Scorepochs: A Computer-Aided Scoring Tool for Resting-State M/EEG Epochs

Matteo Fraschini [1,*], Simone Maurizio La Cava [1], Giuseppe Rodriguez [2], Andrea Vitale [3] and Matteo Demuru [4]

[1] Department of Electrical and Electronic Engineering, University of Cagliari, Cagliari, Italy
[2] Department of Mathematics and Computer Science, University of Cagliari, Cagliari, Italy
[3] Laboratory for Autism and Neurodevelopmental Disorders, Center for Neuroscience and Cognitive Systems @UniTn, Istituto Italiano di Tecnologia, Rovereto, Italy
[4] Stichting Epilepsie Instellingen Nederland (SEIN), Heemstede, The Netherlands

Abstract: M/EEG resting-state analysis often requires the definition of the epoch length and the criteria to select which epochs to include in the subsequent steps. However, the effects of epoch selection remain scarcely investigated and the procedure used to (visually) inspect, label, and remove bad epochs is often not documented, thus hindering the reproducibility of the reported results. In this study, we present Scorepochs, a simple and freely available tool for the automatic scoring of resting-state M/EEG epochs that aims to provide an objective method to aid M/EEG experts during epoch selection procedure. We tested our approach on a freely available EEG dataset containing recordings from 109 subjects using the BCI2000 64-channel system.

Keywords: M/EEG; epoch selection; resting-state

1. Introduction

Task-free resting-state recording from M/EEG (magneto/electroencephalogram) represents one of the most used experimental paradigms to investigate the baseline level of brain activity in healthy subjects and patients [1]. However, the resting-state condition is an elusive concept which is influenced by different states of vigilance that are usually out of control from the experimenter [2]. Generally, the first steps performed during an M/EEG resting-state analysis consist of (i) segment the raw/filtered EEG traces into a set of non-overlapping epochs and (ii) select a number of artifacts-free epochs to be used during the subsequent steps of the pipeline. These steps require the definition of the epoch length and the criteria to select which epochs to include in the subsequent analysis. The effects of the epoch length have been previously investigated [3], while the effects of epoch selection, induced by the inter-observer variability and unclear criteria used for this task, remain scarcely investigated. Epoch selection is performed at individual level (independently for each subject) and usually is conducted by one or more experts. Some kind of procedure to detect and mitigate EEG artifacts may be applied before this step. However, the precise procedure used to (visually) inspect, label, and remove bad epochs is often not documented [4], thus hindering the reproducibility of the reported results. Most importantly, especially in the case the selection procedure is performed by different experts, it would be of relevance to assure that homogeneous criteria were used. In this context, most of the studies using resting-state paradigms make assumptions on stationarity of EEG signal and perform averaging of individual features (extracted at subject-level) to make inferences at group-level. In short, this means that it is assumed a strong within-subject stability of M/EEG features and that these individual characteristics may be consistent within a group. Subjective visual scoring and/or inter-observer variability poses a possible threat to the validity of these assumptions, although some studies have reported that the subjective influence may lead to minimal changes when a sufficient number of epochs are selected [5,6]. Nevertheless, it is still far to be clear how to quantify this sufficient number of epochs. In this context, it will assume a very important role the possibility

to develop some kind of automated analysis with the aim to help clinicians during this very crucial step. Few recent studies used computer-assisted tool to allow interpreting EEG background patterns [7,8] but, probably, as suggested by van Diessen et. al. [4], these methods were not applied to large scale also because inherent complexity or limited transparency. On the other hand, other studies have suggested the use of automatic artifacts suppression [9] that would be indirectly of help to limit the uncertainty induced by inter-observer variability. Most of these approaches are based on independent component analysis (ICA) [10], which however requires a great amount of EEG data to obtain an acceptable decomposition. In this study, we present Scorepochs, a simple and freely available tool for automatic scoring of resting-state M/EEG epochs that aims to provide an objective method to aid M/EEG experts during epoch selection procedure. Our approach, which works at subject-level, provides a score for each epoch within a single M/EEG recording, with the attempt to make this crucial procedure less ambiguous, more objective, and reproducible. Furthermore, Scorepochs is based on the power spectrum of the EEG, does not require any specific assumption on the underlying frequency content and may thus keep all the relevant spectral information contained in the unfiltered raw signal [11].

