Contents lists available at ScienceDirect

SoftwareX

journal homepage: www.elsevier.com/locate/softx

Original software publication

# Cluster permutation analysis for EEG series based on non-parametric Wilcoxon–Mann–Whitney statistical tests

Diego Candia-Rivera [*], Gaetano Valenza

*Bioengineering and Robotics Research Center E. Piaggio and the Department of Information Engineering, School of Engineering, University of Pisa, Pisa, Italy*

ARTICLE INFO

*Article history:*
Received 5 March 2022
Received in revised form 23 May 2022
Accepted 22 July 2022

*Keywords:*
Cluster-based permutation
EEG
Statistical inference
Wilcoxon
Mann–Whitney
t-test

ABSTRACT

Cluster-based permutation tests are widely used in neuroscience studies for the analysis of high-dimensional electroencephalography (EEG) and event-related potential (ERP) data as it may address the multiple comparison problem without reducing the statistical power. However, classical cluster-based permutation analysis relies on parametric t-tests, whose assumptions may not be verified in case of non-normality of the data distribution and alternative options may be considered. To overcome this limitation, here we present a new software for a cluster permutation analysis for EEG series based on non-parametric Wilcoxon–Mann–Whitney tests. We tested both t-test and non-parametric Wilcoxon implementations in two independent datasets of ERPs and EEG spectral data: while t-test-based and non-parametric Wilcoxon-based cluster analyses showed similar results in case of ERP data, the t-test implementation was not able to find clustered effects in case of spectral data. We encourage the use of non-parametric statistics for a cluster permutation analysis of EEG data, and we provide a publicly available software for this computation.

## Code metadata

| | |
|---|---|
| Current code version | *v1* |
| Permanent link to code/repository used for this code version | https://github.com/ElsevierSoftwareX/SOFTX-D-22-00059 |
| Permanent link to reproducible capsule | https://github.com/diegocandiar/eeg_cluster_wilcoxon |
| Legal code license | *GNU General Public License (GPL).* |
| Code versioning system used | *none* |
| Software code languages, tools and services used | *MATLAB* |
| Compilation requirements, operating environments and dependencies | *MATLAB 2017a or superior* |
| If available, link to developer documentation/manual | |
| Support email for questions | diego.candia.r@ug.uchile.cl |

## 1. Motivation and significance

Cluster-based permutation tests of EEG/ERP data are widely used methodologies in the fields of psychophysiology and neuroscience. EEG data are characterized by the spatiotemporal structure in which univariate statistical tests should be conducted at every data sample in time; however, this may lead to misconclusions due to errors related to multiple statistical comparison [1]. Classical correction methods for multiple testing may be used to tackle this issue, nevertheless, they may reduce the measured effect size and the probability of observing the true effect present in the data [2]. To address these concerns and maximize the power in the statistical analysis, a cluster-based analysis has been proposed with the assumption that neural effects are clustered in the different dimensions of interest: space, time, and/or frequency.

Cluster-based analysis leverages on univariate tests, a mask definition based on their a priori significance, and candidate clusters identification by neighboring points in the dimensions of interest. The first proposal of cluster-based permutation tests was introduced for magnetic resonance analysis [3], and since then other generic open-access implementations have been released [4–6]. Classical implementations of cluster-based permutation analysis rely on parametric t-tests in the sample-to-sample analysis to identify the candidate clusters. However, the t-test

[*] Corresponding author.
*E-mail address:* diego.candia.r@ug.uchile.cl (Diego Candia-Rivera).

https://doi.org/10.1016/j.softx.2022.101170

assumes the normality of distributions associated with the random variables that generated the samples; moreover, such distributions are assumed to have equal variances. EEG and/or ERP samples may show non-normal distribution, failing the assumptions of a parametric statistical analysis. Specifically, t-tests fail in the conditions of small sample size, presence of non-normality, or heteroscedasticity [7–9].

An illustrative example on the differences between parametric and non-parametric tests is shown in Fig. 1. We generated samples from two random variables having different mean and opposite skewness. Samples were generated at different sample sizes, and for each generation we tested statistical differences through Wilcoxon test and t-test. While Wilcoxon tests showed the majority of p-values lower than 0.05, the application of parametric t-test was associated with the majority of p-values higher than 0.05. This may also be due to the differences in standard deviation, as revealed by the application of F-tests.

