8037720](https://doi.org/10.1109/EMBC.2017.8037720)
- Efron, B. (2004). Large-Scale Simultaneous Hypothesis Testing: The Choice of a Null Hypothesis. *Journal of the American Statistical Association*, 99(465), 96–104.
- Efron, B. (2007, August). Size, power and false discovery rates. *Ann. Statist.* 35(4), 1351–1377. doi:[10.1214/009053606000001460](https://doi.org/10.1214/009053606000001460)
- Efron, B., Tibshirani, R., Storey, J. D., & Tusher, V. (2001). Empirical Bayes analysis of a microarray experiment. *Journal of the American statistical association*, 96(456), 1151–1160. doi:[10.1198/016214501753382129](https://doi.org/10.1198/016214501753382129)
- Farmer, S. F., Bremner, F. D., Halliday, D. M., Rosenberg, J. R., & Stephens, J. A. (1993). The frequency content of common synaptic inputs to motoneurons studied during voluntary isometric contraction in man. *The Journal of Physiology*, 470(1), 127–155. Retrieved July 2, 2013, from <http://jp.physoc.org/content/470/1/127.short>
- Fisher, K. M., Zaaami, B., Williams, T. L., Baker, S. N., & Baker, M. R. (2012, January 9). Beta-band intermuscular coherence: A novel biomarker of upper motor neuron dysfunction in motor neuron disease. *Brain*, 135(9), 2849–2864. doi:[10.1093/brain/aww150](https://doi.org/10.1093/brain/aww150). pmid: 22734124
- Fisher, N. I. (1995, October 12). *Statistical Analysis of Circular Data*. Cambridge University Press.
- Halliday, D., Rosenberg, J., Amjad, A., Breeze, P., Conway, B., & Farmer, S. (1995). A framework for the analysis of mixed time series/point process data—Theory and application to the study of physiological tremor, single motor unit discharges and electromyograms. *Progress in Biophysics and Molecular Biology*, 64(2–3), 237–278. doi:[10.1016/S0079-6107\(96\)00009-0](https://doi.org/10.1016/S0079-6107(96)00009-0)
- Halliday, D. M., Brittain, J.-S., Stevenson, C. W., & Mason, R. (2018). Adaptive spectral tracking for coherence estimation: The z-tracker. *J. Neural Eng.* 15(2), 026004. doi:[10.1088/1741-2552/aaa3b4](https://doi.org/10.1088/1741-2552/aaa3b4)
- Halliday, D. M. & Rosenberg, J. R. [Jay R.]. (1999). Time and frequency domain analysis of spike train and time series data. In *Modern techniques in neuroscience research* (pp. 503–543). Springer. Retrieved July 1, 2013, from [http://doi.org/10.1007/978-3-642-58552-4\\_18](http://doi.org/10.1007/978-3-642-58552-4_18)
- Hotelling, H. (1931, August). The Generalization of Student's Ratio. *Ann. Math. Statist.* 2(3), 360–378. doi:[10.1214/aoms/1177732979](https://doi.org/10.1214/aoms/1177732979)
- Iyer, P. M., Egan, C., Pinto-Grau, M., Burke, T., Elamin, M., Nasseroleslami, B., ... Hardiman, O. (2015, June 19). Functional Connectivity Changes in Resting-State EEG as Potential Biomarker for Amyotrophic Lateral Sclerosis. *PLoS ONE*, 10(6), e0128682. doi:[10.1371/journal.pone.0128682](https://doi.org/10.1371/journal.pone.0128682)
- Jurečková, J. & Kalina, J. (2012, February). Nonparametric multivariate rank tests and their unbiasedness. *Bernoulli*, 18(1), 229–251. doi:[10.3150/10-BEJ326](https://doi.org/10.3150/10-BEJ326)
- Koles, Z. J. & Flor-Henry, P. (1987, May). The effect of brain function on coherence patterns in the bipolar EEG. *Int J Psychophysiol*, 5(1), 63–71. pmid: 3597171
- Lindsay, K. A. & Rosenberg, J. R. [J R]. (2011, June). Identification of directed interactions in networks. *Biol Cybern*, 104(6), 385–396. doi:[10.1007/s00422-011-0437-7](https://doi.org/10.1007/s00422-011-0437-7). pmid: 21678101
- Mehrkanoon, S., Breakspear, M., & Boonstra, T. W. (2014, October 15). The reorganization of corticomuscular coherence during a transition between sensorimotor states. *NeuroImage*, 100, 692–702. doi:[10.1016/j.neuroimage.2014.06.050](https://doi.org/10.1016/j.neuroimage.2014.06.050)
- Mehrkanoon, S., Breakspear, M., Daffertshofer, A., & Boonstra, T. W. (2013). Non-identical smoothing operators for estimating time-frequency interdependence in electrophysiological recordings. *EURASIP Journal on Advances in Signal Processing*, 2013(1), 1. Retrieved October 9, 2016, from <http://asp.eurasipjournals.springeropen.com/articles/10.1186/1687-6180-2013-73>
- Misiti, M., Misiti, Y., Oppenheim, G., & Poggi, J.-M. (2007). *Wavelets and their applications* (M. Misiti, Y. Misiti, G. Oppenheim, & J.-M. Poggi, Eds.). London: ISTE.