2. Methods

The proposed method is based on a very simple and fast algorithm which takes as input (*i*) a set of M/EEG recordings and (*ii*) the length of the desired epoch. After that, the algorithm provides as output a score for each single M/EEG epoch. A schematic representation of the proposed method is depicted in Figure 1. Furthermore, all the scripts used to perform the analysis are freely available for MATLAB (https://github.com/matteogithub/scorepochs) and for Python (https://github.com/smlacava/scorepochs/tree/master/Python). For each subject, each epoch, and each channel, the algorithm computes the power spectral density (PSD) via the Welch method into a specific range of frequencies (see Figure 1, panels A and B). At channel-level, a similarity score, computed by using the Spearman correlation coefficient, is evaluated between the PSD extracted from all the epochs, thus providing a correlation matrix with *number of epochs x number of epochs* as dimension (see Figure 1, panel C). A first average is now computed over the rows (/columns) of the symmetric matrix to obtain a *score vector* with length equal to the *number of epochs*, where the entries represent the mean similarity score of the corresponding epoch (see Figure 1, panel D). Computing the *score vector* for each channel, and then averaging the *score vectors* across channels, it is possible to obtain a final score for each epoch (see Figure 1, panel E). Finally, for each subject, the score can be sorted in descending order allowing to select the suggested epochs to be included in the subsequent steps of the analysis.

We tested our approach on a freely available EEG dataset [12,13] containing recordings from 109 subjects collected using the BCI2000 64-channel system (http://www.bci2000.org). The EEG dataset is available at the following link: https://physionet.org/content/eegmmidb/1.0.0/. We decided to define an easily interpretable hypothetical scenario where the interest was to contrast two different baseline conditions, namely eyes-open (EO) and eyes-closed (EC) resting-state. To contrast the two conditions the relative alpha power (computed in the range between 8 and 13 Hz) was the perfect candidate, as this property is a very common yet effective feature able to detect macroscopic differences between EO and EC conditions. The analysis was performed on 99 out of the 109 subjects, since some of them were excluded due to differences in recording parameters or overall poor quality. We used an epoch length of 5 seconds and thus segmented the one-minute available recordings into twelve non-overlapping epochs (the results were successively replicated using two different epoch lengths, respectively of 2 and 8 seconds). For each epoch and each channel, we extracted the relative alpha power and the average across channels was successively evaluated for each epoch, thus computing the relative alpha power at global level. To mimic a realistic epoch selection procedure and investigate its possible effect, we decided to select for each subject four of the twelve available epochs, considering 495 different selections, representing all the possible combinations obtained by using the same subset of epochs for all the subjects. We then compared the results in terms of magnitude of effect-size obtained contrasting the two conditions (EO vs EC) on a group-level. We computed 495 t-

tests, where for each test we selected 4 epochs for every subject in sequential order (i.e, for all subjects, for the first test the epochs selected were [1, 2, 3, 4], for the second test the epochs selected were [1, 2, 3, 5], for the third test the epoch selected were [1, 2, 3, 6], …, up to the last test where the selected epochs were [9, 10, 11, 12]). We then compared the magnitude of the paired Cohen's d effect-size obtained using the selection suggested by Scorepochs against the distribution of Cohen's d effect-size based on the sequential random selection. The analysis was later replicated using a completely different method, namely the phase lag index (PLI) [14], to compare the two experimental conditions. We performed this further analysis to investigate if the proposed approach may be successfully applied also to different methods.

Finally, we explored a potential application of Scorepochs as an "EEG quality index" for evaluating the impact of alternative preprocessing steps on the same epoched data. In particular, we compared the average score as obtained after different cleaning procedures included in common semi-automated pipelines: cleanraw method [15] or standard deviation threshold for bad channel detection, and the Artifact subspace reconstruction [16] versus the wavelet-enhanced ICA [17] for the correction of artifact-laden portion of data.