Thus, a rank-based non-parametric statistical analysis, where no assumption on the specific probability density function of the random variables is made, would be more appropriate for a cluster analysis. Despite the vast availability of software implementations, to our knowledge non-parametric Wilcoxon and Mann Whitney tests, paired sign rank or unpaired rank sum, respectively, for mask and candidate cluster definition have not been exploited.

We remark that the application non-parametric tests may not always be optimal or outperform a parametric method [10–13], for instance, in case the random variables have equal median and different higher-order moments (e.g., skewness and kurtosis).

Here we describe a new software implementation of non-parametric cluster analysis and show differences and similarities in the application of a t-test and a Wilcoxon test using real data.

## 2. Software description

### 2.1. Software architecture

This implementation of cluster-based analysis comprises the application of univariate statistical tests for each data sample in time, a mask definition based on their a priori significance, and candidate clusters identification based on neighboring points in the dimensions of interest. Cluster-level statistics are computed by combining all cluster's data samples and univariate tests to multiple random partitions to evaluate the probability of having a true effect.

*Mask definition*

A mask is defined as the set of statistical tests associated with a significant *p*-value, i.e., p<$\boldsymbol{\alpha_c}$ where $\boldsymbol{\alpha_c}$ is the significance that is typically set at 0.05 or 0.01 [14]. A preliminary mask is constructed by performing univariate statistical tests at individual data points, with consecutive candidate clusters definition based on neighboring points in the dimensions studied, i.e., time, time–space, and time–frequency–space. The mask is defined by performing first-level statistical tests comparing two conditions or two independent groups of participants, i.e., paired or unpaired statistical test. To illustrate, the null hypothesis $H_0$ for an unpaired test refers to two groups of subjects exposed to the same experimental condition; under $H_0$, while the difference in the sample mean/median is likely different from zero, the difference between the two random variables' mean/median is precisely 0. For a paired test, the null hypothesis $H_0$ refers to, e.g., one group of subjects exposed to two different experimental conditions, resulting in random variables' mean/median difference between the two conditions equal to 0.

In this study we compare the clustered effects found by the algorithm using parametric t-tests and non-parametric Wilcoxon tests. The chosen statistical test is repeated for all possible data points in the dimension under study (e.g., time). The tests resulting in a *p*-value lower than a significance $\boldsymbol{\alpha_c}$ will be included in the preliminary mask. Given the high-dimensional data and associated multiple comparison issue, it is expected to observe significant p-values even in case the samples are derived from the very same random variable, i.e., the null hypothesis $H_0$ should be accepted. This issue is resolved by identifying the clusters as described below.

*Candidate cluster identification*

Different methods for multiple comparison correction exist, such as error rate control (false discovery rate, or family-wise error rate), or cluster-based permutations [1], which are the most widely used in EEG/ERP analyses [14]. Cluster-based permutation tests group neighbor data samples if present in the preliminary mask. In one-dimension data, i.e., 1-D time series without spatial definition, the grouping is considered for adjacent data points over time. In 2-D data, the time series have also a spatial dimension, and their neighboring definition depends on the EEG system (e.g., number of channels and electrode positioning system) and should be properly defined by the user along with a minimum cluster size ($\mathbf{min}\, \boldsymbol{n_{channel}}$). Two or more clusters can be combined if they have a minimum number of channels in common at the same timestamp ($\mathbf{min}\, \boldsymbol{n_{channel}}$). In 3-D data, the power at defined frequencies is retained for further analyses along with the existing time and spatial dimensions. In this case, the candidate clusters could be constituted first on individual time stamps and separately for each frequency band. The combination of clusters can be performed if they intersect in a minimum of 2-D points, in time and space dimensions ($\mathbf{min}\, \boldsymbol{n_{channel\ x\ time}}$), at the same frequency. An additional parameter to set is the threshold for the cluster's minimum duration ($\mathbf{min}\, \boldsymbol{T_{cluster}}$), i.e., the overall duration of the cluster considering the latency between the earliest point in the cluster and the latest.

