Clinical Neurophysiology 174 (2025) 61–72

Contents lists available at ScienceDirect

Clinical Neurophysiology

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

Estimating sensor-space EEG connectivity PART 2: Identifying optimal artifact reduction techniques for functional connectivity in real data

Aleksandra Miljevic [a,*], Oscar W. Murphy [a,b], Paul B. Fitzgerald [a,c], Neil W. Bailey [a,c]

[a] Department of Psychiatry, Central Clinical School, Monash University, Melbourne, VIC, Australia
[b] Bionics Institute, Melbourne, VIC, Australia
[c] School of Medicine and Psychology, Australian National University, Canberra, ACT, Australia

ARTICLE INFO

Keywords:
EEG
Functional connectivity
FC
Lifespan
Alpha frequency
Electroencephalography

ABSTRACT

Objectives: Electroencephalography (EEG) can be used to assess functional brain connectivity (FC). However, there is considerable variability in the methods used for FC measurement across different studies, which may contribute to heterogeneity in research outcomes. We aimed to assess how different EEG pre-processing steps impact EEG-FC measurement when applied to real EEG data.

Methods: Using the BrainClinics.com open-source EEG data repository we investigated how different pre-processing steps impacted the ability to detect age-related differences in alpha band FC and the test–retest reliability of FC measures. The pre-processing steps tested included artifact reduction techniques (Independent Component Analysis (ICA), wavelet-enhanced ICA (wICA), and Multi-channel Wiener Filters (MWF)), different epoch lengths (epochs that were 2 s versus 6 s in length), and different re-referencing montages (the common average reference (CAR) versus current source density (CSD) re-referencing). We also assessed different FC metrics including imaginary coherence (iCOH), real magnitude squared coherence (rMSC), and weighted phase lag index (wPLI) metrics.

Results: The best performing pipeline at detecting age-related differences in alpha FC and providing high test–retest reliability included artifact reduction by ICA or wICA, data re-referenced using the CSD method, and FC measured by rMSC.

Conclusion & significance: This paper presents evidence for an EEG pre-processing pipeline that provides good detection of meaningful effects and high test–retest reliability for sensor space EEG alpha frequency FC.

1. Introduction

Electroencephalography (EEG) enables the detection and analysis of electrical signals produced by networks of neuronal populations in the brain (Panteliadis, 2021). The relationships between signals recorded by individual EEG electrodes can be evaluated to provide information relating to the brain's *functional* connectivity (FC). FC is defined as the statistically significant synchronous activity between neural signals (recorded by scalp electrodes above certain brain regions in the case of EEG) (Cohen, 2014).

The recorded EEG signal represents the difference between electrical potentials obtained from the active electrodes, relative to a "reference" electrode (Hagemann et al., 2001; Yao, 2001; Da Silva, 1999). This raw EEG signal is comprised of the signal generated by brain activity, as well as noise (non-brain) signals from the participant and the external recording environment (for example, muscle movements, eye blink, circadian rhythm, and environmental electrical noise). Electrical signals from both brain activity and artifacts travel through the brain and scalp via volume conduction and can reach many EEG electrodes with a near-instantaneous delay due to the conductibility of brain tissue (van Diessen et al., 2015). While the brain contains monosynaptic connections that can produce zero-lag connectivity (Sabatini and Regehr, 1999), given the current EEG methods and measures available, it is not possible to differentiate these zero-lag connections from the effects of volume conduction. As such, if methods to protect against the effects of these volume conducted signals are not implemented, then high FC estimates can be obtained because of volume conduction, without meaningful FC being present (Miljevic et al., 2025). Therefore, to reliably

* Corresponding author at: Epworth Centre for Innovation in Mental Health, Epworth HealthCare, 888, Toorak Rd, Camberwell, Victoria 3124, Australia.
E-mail addresses: aleksandra.miljevic@monash.edu (A. Miljevic), omurphy@bionicsinstitute.org (O.W. Murphy), paul.fitzgerald@monash.edu (P.B. Fitzgerald), neil.bailey@monash.edu (N.W. Bailey).

https://doi.org/10.1016/j.clinph.2025.03.042
Accepted 29 March 2025
Available online 8 April 2025
1388-2457/

estimate FC from the EEG signal, the raw data must be (pre-)processed to remove artificial, non-brain information (Bailey et al., 2023a), and steps to protect against the inflation of FC estimates due to volume conduction are critical (Miljevic et al., 2025).