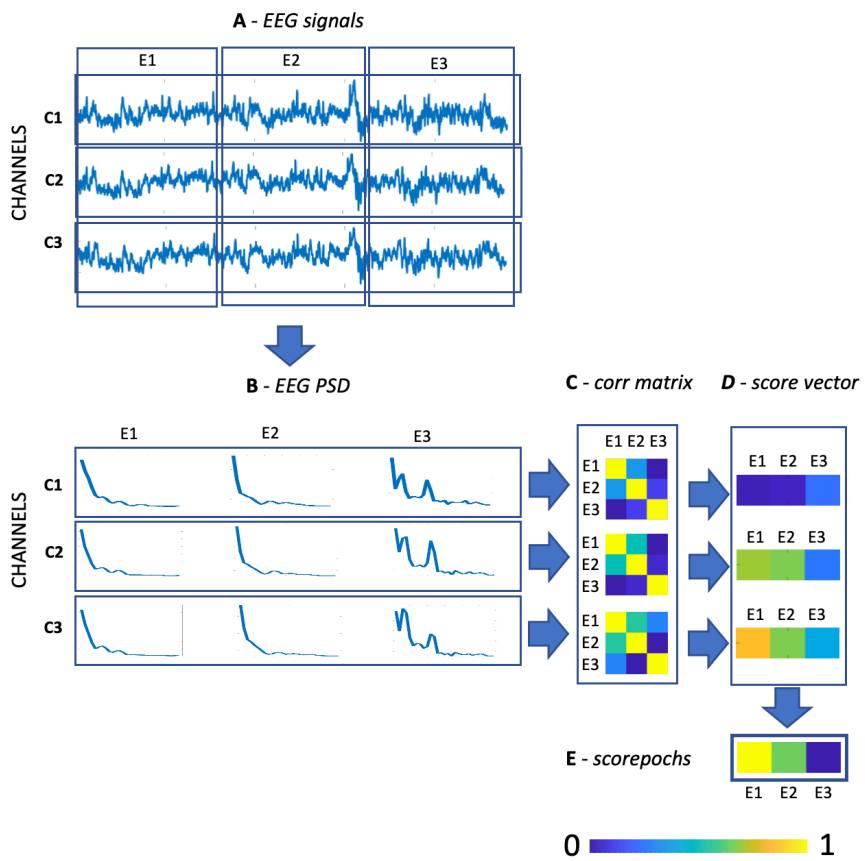

Figure 1. A schematic representation of the algorithm used to compute Scorepochs. Panel A: EEG raw signals of three channels with epoching scheme; Panel B: Power spectral density plots for each channel and each epoch; Panel C: correlation matrices for each channel; Panel D: score vector for each channel and for each epoch; Panel E: score for each epoch.

3. Results

All the results are summarized in Figure 2, which shows the comparison, in terms of Cohen's d effect-size values, between the described sequential random selection and the selection suggested by our approach. In particular, Figure 2, Panel A, depicts the 'effect-size time-course' using this random

selection with the result derived by the application of our method, represented by the green dashed line. A decreasing trend in terms of effect-size values can be observed. Figure 2, Panel B, shows the distribution of the effect-size values (independently of the sequential order), where the vertical green dashed line represents the value of the effect-size obtained using the epochs suggested by our approach. The Cohen's d value for the Scorepochs algorithm is 1.4512, which is around the 75th percentile of the random epoch selection distribution. It is worth noting that the minimum effect-size is bigger than 1, meaning that, as expected, we are in a situation where the difference is largely independent of the epoch selection strategy (i.e., the difference between the two conditions EO and EC is reliably measurable).

With the aim to investigate the possible effects of the time-window on the reported results, we reproduced all the analysis using two different epoch lengths of 2 and 8 seconds. The results obtained from this new analysis are represented in Figure 3 and Figure 4 respectively.

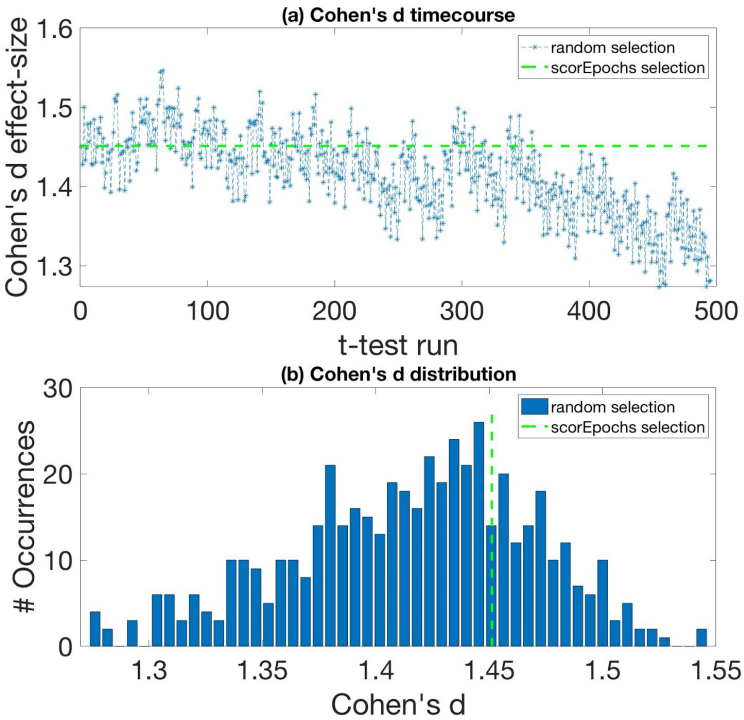

Figure 2. Cohen's d effect-size for random selection and Scorepoch selection using a time window equal to 5 seconds. (a) shows the 'time course' of the effect-size computed using a sequential random selection. The effect-size values are reported in the y-axis, while the x-axis indicates the t-test with a sequential random selection. The green dashed line represents Cohen's d value obtained by selecting the epochs using Scorepochs. (b) Cohen's d effect-size distribution for random epoch selection and the Scorepoch selection. The effect-size values are reported on the x-axis, while the y-axis indicates the occurrences of specific effect-size values. The vertical green dashed line represents Cohen's d value selecting the 4 epochs suggested by Scorepochs.