*Cluster statistics*

Statistical inference from the clusters can be performed through permutation tests [15]. The data points defined by the cluster's mask in space, time and frequency dimensions are averaged, and $\boldsymbol{n_{rand}}$ random partitions are constructed to compute the Monte Carlo *p*-value ($p_{mc}$) [14]. The null hypothesis $H_0$ of such second-level statistics is that the identified cluster effect is exchangeable between the conditions. Therefore, the statistical test on the random partitions aims to represent the empirical cumulative density function of all possible null values. A number of random permutations $\boldsymbol{n_{rand}} \geq 800$ is recommended [16]. The samples of the two conditions are randomly permutated and the statistical test is performed over every single random partition. The computation of the Monte Carlo *p*-value considers the proportion of random partitions that resulted in a lower *p*-value (or higher effect size) than the observed one. The significance of the cluster's Monte Carlo *p*-value is defined as p < $\boldsymbol{\alpha_{cluster}}$. The cluster statistic presented in this study is the maximum effect size (t-stat or z-value) found within the cluster during the first-level univariate statistical tests stage.

### 2.2. Software functionalities

The algorithm reads the data and channel neighboring information as structured in Fieldtrip Toolbox [6] for MATLAB, but the Toolbox is not needed to run the MATLAB functions. This implementation of cluster permutation analysis can be applied to any EEG data or EEG-derived markers. This implementation was successfully applied to describe brain-heart interplay under thermal stress [17] and emotion elicitation [18].

A summary of the parameter to set is shown in Table 1. These parameters must be indicated in the configuration struct required as an input in the MATLAB function.