- Mohr, K., Nasseroleslami, B., Iyer, P. M., Hardiman, O., & Lalor, E. C. (2017, May 6). EyeBallGUI: A Tool for Visual Inspection and Binary Marking of Multi-Channel Bio-Signals. *bioRxiv*, 129437. doi:[10.1101/129437](https://doi.org/10.1101/129437)
- Muthuraman, M., Raethjen, J., Koirala, N., Anwar, A. R., Mideksa, K. G., Elble, R., ... Deuschl, G. (2018). Cerebello-cortical network fingerprints differ between essential, Parkinson's and mimicked tremors. *Brain*. doi:[10.1093/brain/awy098](https://doi.org/10.1093/brain/awy098)
- Nasseroleslami, B. (2018, June 8). An Implementation of Empirical Bayesian Inference and Non-Null Bootstrapping for Threshold Selection and Power Estimation in Multiple and Single Statistical Testing. *bioRxiv*, 342964. doi:[10.1101/342964](https://doi.org/10.1101/342964)
- Nasseroleslami, B., Dukic, S., Broderick, M., Mohr, K., Schuster, C., Gavin, B., ... Hardiman, O. (2017, November). Characteristic Increases in EEG Connectivity Correlate With Changes of Structural MRI in Amyotrophic Lateral Sclerosis. *Cereb Cortex*, 1–15. doi:[10.1093/cercor/bhx301](https://doi.org/10.1093/cercor/bhx301)
- Nasseroleslami, B., Lakany, H., & Conway, B. A. (2014, April 15). EEG signatures of arm isometric ex-

- ertions in preparation, planning and execution. *NeuroImage*, 90, 1–14. doi:10.1016/j.neuroimage.2013.12.011
- Niinimaa, A. & Oja, H. [H.]. (2006). Multivariate Median. In *Encyclopedia of Statistical Sciences*. American Cancer Society. doi:10.1002/0471667196.ess1107.pub2
- Nolan, H., Whelan, R., & Reilly, R. B. (2010, September 30). FASTER: Fully Automated Statistical Thresholding for EEG artifact Rejection. *Journal of Neuroscience Methods*, 192(1), 152–162. doi:10.1016/j.jneumeth.2010.07.015
- Nordhausen, K., Mottonen, J., & Oja, H. (2018, March 3). MNM: Multivariate Nonparametric Methods. An Approach Based on Spatial Signs and Ranks (Version 1.0-3). Retrieved May 21, 2018, from <https://CRAN.R-project.org/package=MNM>
- Nordhausen, K. & Oja, H. [Hannu]. (2011). Multivariate L1 Methods: The Package **MNM**. *Journal of Statistical Software*, 43(5). doi:10.18637/jss.v043.i05
- Oja, H. [Hannu]. (2010, March 25). *Multivariate Nonparametric Methods with R: An approach based on spatial signs and ranks*. Springer.
- Oja, H. [Hannu] & Randles, R. H. (2004, November). Multivariate Nonparametric Tests. *Statistical Science*, 19(4), 598–605. doi:10.1214/088342304000000558
- Person, R. S. & Kudina, L. P. (1968). Cross-correlation of electromyograms showing interference patterns. *Electroencephalography and clinical neurophysiology*, 25(1), 58–68. Retrieved November 28, 2016, from <http://www.sciencedirect.com/science/article/pii/0013469468900874>
- Priestley, M. B. (1982). *Spectral Analysis and Time Series*. Academic Press.
- Puri, M. L. & Sen, P. K. [Pranab Kumar]. (1971). *Nonparametric methods in multivariate analysis*. Krieger Pub. Co.