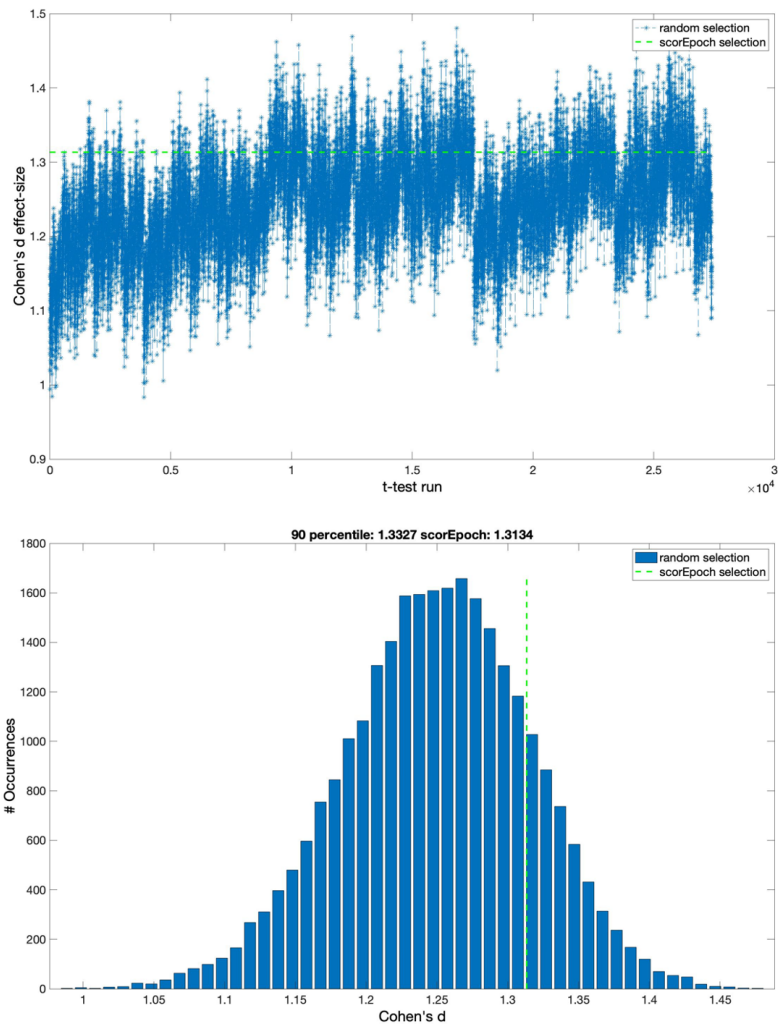

Figure 3. Cohen's d effect-size for random selection and Scorepoch selection using a time window equal to 2 seconds. (a) shows the 'time course' of the effect-size computed using a sequential random selection. The effect-size values are reported in the y-axis, while the x-axis indicates the t-test with a sequential random selection. The green dashed line represents Cohen's d value obtained by selecting the epochs using Scorepochs. (b) Cohen's d effect-size distribution for random epoch selection and the Scorepoch selection. The effect-size values are reported on the x-axis, while the y-axis indicates the occurrences of specific effect-size values. The vertical green dashed line represents Cohen's d value selecting the 4 epochs suggested by Scorepochs.