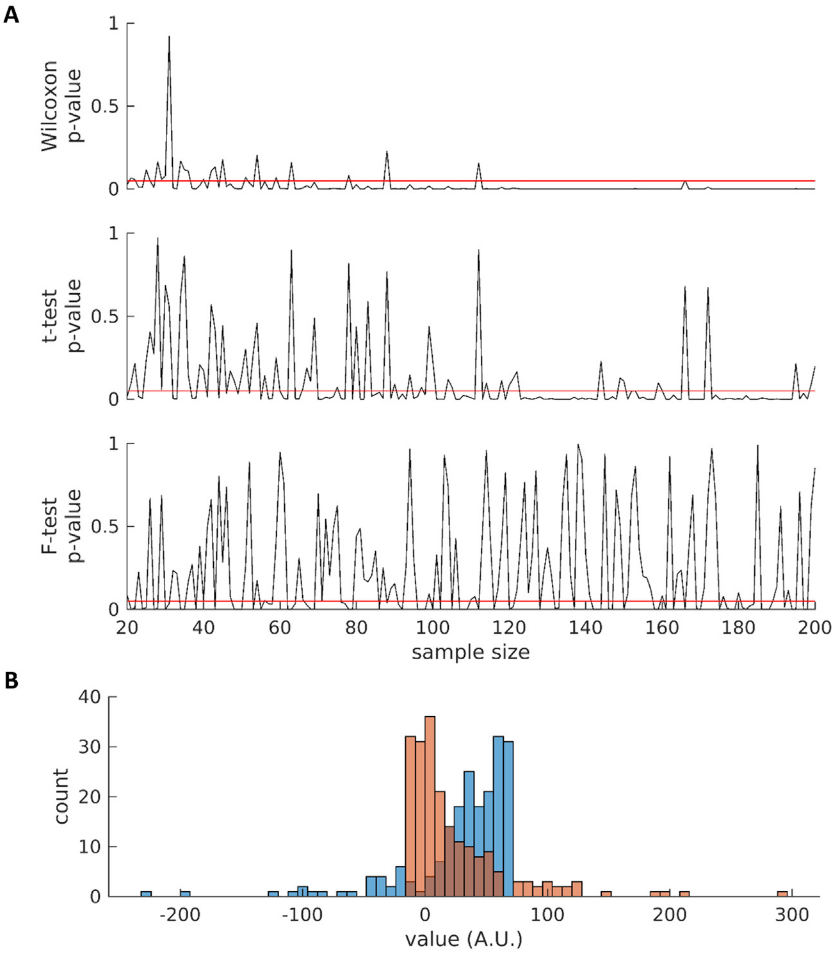

**Fig. 1.** Simulation analysis comparing two random vectors. The random vectors X and Y were generated with a Pearson system with the following parameters for mean, variance, skewness and kurtosis: {35, 40, −3, 20} and {20, 40, 3, 20}, respectively. (A) The random vectors were compared with a Wilcoxon test, t-test, and F-test. P-values are reported for each test as a function of the sample size, ranging from 20 to 200. The red line indicates $p$-value = 0.05. (B) Exemplary histogram distribution for sample size 200. The resulting p-values for Wilcoxon, t-test and F-test are 0.0006, 0.1978 and 0.8531, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

**Table 1**
Summary of parameters and configuration options for cluster-based permutation analysis.

| Configuration/Parameters | Options/Default | In this study |
|---|---|---|
| Dependency | Paired or unpaired | Paired |
| Statistical test | t-test or Wilcoxon test | t-test and Wilcoxon test |
| Number of randomizations | $\geq$ 800 | 10000 |
| Critical alpha ($\boldsymbol{\alpha_c}$) | $\leq$ 0.05 | 0.01 |
| Minimum cluster size (**min** $\boldsymbol{n_{channel}}$) | $\geq$ 2 channels | 3 channels |
| Minimum 2D neighboring points (**min** $\boldsymbol{n_{channel\ x\ time}}$) | $\geq$ 1 | 1 |
| Cluster alpha ($\boldsymbol{\alpha_{cluster}}$) | $\leq$ 0.05 | 0.01 |
| Cluster duration (**min** $\boldsymbol{T_{cluster}}$) | $\geq$ 2 samples | 5 samples |

## 3. Illustrative examples

In these examples we describe the software application in two independent datasets. The inclusion criteria were EEG data publicly available, one with the possibility to perform ERP analysis (i.e., the stimuli have a short duration, and evoke a cortical potential associated to a sensory, motor, or cognitive event), and another in which spectrogram analyses is suitable (i.e., audiovisual stimuli with at least 60-second-long duration).

(1) ERP dataset: Brain invaders P300 dataset, consisting in visual stimuli (Space invaders-like images, 1978 video game by Taito) with or without presence of a flashing icon (white colored). 32 channel EEG data were gathered from 41 participants (age range 19–32 years, 30 males). For further details on this dataset, please see [19].

(2) Spectrogram dataset: MAHNOB-HCI dataset of emotion elicitation, consisting in visual stimuli through video clips with affective content. 32 channel EEG data were gathered from 26 healthy subjects (age range 20–37 years, 11 males). For further details on this dataset, please see [20]. The selected trial is an excerpt from the 2002 film The Pianist, by Roman Polanski (distributed by BAC Films in France, Tobis Film in Germany, Syrena Entertainment Group in Poland, and Pathé Distribution in United Kingdom). The emotions associated are anger and sadness. The trial corresponds to the highest arousal score among all trials. The approximate duration of the video clip is 80 s.

The study was performed in accordance with the Declaration of Helsinki.

EEG data processing was performed using Fieldtrip toolbox [6]. The pre-processing consisted in 0.5–45 Hz bandpass frequency