However, (pre-)processing and later analytical steps have the potential to significantly affect EEG FC estimation (see Miljevic et al., 2023 for a detailed discussion of these issues). The key pre-processing steps that hold the greatest potential to affect the assessment of EEG FC include (1) the choice of re-referencing montage, where during pre-processing, the EEG data from active electrodes are usually re-referenced to an alternative reference scheme (for example, the average signal from all available electrodes), (2) the length of EEG data segments or epochs used to estimate FC, and (3) the FC metric used to estimate FC. These FC metrics can include measures that only assess commonalities in phase shifts to estimate FC, for example, weighted phase lag index (wPLI) and imaginary coherence (iCOH), as well as FC measures that assess temporal commonalities in variation in frequency power, for example magnitude squared coherence (rMSC) (Miljevic et al., 2022). In Part 1 of our connected studies (Miljevic et al., in review), we utilized simulated EEG data to test the effects of different pre-processing steps on FC estimation in simulated data (where the ground truth FC was known). Our results showed that pipelines that included 40 or more epochs that were > 6 s (s) in length enabled the best ability to provide high FC values when physiologically plausible FC was simulated, and low FC values when only volume conduction-like connectivity was simulated. Our results also indicated that FC detection was better when these epoch lengths were combined with 1) a Reference Electrode Standardization Technique (REST) or common average reference (CAR) re-reference and the use of imaginary coherence (iCOH) or weighted phase lag index (wPLI) FC metrics, or 2) the current source density (CSD) re-reference with the rMSC FC metric (Miljevic et al., 2024).

Despite the utility of these tests of simulated data (where the ground-truth FC is known), another important (pre-)processing step that has the potential to significantly alter the EEG FC estimate was not addressed: the reduction of artifacts in the EEG data. Many different approaches have been used to remove artifacts from EEG data. Currently, the most commonly implemented approach is to use Independent Component Analysis (ICA) to decompose data into statistically independent components, followed by the removal of components deemed to be artifacts, before reconstructing data back into the scalp space (Iriarte et al., 2003, Pion-Tonachini et al., 2019). ICA iteratively models putative source signals within the raw EEG to obtain statistically independent source components, commonly using distribution entropy or kurtosis to estimate the independence of source components (Pion-Tonachini et al., 2019). ICA assumes that recorded signals result from a linear combination (summed voltage) of independent sources, both neural and non-neural. By decomposing EEG data into these putative source components, ICA aims to separate neural activity from artifacts, such as eye blinks or muscle movements. The process involves applying the ICA decomposition, labelling each component based on its temporal and spatial characteristics, and then selectively subtracting components labelled as artifacts before reconstructing data back to the source space. As such, ICA acts as a spatial filter to remove artifacts from the data. For more information on the ICA technique and its applicability to EEG data cleaning see Iriarte et al. (2003) or Stone (2002).