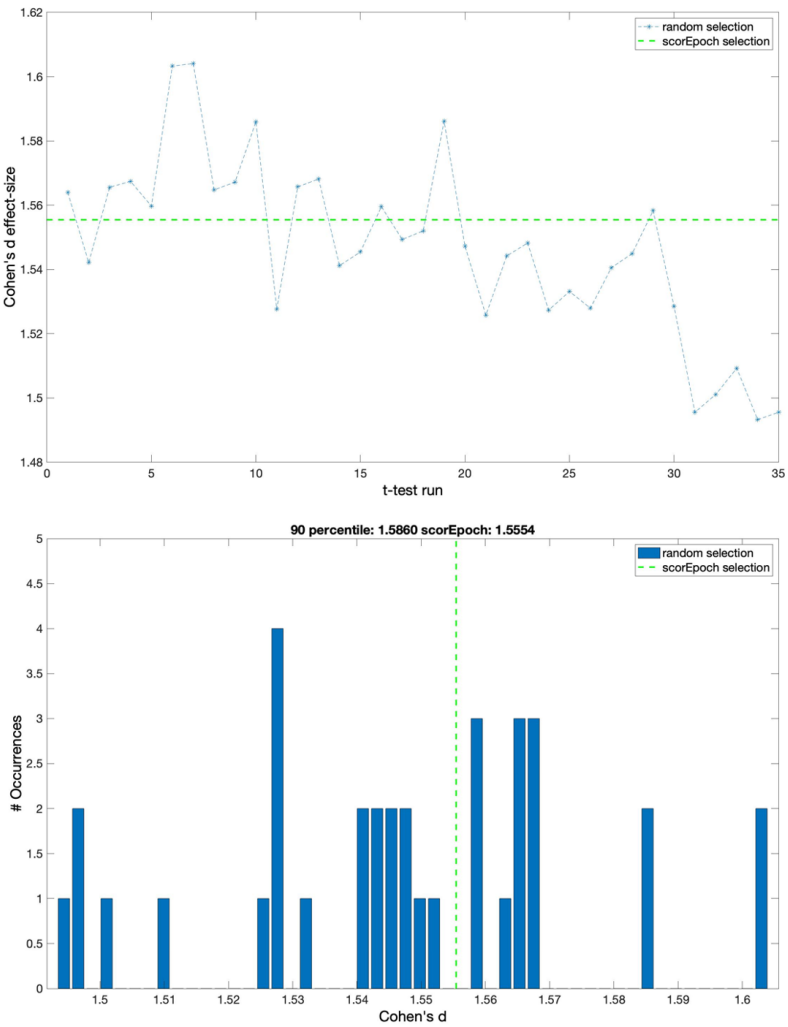

Figure 4. Cohen's d effect-size for random selection and Scorepoch selection using a time window equal to 8 seconds. (a) shows the 'time course' of the effect-size computed using a sequential random selection. The effect-size values are reported in the y-axis, while the x-axis indicates the t-test with a sequential random selection. The green dashed line represents Cohen's d value obtained by selecting the epochs using Scorepochs. (b) Cohen's d effect-size distribution for random epoch selection and the Scorepoch selection. The effect-size values are reported on the x-axis, while the y-axis indicates the occurrences of specific effect-size values. The vertical green dashed line represents Cohen's d value selecting the 4 epochs suggested by Scorepochs.

The results derived from the application of the PLI method are summarized in Figure 5 (a and (b.

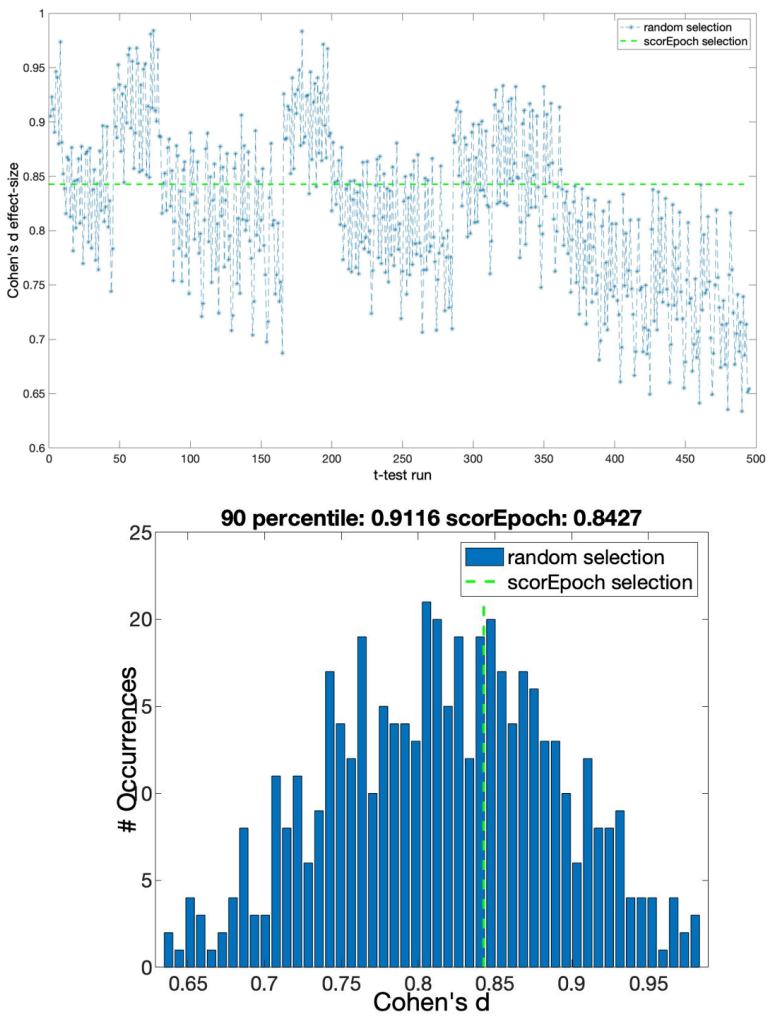

Figure 5. Cohen's d effect-size for random selection and Scorepoch selection using a time window equal to 5 seconds for the PLI method. (a) shows the 'time course' of the effect-size computed using a sequential random selection. The effect-size values are reported in the y-axis, while the x-axis indicates the t-test with a sequential random selection. The green dashed line represents Cohen's d value obtained by selecting the epochs using Scorepochs. (b) Cohen's d effect-size distribution for random epoch selection and the Scorepoch selection. The effect-size values are reported on the x-axis, while the y-axis indicates the occurrences of specific effect-size values. The vertical green dashed line represents Cohen's d value selecting the 4 epochs suggested by Scorepochs.