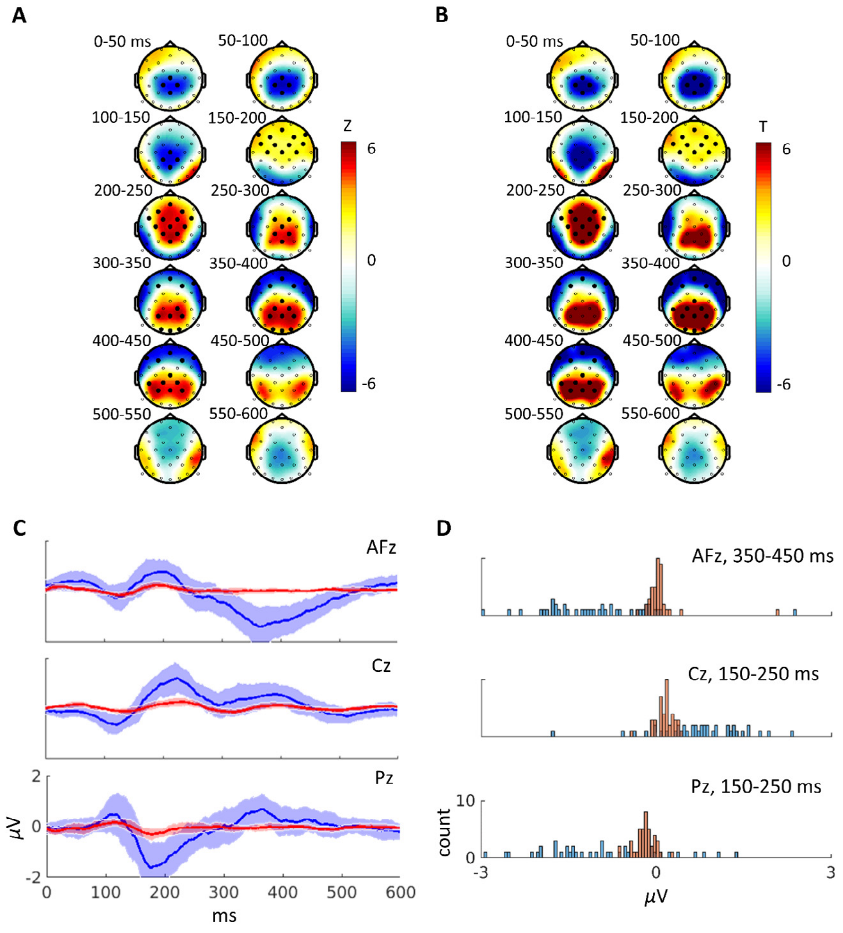

**Fig. 2.** Cluster permutation analysis on the ERP dataset, (A) based on Wilcoxon test and (B) based on t-test. Scalp topographies show the statistic, Wilcoxon's Z-value or T-stat, comparing flashing light vs without flashing light. Thick electrodes show presence of an observed effect in cluster-permutation analysis. (C) Time course of three channels (median ± median absolute deviation) during the visualization of the flashing light (in blue) and without flashing light (in red). (D) Distribution of the selected three channels AFz, Cz and Pz, averaged at 350–450 ms, 150–250 ms and 150–250 ms, respectively, for the ERP during the visualization of the flashing light (in blue) and without flashing light (in red). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

filtering, large artifact removal, eye movements and cardiac-field artifacts removal through independent component analysis, interpolation of contaminated channels, and common average re-reference [21]. The EEG spectrogram was computed using a short-time Fourier transform to gather time series integrated within the delta band ($\delta$; 0–4 Hz). For a more comprehensive review of EEG pre-processing, please see [21].

We performed a cluster permutation analysis based on t-test and Wilcoxon test in the ERP dataset. The clusters did not differ significantly between the two methods. Four clusters were found when comparing the ERPs when visualizing flashing icons vs the ERPs without flashing icons.

The clusters identified through a Wilcoxon and T-test are presented in Table 2. Both methods found the same clusters, and differences are mainly related to the latency, where t-test repeatedly found the clusters at a later latency with respect to Wilcoxon test. Figs. 2A and 2B show the results from a cluster permutation analysis based on Wilcoxon tests and t-test, respectively. Fig. 2C shows the time course of three EEG channels: AFz, Cz and Pz, and Fig. 2D shows the distribution of all data points for the same channels.

Next, we show the same cluster permutation analysis on the spectrogram dataset. The method based on Wilcoxon test found 4 clusters, as shown in Fig. 3A, whereas the method based on t-test did not find clusters. The distribution of the averaged delta power in frontal channels is shown in Fig. 3B.

## 4. Impact

Neural data distributions may present non-normality and might show very skewed distributions or heavy tails, therefore the use of non-parametric tests for statistical inference is recommended [7,22].

The use of non-parametric methods for statistical inference may have a positive impact in brain research, the validation of biomarkers developed for cognitive paradigms, brain–computer interfaces, and diagnostic/therapeutic applications based on ERP data analysis [23–25].