ICA has been shown to effectively remove artifacts, but some research has also suggested that ICA may also remove neural signals, which can paradoxically inflate FC estimates. This is because the removal of artifacts within the component space necessitates the removal of a common signal contribution to multiple electrodes, artificially producing a common shift in signals across multiple electrodes (Castellanos and Makarov, 2006, Bailey et al., 2024). To address this potential issue, Wavelet-enhanced ICA (wICA) applies wavelet transforms to artifact components from the ICA decomposition prior to their removal. This is proposed to enable a time–frequency representation of the only artifact data to be removed (Castellanos and Makarov, 2006). This is particularly beneficial for analyzing EEG signals with non-stationary characteristics (where the frequency power spectrum changes over time). By offering time–frequency based artifact removal, wICA is believed to perform better at identifying and separating artifact signals from neural activity in comparison to ICA which decomposes data based only on spatial relationships (Bailey et al., 2023a). This has been suggested to protect against the inflated FC estimate suggested to be caused by ICA (Castellanos and Makarov, 2006). Additionally, it is suggested that ICA subtraction may invalidate FC estimates by reducing the rank of the data prior to connectivity analysis, whereas wICA does not (Castellanos and Makarov, 2006). For more information on wICA and its applicability to EEG data cleaning see Inuso et al. (2007) or Castellanos and Makarov (2006).

Recent research has shown the Multi-channel Wiener Filter (MWF) to be an effective artifact reduction technique for EEG (Somers et al., 2018, Bailey et al., 2023a). The MWF approach treats EEG signals at each electrode as a linear combination of the desired brain signal and unwanted noise. Based on this assumption, MWF uses delay embedded (i.e. stacked time-lagged) artifact and non-artifact periods from the EEG data as templates to construct artifact and non-artifact covariance matrices, enabling estimation of spatio-temporal weightings representing the multi-channel artifact signal (based on optimisation of the minimum mean squared error). A generalized eigenvector decomposition is then performed on the artifact covariance matrices, and positive rank components are selected as spatio-temporal sources for removal (as these components are likely to provide the best model of the artifact). This process enables construction of a spatio-temporal filter that can be applied to the delay embedded version of the original data to reduce artifacts prior to reconstruction of the scalp space data cleaned of artifacts (see Somers et al., 2018 for the full details of this approach). While ICA decomposes data purely based on spatial information, MWF incorporates temporal information and has therefore been suggested to offer advantages for artifact reduction prior to FC estimation (Bailey et al., 2024). As far as we are aware, the effects of the MWF cleaning approach on FC estimation has not yet been systematically tested.

1.1. Study aims and predictions

The overarching aim of the present study was to test for the highest performing combination of EEG pre-processing steps (i.e., a pipeline) out of likely candidate pipelines for robust identification of EEG FC in sensor-space. The aim was assessed through two primary comparison approaches. Firstly, we aimed to evaluate how FC estimates from real data are affected by a range of pre-processing steps, including the length of epochs used to estimate FC, the re-referencing technique used, and the artifact cleaning technique applied, as well as how these pre-processing steps interacted with three common FC metrics: wPLI, rMSC, and iCOH. Ideally, this aim would be addressed using data where the "ground truth" connectivity signal was known, and our measures would assess how accurately each approach estimated this ground truth connectivity. However, while it is possible to simulate EEG-like artifacts, simulated EEG artifacts are not truly reflective of the variability and uniqueness of artifacts obtained from real EEG recordings, and simulations do not replicate the interconnection between artifacts and neural activity (for example, blink artifacts are concurrent with the neural activity that both produces and reacts to the blink). Therefore, to achieve this aim we used a large open-source repository of real EEG datasets. Within this dataset, we examined the ability of EEG FC pre-processing approaches to detect differences in FC with age which have been reliably reported in previous literature, namely the increases in alpha FC reported during the transition from childhood to early adulthood (Chung et al., 2022; Kavčič et al., 2023; Michels et al., 2013), and the reductions in alpha FC reported in older adults (Scally et al., 2018; Smit et al., 2016, Duffy et al., 1996, Vysata et al., 2014). We applied a simple support vector machine (SVM) learning algorithm to investigate how well each of the investigated FC pipelines performed at classifying different age groups based on frontal alpha connectivity, since effective FC estimation should provide FC values that allow well-performing classification of younger and older adults based on differences in FC across the lifespan.