Finally, the application of Scorepochs as a metrics for evaluating the impact of alternative preprocessing steps at different levels of a pipeline are reported in Figure 6.

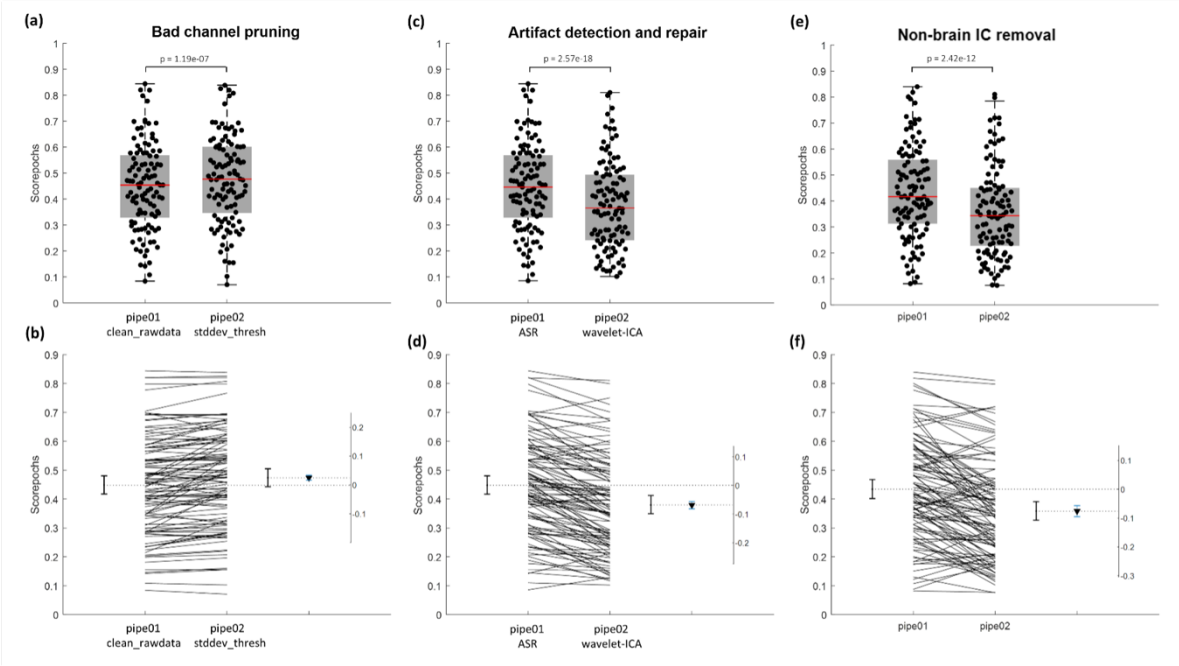

Figure 6. Scorepochs as an "EEG quality metrics" to evaluate and compare the impact of alternative preprocessing steps. (a) The same epoched data (5 seconds length) from the eyes-open condition were submitted to two different methods for bad (flat or noisy) channel detection: cleanraw algorithm [15] based on the correlation with nearby channels (pipe01) or >3 standard deviations from the mean (pipe02). The score values for the alpha frequency band (8-13 Hz) averaged at individual level are reported and (b) an estimation graphic was computed to investigate the direction and size of the effect [18]. (c) Comparison between two common approaches for artifact detection and repair: Artifact Subspace Reconstruction [16] (pipe01) versus wavelength-enhanced ICA [17] (pipe02). Paired-sample t-test as well as (d) the size of the confidence intervals shows a more effective/efficient performance obtained by ASR for the repair of artifact-laden data segment. (e-f) Scorepochs values were compared at the end of the preprocessing procedure after a final step of independent component analysis and the removal of ICs classified as of unlikely brain source (< 20% according to the ICLabel algorithm [19]). A similar impact of the preprocessing steps was reproduced by considering a shorter epoch length (2 seconds) and an extended power spectrum (1-80 Hz), with the exception of the bad channel detection where none of the two alternative methods had a significant influence on the Scorepochs results.