In the examples above, we showed minor differences when comparing t-test vs Wilcoxon test in ERP data. However, when performing the same comparison on EEG spectral data, the algorithm based on parametric t-test did not find any clustered effect, whereas the algorithm based on non-parametric Wilcoxon test found four clusters.

As previously recommended, the results obtained from a cluster permutation analyses should have a cautious interpretation [26] as they depend on a specific threshold set to accept or reject the null hypothesis. We encourage to follow those guidelines for a proper interpretation of the results obtained through a cluster analysis, with special regard to the use of non-parametrical statistics when analyzing neural data.

**Table 2**
Clustered effects found in the ERP dataset using cluster permutation analysis based on Wilcoxon and T-test.

| | Wilcoxon implementation | | | | T-test implementation | | | |
|---|---|---|---|---|---|---|---|---|
| | $p_{mc}$ | Stat (Z) | $p_{cluster}$ | Latency (ms) | $p_{mc}$ | Stat (T) | $p_{cluster}$ | Latency (ms) |
| Negative clusters | < 0.0001 | −5.46 | $6.77 \cdot 10^{-8}$ | 15–101 | < 0.0001 | −9.67 | $7.48 \cdot 10^{-12}$ | 27–95 |
| | < 0.0001 | −5.54 | $5.85 \cdot 10^{-8}$ | 336–422 | < 0.0001 | −10.26 | $1.08 \cdot 10^{-10}$ | 338–422 |
| Positive clusters | < 0.0001 | 5.33 | $7.82 \cdot 10^{-8}$ | 172–250 | < 0.0001 | 8.6 | $1.38 \cdot 10^{-10}$ | 173–240 |
| | < 0.0001 | 5.57 | $3.03 \cdot 10^{-8}$ | 332–422 | < 0.0001 | 13.4 | $5.19 \cdot 10^{-15}$ | 345–423 |

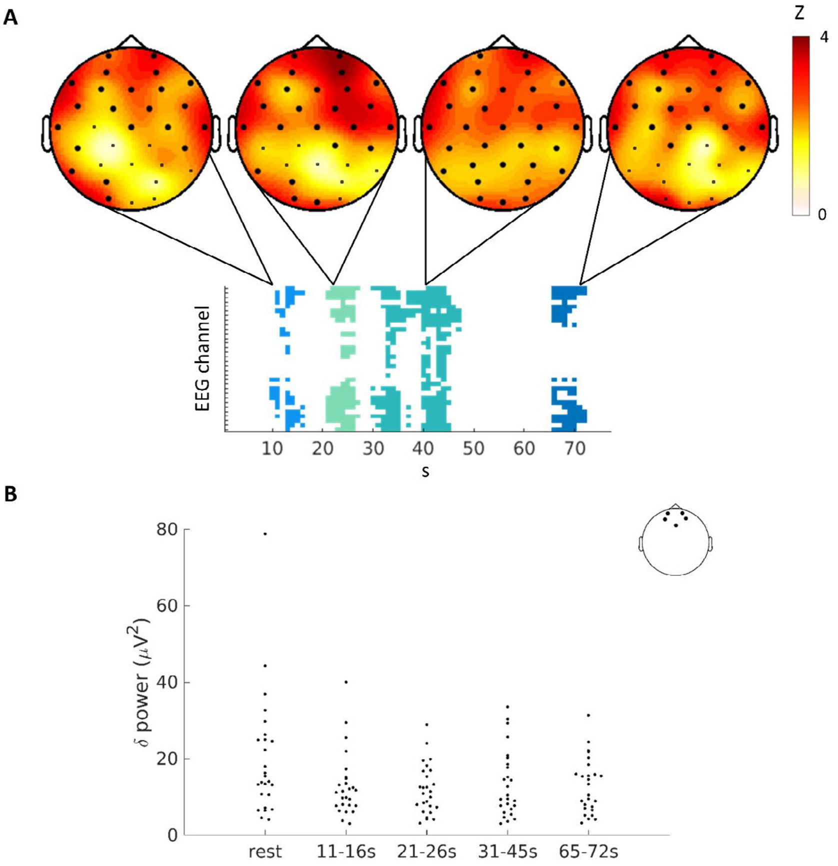

**Fig. 3.** Cluster permutation analysis based on Wilcoxon test in EEG spectral data. (A) The four clusters found are represented in different colors in the bottom graph. Each colored pixel represents a point in the channel and time dimensions belonging to a specific cluster. Scalp topographies represent the averaged Wilcoxon's Z-values on the clusters' latencies, and thick electrodes show presence of an observed effect in cluster-permutation analysis. (B) Distribution of the delta power average in frontal channels for rest and four time windows in which a clustered effect was found using the Wilcoxon implementation.