Secondly, we aimed to examine the test–retest reliability of FC estimates associated with the application of different pre-processing steps. Previous research indicates high test–retest reliability for FC measures in resting-state scalp-based EEG recordings across a range of timeframes (i.e., 1 week–3 months) (Popov et al., 2023; Rolle et al., 2022). However, the test–retest reliability of EEG FC has not been compared across different EEG pre-processing approaches. Our rationale for these tests was that an optimal FC estimation method should detect meaningful effects in real data (for example differences in FC with age) and be consistent within individuals between repeated testing sessions weeks apart. A subset of the data was identified within which two EEG recordings from each individual were available, with recordings obtained at least one week apart. To provide maximal utility, a pipeline used for FC estimation should provide high test–retest reliability, thus, the pipeline providing the highest test–retest reliability was assumed to reflect the best pipeline for assessing FC out of the likely candidate pipelines.

2. Methods

2.1. Participants

Our investigations were conducted using the large opensource repository of EEG datasets called 'TD-BRAIN' available through Brain-Clinics (https://brainclinics.com/resources/; van Dijk et al. 2022). The full TD-BRAIN dataset consists of 1274 datasets from participants of varying ages, backgrounds, and health histories. Most of the sample consists of clinical populations with psychiatric illnesses including major depression, attentional deficit disorder, and obsessive–compulsive disorder. For a full breakdown of sample characteristics, see Supplementary Materials 1: S1. A total of 73 datasets were excluded for the purposes of the current analyses, primarily from participants with childhood or severe adulthood neurological or neurodegenerative disorders or traumas. This included participants with an acquired brain injury, a tumour, with autism spectrum disorder, Tourette Syndrome, Asperger's, epilepsy, Parkinson's, and stroke, as well as other conditions.

Of the remaining 1201 datasets, a further 14 were excluded due to a lack of age information, and 13 excluded due to too few epochs remaining for connectivity analysis (<10 epochs) which would result in inaccurate connectivity estimates. Overall, a total of 1174 datasets were used for our analyses of age-related FC, with an age range of 5 to 83 years of age. These participants were categorized into four age groups: children $\leq$ 18, young adults 19–35, adults 36–59, and older adults $\geq$ 60 years of age. A breakdown of key participant characteristics is provided in Table 1.

Our investigations of test–retest reliability were conducted on a subset of the TD-BRAIN EEG open-source data repository where two EEG recordings from the same individual were available, recorded more than one week apart. Of the full 1274 datasets, a total of 69 individuals attended two EEG recording sessions at separate timepoints. Of these 69 individuals, three were excluded as their second EEG dataset was unable to be located within the broader EEG database, two were excluded due to having a Parkinson's disease diagnosis, and two were excluded due to information about their age being absent from the dataset. Finally, seven participants were excluded due to being under the age of 18 years – an age limit was set for the test–retest tests to account for the more rapid changes that occur in the developing brain compared to the adult brain (Hill et al., 2022, 2023). Thus, a total of 55 participants were included in our analyses of test–retest reliability of FC from each of our pipelines. The average age of participants was 49.65 years, with a range from 19.24 to 83.31 years of age, and 41.82 % of participants were female. The time between session 1 and 2 varied from one week to one month.

Table 1
TD-BRAIN participant characteristics.

Age Groups	Female/Male	Number of participants in group
Children (Group 1)		
18 years or younger	69.74 % F	228
Young Adults (Group 2)		
19–35 years of age	50.86 % F	291
Adults (Group 3)		
36–59 years of age	48.33 % F	509
Older Adults (Group 4)		
60 years or older	49.32 % F	146
TOTAL	53.24 % F	1174

Note. No significant differences in FC between gender is present in Group 1, see discussion for details. *Abbreviations:* F = female.

2.2. Data Acquisition

EEG data was recorded using BrainVision Ag/AgCl cap with a total of 26 recording electrodes, with linked mastoids used as the online reference. EEG data were recorded for 2 min (+/- 5 s) across all individuals using both eyes open and eyes closed resting state conditions. Here, we focused only on the analysis of the resting state, eyes closed conditions. All data was sampled at 500 Hz, with a 50 Hz power line frequency (which was notch filtered out during the pre-processing steps described in the next section). When downloaded from the BrainClinics repository, all data was of a continuous nature.