4. Discussion

In this study, we propose an automatic method to assist M/EEG experts during the epoch selection procedure in resting-state analysis. Our method represents an objective approach (or less subjective) to perform epochs selection, if compared to the potential arbitrariness introduced by human observers and the lack of clear and shared criteria used to accomplish this crucial task. We showed, in a prototypical scenario of a group comparison between two resting-state conditions (EO vs EC), that the effect-size varied extensively depending on the epochs included in the analysis. In fact, even if it is possible to detect an effect between the two conditions EO and EC almost independently of the selected epochs (since the detection of an effect is highly probable as Cohen's d > 1 for every epoch selection), there is a considerable variation on the effect-size depending on the actual selection. Specifically, we showed a decreasing trend (see Figure 2, Panel A) in terms of the effect-size with respect to the time in which the selection occurred (i.e., selecting epochs at the beginning of the experiment gives a higher effect-size compared to the end of the experiment). This trend may be influenced by different states of vigilance related to tiredness or drowsiness that are

reflected in the recorded signals. The magnitude of the effect-size obtained using our proposed epoch selection is close to the mean effect-size (which should represent the best estimate of the population effect-size) and, more importantly, is based on a quantifiable, objective, and replicable strategy (i.e., scores computed on the PSD). It is also relevant to highlight that our results successfully replicated using different size of time windows. Finally, our results suggest that the proposed approach may be easily extended to other methods as the one based on connectivity metrics. Compared to other semi-automatic procedures for the selection of the 'best artefact-free epochs' suitable for the analysis (e.g., independent component analysis, ICA), our method is completely data-driven, and it does not require any intervention or particular skills of the user, compared to other selection strategies (e.g., knowledge of stereotypical EEG pattern related to artefact components using ICA). Moreover, the Scorepochs method, because of the small number of requirements (i.e., computation of the PSD), is not computationally expensive. Although its simplicity, this method is well grounded in physiological terms. In fact, it has been observed how the computation of simple statistics based on the PSD reflects intrinsic properties of excitatory/inhibitory levels of neuronal populations [20]. Furthermore, the PSD is able to capture different dynamics modulated by external stimuli and provides insight into sensory neural representation [21]. Finally, recently it has been reported how different behavioral states are reflected in different properties of the PSD [22]. In no way, the proposed approach aims to replace the work to be performed by experts (alone or using other automatic or semi-automatic methods) during visual inspections of real M/EEG data. Scorepochs guided selection should be complementary to the human activity or to any other selection method.

Acknowledgments: The authors have no conflict of interest to disclose. Simone Maurizio La Cava was supported by Regione Autonoma della Sardegna POR FESR Sardegna 2014-2020 Asse 1 Azione 1.2.2 CC F26C18000500006. Giuseppe Rodriguez and Matteo Fraschini were in part supported by Regione Autonoma della Sardegna research project "Algorithms and Models for Imaging Science [AMIS]", FSC 2014-2020 - Patto per lo Sviluppo della Regione Sardegna.

References

1. Stam, C.J. Modern Network Science of Neurological Disorders. *Nature Reviews Neuroscience* **2014**, *15*, 683–695, doi:10.1038/nrn3801.

2. Diaz, B.A.; Van Der Sluis, S.; Moens, S.; Benjamins, J.S.; Migliorati, F.; Stoffers, D.; Den Braber, A.; Poil, S.-S.; Hardstone, R.; Van't Ent, D.; et al. The Amsterdam Resting-State Questionnaire Reveals Multiple Phenotypes of Resting-State Cognition. *Front Hum Neurosci* **2013**, *7*, doi:10.3389/fnhum.2013.00446.

3. Fraschini, M.; Demuru, M.; Crobe, A.; Marrosu, F.; Stam, C.J.; Hillebrand, A. The Effect of Epoch Length on Estimated EEG Functional Connectivity and Brain Network Organisation. *J Neural Eng* **2016**, *13*, 036015, doi:10.1088/1741-2560/13/3/036015.

4. van Diessen, E.; Numan, T.; van Dellen, E.; van der Kooi, A.W.; Boersma, M.; Hofman, D.; van Lutterveld, R.; van Dijk, B.W.; van Straaten, E.C.W.; Hillebrand, A.; et al. Opportunities and Methodological Challenges in EEG and MEG Resting State Functional Brain Network Research. *Clinical Neurophysiology* **2015**, *126*, 1468–1481, doi:10.1016/j.clinph.2014.11.018.

5. Douw, L.; de Groot, M.; van Dellen, E.; Aronica, E.; Heimans, J.J.; Klein, M.; Stam, C.J.; Reijneveld, J.C.; Hillebrand, A. Local MEG Networks: The Missing Link between Protein Expression and Epilepsy in Glioma Patients? *NeuroImage* **2013**, *75*, 195–203, doi:10.1016/j.neuroimage.2013.02.067.

6. Olde Dubbelink, K.T.E.; Hillebrand, A.; Stoffers, D.; Deijen, J.B.; Twisk, J.W.R.; Stam, C.J.; Berendse, H.W. Disrupted Brain Network Topology in Parkinson's Disease: A Longitudinal Magnetoencephalography Study. *Brain* **2014**, *137*, 197–207, doi:10.1093/brain/awt316.

7. Anderson, N.R.; Doolittle, L.M. Automated Analysis of EEG: Opportunities and Pitfalls. *Journal of Clinical Neurophysiology* **2010**, *27*, 453–457, doi:10.1097/WNP.0b013e3181fe0b6f.