- R Core Team. (2016). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org>
- Ramírez, R. R., Wipf, D., & Baillet, S. (2010). Neuroelectromagnetic Source Imaging of Brain Dynamics. In W. Chaovalitwongse, P. M. Pardalos, & P. Xanthopoulos (Eds.), *Computational Neuroscience* (Vol. 38, pp. 127–155). New York, NY: Springer New York. Retrieved September 28, 2015, from [doi.org/10.1007/978-0-387-88630-5\\_8](https://doi.org/10.1007/978-0-387-88630-5_8)
- Rosenberg, J. R. [J. R.], Halliday, D. M., Breeze, P., & Conway, B. A. (1998, August 31). Identification of patterns of neuronal connectivity—partial spectra, partial coherence, and neuronal interactions. *Journal of Neuroscience Methods*, 83(1), 57–72. doi:10.1016/S0165-0270(98)00061-2
- Sen, P. K. [Pranab K.]. (2005, July 15). Multivariate Median and Rank Sum Tests. In P. Armitage & T. Colton (Eds.), *Encyclopedia of Biostatistics*. Chichester, UK: John Wiley & Sons, Ltd. doi:10.1002/0470011815.b2a13052
- Sen, P. K. [Pranab Kumar] & Puri, M. L. (1967). On the theory of rank order tests for location in the multivariate one sample problem. *The Annals of Mathematical Statistics*, 1216–1228.
- Siegel, M., Donner, T. H., & Engel, A. K. (2012, February). Spectral fingerprints of large-scale neuronal interactions. *Nat Rev Neurosci*, 13(2), 121–134. doi:10.1038/nrn3137
- Siegmund, D. O. & Worsley, K. J. (1995, April). Testing for a Signal with Unknown Location and Scale in a Stationary Gaussian Random Field. *Ann. Statist.* 23(2), 608–639. doi:10.1214/aos/1176324539
- Stouffer, S. A. (1977). *The American Soldier: Adjustment during army life*. Military Affairs/Aerospace Historian.
- Tatum, W. O., Dworetzky, B. A., & Schomer, D. L. (2011, June). Artifact and recording concepts in EEG. *J Clin Neurophysiol*, 28(3), 252–263. doi:10.1097/WNP.0b013e31821c3c93
- Terry, K. & Griffin, L. (2008, February). How computational technique and spike train properties affect coherence detection. *Journal of Neuroscience Methods*, 168(1), 212–223. doi:10.1016/j.jneumeth.2007.09.014
- van Ede, F., Quinn, A. J., Woolrich, M. W., & Nobre, A. C. (2018, May 5). Neural Oscillations: Sustained Rhythms or Transient Burst-Events? *Trends in Neurosciences*. doi:10.1016/j.tins.2018.04.004
- Vuckovic, A. [A.], Hasan, M. A., Osuagwu, B., Fraser, M., Allan, D. B., Conway, B. A., & Nasserolelami, B. (2015, November). The influence of central neuropathic pain in paraplegic patients on performance of a motor imagery based Brain Computer Interface. *Clinical Neurophysiology*, 126(11), 2170–2180. doi:10.1016/j.clinph.2014.12.033
- Vuckovic, A. [Aleksandra], Hasan, M. A., Fraser, M., Conway, B. A., Nasserolelami, B., & Allan, D. B. (2014, June). Dynamic Oscillatory Signatures of Central Neuropathic Pain in Spinal Cord Injury. *The Journal of Pain*, 15(6), 645–655. doi:10.1016/j.jpain.2014.02.005
- Waldert, S., Pistohl, T., Braun, C., Ball, T., Aertsen, A., & Mehring, C. (2009, May). A review on directional information in neural signals for brain-machine interfaces. *Journal of Physiology-Paris. Neurorobotics*, 103(3–5), 244–254. doi:10.1016/j.jphysparis.2009.08.007
- Walter, D. O. (1963, August). Spectral analysis for electroencephalograms: Mathematical determination of neurophysiological relationships from records of limited duration. *Experimental Neurology*, 8(2), 155–181. doi:10.1016/0014-4886(63)90042-6

- Watson, G. S. & Williams, E. J. (1956). On the Construction of Significance Tests on the Circle and the Sphere. *Biometrika*, 43(3/4), 344–352. doi:10.2307/2332913. JSTOR: 2332913
- Westfall, P. H. (2014). Combining P-values. In *Wiley StatsRef: Statistics Reference Online*. John Wiley & Sons, Ltd. doi:10.1002/9781118445112.stat05891
- Wijk, V., M, B. C., Beek, P. J., & Daffertshofer, A. (2012). Neural synchrony within the motor system: What have we learned so far? *Front. Hum. Neurosci.* 6. doi:10.3389/fnhum.2012.00252
- Xu, R., JIANG, N., Vuckovic, A., Hasan, M., Mrchacz-Kersting, N., Allan, D., ... Farina, D. (2014, August 27). Movement-related cortical potentials in paraplegic patients: Abnormal patterns and considerations for BCI-rehabilitation. *Front. Neuroeng.* 7, 35. doi:10.3389/fneng.2014.00035