2.3. Data pre-processing

The aim of this paper was to assess the effects of various artifact cleaning techniques on EEG FC estimation. As such, EEG pre-processing involved first rejecting noisy channels and epochs according to the default settings provided in the RELAX automated EEG pre-processing pipeline, which are based on typical approaches applied in the field (Bailey et al., 2023a, Bailey et al., 2023b). The following three artifact cleaning pipelines were then applied to enable comparison of their performance in estimating FC: (1) ICA with artifact component subtraction, with artifact components automatically classified using ICLabel, a machine learning algorithm trained to match expert identification of ICA components (Pion-Tonachini et al., 2019); (2) wICA, with artifact components automatically classified using ICLabel (Pion-Tonachini et al., 2019); and (3) MWF. The RELAX automated pipeline and toolbox provides the ability to select and automatically apply each one of these three commonly used pre-processing approaches individually using automated approaches that are commonly used in research (Bailey et al., 2023a), hence RELAX was used to clean the EEG data according to these three techniques (ICA, wICA, and MWF). For a more detailed account of these artifact reduction techniques, see Supplementary Materials 1: S2.

The cleaned data were then epoched into 2 and 6 s segments and re-referenced according to the common average reference (CAR) technique or the Current Source Density (CSD) technique (also known as the surface Laplacian). These are steps indicated to perform well at EEG FC identification in our adjoining paper (Part 1; Miljevic et al., 2025). These steps provided a minimum of thirty 2 s epochs from all participants that were included in our analysis. Due to the length of the TD-BRAIN EEG recordings, fewer epochs were available compared to the numbers of epochs that have been suggested to provide better performance by our companion article, where > 40 epochs were suggested to provide better FC estimation (Part 1; Miljevic et al., 2025). As per the guidelines and validation steps presented in Miljevic et al. (2023), we encourage future researchers to record longer segments of resting EEG data.

Fig. 1 depicts the data (pre-)processing pipelines employed in this study.

All pipelines were applied in our tests of differences between age groups to evaluate EEG FC and how the various cleaning steps may affect