## 5. Conclusions

Cluster-based permutation tests address the multiple comparison problem without reducing the statistical power. This new software implementation offers a cluster permutation analysis for EEG series based on non-parametric Wilcoxon–Mann–Whitney tests, as well as the standard t-test. The results using t-test and non-parametric Wilcoxon implementations in two independent datasets showed that the Wilcoxon implementation may outperform the t-test version in finding underlying clustered effects in the data. We encourage the use of fully non-parametric statistics for a cluster permutation analysis applied on EEG data.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

The research leading to these results has received partial funding from the European Commission - Horizon 2020 Program under grant agreement no 813234 of the project "RHUMBO" and by the Italian Ministry of Education and Research (MIUR) in the framework of the CrossLab project (Departments of Excellence).

Portion of the research in this paper uses the MAHNOB Database collected by Professor Pantic and the iBUG group at Imperial College London, and in part collected in collaboration with Prof. Pun and his team of University of Geneva, in the scope of MAHNOB project financially supported by the European Research Council under the European Community's 7th Framework programme (FP7/2007-2013)/ERC Starting grant agreement No. 203143.

## References

[1] Groppe DM, Urbach TP, Kutas M. Mass univariate analysis of event-related brain potentials/fields I: a critical tutorial review. Psychophysiology 2011;48:1711–25. http://dx.doi.org/10.1111/j.1469-8986.2011.01273.x.

[2] Button KS, Ioannidis JPA, Mokrysz C, Nosek BA, Flint J, Robinson ESJ, et al. Power failure: why small sample size undermines the reliability of neuroscience. Nat Rev Neurosci 2013;14:365–76. http://dx.doi.org/10.1038/nrn3475.

[3] Bullmore E, Brammer M, Williams SCR, Rabe-Hesketh S, Janot N, David A, et al. Statistical methods of estimation and inference for functional MR image analysis. Magn Reson Med 1996;35:261–77. http://dx.doi.org/10.1002/mrm.1910350219.

[4] Delorme A, Makeig S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J Neurosci Methods 2004;134:9–21. http://dx.doi.org/10.1016/j.jneumeth.2003.10.009.

[5] Gramfort A, Luessi M, Larson E, Engemann DA, Strohmeier D, Brodbeck C, et al. MEG and EEG data analysis with MNE-python. Front Neurosci 2013;7. http://dx.doi.org/10.3389/fnins.2013.00267.

[6] Oostenveld R, Fries P, Maris E, Schoffelen J-M. FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput Intell Neurosci 2011;2011:9. http://dx.doi.org/10.1155/2011/156869.

[7] Fay MP, Proschan MA. Wilcoxon-Mann–Whitney or t-test? On assumptions for hypothesis tests and multiple interpretations of decision rules. Stat Surv 2010;4:1–39. http://dx.doi.org/10.1214/09-SS051.

[8] Wiedermann W, von Eye A. Robustness and power of the parametric t test and the nonparametric Wilcoxon test under non-independence of observations. Psychol Test Assess Model 2013;55:39–61.

[9] Zimmerman DW. Comparative power of student T test and Mann–Whitney U test for unequal sample sizes and variances. J Exp Educ 1987;55:171–4. http://dx.doi.org/10.1080/00220973.1987.10806451.

[10] Divine G, Norton HJ, Hunt R, Dienemann J. A review of analysis and sample size calculation considerations for Wilcoxon tests. Anesth Analg 2013;117:699–710. http://dx.doi.org/10.1213/ANE.0b013e31827f53d7.