8. Lodder, S.S.; Askamp, J.; Putten, M.J.A.M. van Computer-Assisted Interpretation of the EEG Background Pattern: A Clinical Evaluation. *PLOS ONE* **2014**, *9*, e85966, doi:10.1371/journal.pone.0085966.

9. Jiang, X.; Bian, G.-B.; Tian, Z. Removal of Artifacts from EEG Signals: A Review. *Sensors (Basel)* **2019**, *19*, doi:10.3390/s19050987.

10. Groppe, D.M.; Makeig, S.; Kutas, M. Identifying Reliable Independent Components via Split-Half Comparisons. *NeuroImage* **2009**, *45*, 1199–1211, doi:10.1016/j.neuroimage.2008.12.038.

11. Donoghue, T.; Haller, M.; Peterson, E.J.; Varma, P.; Sebastian, P.; Gao, R.; Noto, T.; Lara, A.H.; Wallis, J.D.; Knight, R.T.; et al. Parameterizing Neural Power Spectra into Periodic and Aperiodic Components. *Nature Neuroscience* **2020**, *23*, 1655–1665, doi:10.1038/s41593-020-00744-x.

12. Goldberger Ary L.; Amaral Luis A. N.; Glass Leon; Hausdorff Jeffrey M.; Ivanov Plamen Ch.; Mark Roger G.; Mietus Joseph E.; Moody George B.; Peng Chung-Kang; Stanley H. Eugene PhysioBank, PhysioToolkit, and PhysioNet. *Circulation* **2000**, *101*, e215–e220, doi:10.1161/01.CIR.101.23.e215.

13. Schalk, G.; McFarland, D.J.; Hinterberger, T.; Birbaumer, N.; Wolpaw, J.R. BCI2000: A General-Purpose Brain-Computer Interface (BCI) System. *IEEE Transactions on Biomedical Engineering* **2004**, *51*, 1034–1043, doi:10.1109/TBME.2004.827072.

14. Stam, C.J.; Nolte, G.; Daffertshofer, A. Phase Lag Index: Assessment of Functional Connectivity from Multi Channel EEG and MEG with Diminished Bias from Common Sources. *Human Brain Mapping* **2007**, *28*, 1178–1193, doi:10.1002/hbm.20346.

15. Kothe, C.A.; Makeig, S. BCILAB: A Platform for Brain–Computer Interface Development. *J. Neural Eng.* **2013**, *10*, 056014, doi:10.1088/1741-2560/10/5/056014.

16. Mullen, T.R.; Kothe, C.A.E.; Chi, Y.M.; Ojeda, A.; Kerth, T.; Makeig, S.; Jung, T.-P.; Cauwenberghs, G.

Real-Time Neuroimaging and Cognitive Monitoring Using Wearable Dry EEG. *IEEE Transactions on Biomedical Engineering* **2015**, *62*, 2553–2567, doi:10.1109/TBME.2015.2481482.

17. Castellanos, N.P.; Makarov, V.A. Recovering EEG Brain Signals: Artifact Suppression with Wavelet Enhanced Independent Component Analysis. *Journal of Neuroscience Methods* **2006**, *158*, 300–312, doi:10.1016/j.jneumeth.2006.05.033.

18. Ho, J.; Tumkaya, T.; Aryal, S.; Choi, H.; Claridge-Chang, A. Moving beyond P Values: Data Analysis with Estimation Graphics. *Nat Methods* **2019**, *16*, 565–566, doi:10.1038/s41592-019-0470-3.

19. Pion-Tonachini, L.; Kreutz-Delgado, K.; Makeig, S. ICLabel: An Automated Electroencephalographic Independent Component Classifier, Dataset, and Website. *NeuroImage* **2019**, *198*, 181–197, doi:10.1016/j.neuroimage.2019.05.026.

20. Gao, R.; Peterson, E.J.; Voytek, B. Inferring Synaptic Excitation/Inhibition Balance from Field Potentials. *NeuroImage* **2017**, *158*, 70–78, doi:10.1016/j.neuroimage.2017.06.078.

21. Barbieri, F.; Mazzoni, A.; Logothetis, N.K.; Panzeri, S.; Brunel, N. Stimulus Dependence of Local Field Potential Spectra: Experiment versus Theory. *J. Neurosci.* **2014**, *34*, 14589–14605, doi:10.1523/JNEUROSCI.5365-13.2014.

22. Zilio, F.; Gomez-Pilar, J.; Cao, S.; Zhang, J.; Zang, D.; Qi, Z.; Tan, J.; Hiromi, T.; Wu, X.; Fogel, S.; et al. Are Intrinsic Neural Timescales Related to Sensory Processing? Evidence from Abnormal Behavioral States. *NeuroImage* **2021**, *226*, 117579, doi:10.1016/j.neuroimage.2020.117579.