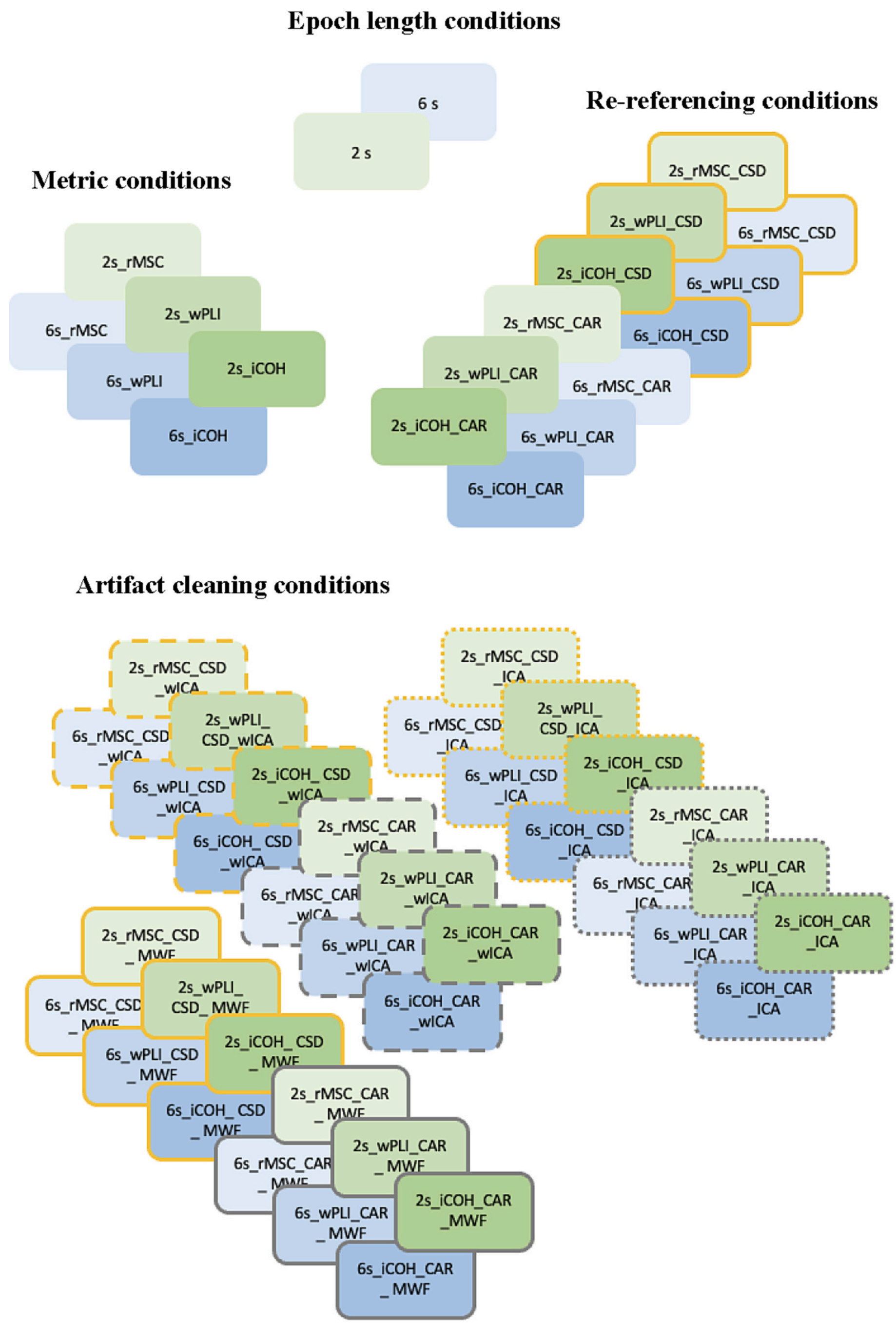

Fig. 1. Visual depiction of pipeline (pre-)processing steps and comparisons. *Note.* 2 s/6 s = two/six second epoch length; CAR = common average reference; CSD = current source density; ICA = independent component analysis; ICC = intraclass correlation coefficient; iCOH = imaginary coherence; MWF = multiple Wiener filters; N = number of participants; rMSC = real magnitude squared coherence; wICA = wavelet-enhanced ICA; wPLI = weighted phase lag index.

FC measurement in relation to the 'ground truth' that age-related differences in alpha FC were expected across the lifespan.

2.4. Functional connectivity estimation

We assessed alpha frequency FC for each combination of epoch length, artifact reduction approaches, and re-referencing montages. To assess FC, we used the commonly applied rMSC, iCOH, and wPLI metrics. These metrics all provide a measure of the strength of FC between two EEG signals, with values ranging from 0 to 1. A value closer to 1 indicates stronger FC while a value closer to 0 indicates weaker FC. For a detailed description of these FC metrics and methods, we refer the reader to the companion article Miljevic et al. (in submission).

FC was averaged across all values obtained from the 10 × 10 frontal electrode pairs available within the dataset prior to the ANOVA and test–retest reliability assessments. The 10 individual electrodes included in these averages were AFp1, AFp2, F9, F3, Fz, F4, F10, FC3, FCz, and FC4.

2.5. Design and statistical analysis

First, we examined which cleaning pipelines provided high performance at identifying EEG FC differences with age, with the expectation that alpha FC would increase in frontal regions from childhood to adulthood and decrease with increasing age in adulthood. All statistical analyses were conducted using JASP software version 0.17.1.0 (https://jasp-stats.org/). Scikit-Learn, a toolbox implemented in Python, was used for machine learning classification (Pedregosa et al., 2011).