[11] Divine G, Norton HJ, Barón AE, Juarez-Colunga E. The Wilcoxon–Mann–Whitney procedure fails as a test of medians. Am Stat 2018;72:278–86. http://dx.doi.org/10.1080/00031305.2017.1305291.

[12] Fagerland MW, Sandvik L. The Wilcoxon–Mann–Whitney test under scrutiny. Stat Med 2009;28:1487–97. http://dx.doi.org/10.1002/sim.3561.

[13] Zimmerman DW, Zumbo BD. Relative power of the Wilcoxon test, the friedman test, and repeated-measures ANOVA on ranks. J Exp Educ 1993;62:75–86. http://dx.doi.org/10.1080/00220973.1993.9943832.

[14] Maris E, Oostenveld R. Nonparametric statistical testing of EEG- and MEG-data. J Neurosci Methods 2007;164:177–90. http://dx.doi.org/10.1016/j.jneumeth.2007.03.024.

[15] Ernst MD. Permutation methods: A basis for exact inference. Statist Sci 2004;19:676–85. http://dx.doi.org/10.1214/088342304000000396.

[16] Pernet CR, Latinus M, Nichols TE, Rousselet GA. Cluster-based computational methods for mass univariate analyses of event-related brain potentials/fields: A simulation study. J Neurosci Methods 2015;250:85–93. http://dx.doi.org/10.1016/j.jneumeth.2014.08.003.

[17] Candia-Rivera D, Catrambone V, Barbieri R, Valenza G. Functional assessment of bidirectional cortical and peripheral neural control on heartbeat dynamics: a brain-heart study on thermal stress. NeuroImage 2022;251:119023. http://dx.doi.org/10.1016/j.neuroimage.2022.119023.

[18] Candia-Rivera D, Catrambone V, Thayer JF, Gentili C, Valenza G. Cardiac sympathetic-vagal activity initiates a functional brain-body response to emotional arousal. Proc Natl Acad Sci 2022;119(21). http://dx.doi.org/10.1073/pnas.211959911, e2119599119.

[19] Korczowski L, Cederhout M, Andreev A, Cattan G, Coelho Rodrigues PL, Gautheret V, et al. Brain invaders calibration-less P300-based BCI with modulation of flash duration dataset. 2019, http://dx.doi.org/10.5281/zenodo.3266930.

[20] Soleymani M, Lichtenauer J, Pun T, Pantic M. A multimodal database for affect recognition and implicit tagging. IEEE Trans Affect Comput 2012;3:42–55. http://dx.doi.org/10.1109/T-AFFC.2011.25.

[21] Candia-Rivera D, Catrambone V, Valenza G. The role of electroencephalography electrical reference in the assessment of functional brain–heart interplay: From methodology to user guidelines. J Neurosci Methods 2021;360:109269. http://dx.doi.org/10.1016/j.jneumeth.2021.109269.

[22] Blair RC, Higgins JJ. The power of t and Wilcoxon statistics: A comparison. Eval Rev 1980;4:645–56. http://dx.doi.org/10.1177/0193841X8000400506.

[23] Guger C, Daban S, Sellers E, Holzner C, Krausz G, Carabalona R, et al. How many people are able to control a P300-based brain–computer interface (BCI)? Neurosci Lett 2009;462:94–8. http://dx.doi.org/10.1016/j.neulet.2009.06.045.

[24] Xu R, Spataro R, Allison BZ, Guger C. Brain–computer interfaces in acute and subacute disorders of consciousness. J Clin Neurophysiol 2022;39:32–9. http://dx.doi.org/10.1097/WNP.0000000000000810.

[25] Candia-Rivera D, Annen J, Gosseries O, Martial C, Thibaut A, Laureys S, et al. Neural responses to heartbeats detect residual signs of consciousness during resting state in postcomatose patients. J Neurosci 2021;41:5251–62. http://dx.doi.org/10.1523/JNEUROSCI.1740-20.2021.

[26] Sassenhagen J, Draschkow D. Cluster-based permutation tests of MEG/EEG data do not establish significance of effect latency or location. Psychophysiology 2019;56:e13335. http://dx.doi.org/10.1111/psyp.13335.