For our tests of the differences in FC between age groups, the independent variable was age group, and the dependent variables were the different average alpha frequency FC metrics across all frontal electrodes produced by the different cleaning pipelines (as depicted in Fig. 1). A total of 36 independent one-way analysis of variance (ANOVA) tests were conducted to test for differences in age groups after applying each of the EEG pre-processing pipelines and FC metrics. Since our aim was to identify the pipelines providing high performing FC estimation (rather than to detect an effect between different conditions or groups), we did not apply multiple comparison controls, emphasizing sensitivity to differences between pipelines in FC estimation rather than protection against potential false positives (Bailey et al., 2023b). After the top preforming pipelines were identified based on highest effect sizes, additional paired-samples t-tests were conducted to assess if the top performing pipelines significantly differed from one another overall.

Next, we tested the consistency of FC estimates within the same individuals over time using the test–retest reliability provided by the different EEG FC pipelines. The Intraclass Correlation Coefficient (ICC) was used to assess the test–retest reliability of each pipeline. The ICC is a correlation coefficient that evaluates the consistency between measures of the same class (i.e., repeated measurements of the same dependent variable). The ICC is commonly used to evaluate inter-rater (or inter-measure or inter-test) reliability and consistency (Field et al., 2012; Shrout and Fleiss, 1979; Koo and Li, 2016). The test produces a result between 0 and 1. Values < 0.5 indicate poor reliability, values between 0.5 and 0.75 indicate moderate reliability, values between 0.75 and 0.9 indicate good reliability, and values > 0.9 indicate excellent reliability (Koo and Li, 2016).

2.6. Machine learning and classification

We also employed simple support vector machine (SVM) classification approaches to evaluate the performance of each individual pipeline at classifying individuals to the different age groups using the FC values from each individual pair of electrodes as the input into the SVM. The SVM classifiers were configured with a radial basis function (RBF) kernel, as seen in previous literature (Meier et al., 2012).

To evaluate the pipelines and ensure reliable results, we performed the following binary comparisons for each of the pipelines: children (Group 1; G1) versus adults (Group 3; G3), children (G1) versus older adults (Group 4; G4), and young adults (Group 2; G2) versus older adults (G4). Comparing these groups allowed for a larger age gap between participants, which avoided possible effects of inter-individual variability in development, transitions from one age group to another, and probable similarities between the upper boundary of a younger group and the lower boundary of an older groups as would be the case if two adjacent age groups were compared (i.e., G1 and G2 vs. G1 and G3).

A dataset consisting of all frontal individual electrode pair FC and corresponding age-group labels was subjected to a 5-fold cross-validation procedure, with each iteration involving the training and evaluation of a SVM classifier. The utilization of stratified 5-fold cross-validation (i.e., the data is divided in such a way that each fold maintains the same proportion of classes as the original dataset) ensured robustness in assessing the SVM model's ability to generalize to unseen data. The 5-fold cross-validation also ensured that optimistic classification accuracies could not be obtained as a result of the SVM model simply memorizing the training data, since doing so would represent over-fitting to the training data and result in poor model fits in the test data.

Finally, due to imbalanced group sizes, the model might become biased toward the majority class, leading to poor performance on minority classes. Therefore, to test classification performance we used the balanced accuracy measure. Balanced accuracy gives equal weight to each class, regardless of its size to account for bias. Using this metric, a perfect classifier has a score of 1 whereas a classifier operating at random has a score of 0.5. As such, the model was assessed on each fold in terms of balanced accuracy and through the generation of classification reports.

We also examined Receiver Operating Characteristic (ROC) curves which illustrate the performance of the binary classification model across different discrimination thresholds. ROC curves are created by plotting the true positive rate (sensitivity) against the false positive rate (1-specificity) for various threshold values. The area under the ROC curve is a common metric used to quantify the overall performance of a classification model. In the context of the SVM, ROC curves provide valuable insights into the trade-off between sensitivity and specificity and allow for visual assessment of how well the SVM model distinguishes between the positive and negative classes at different decision thresholds. A model with a higher area under the ROC value generally indicates better discriminative power.

3. Results

3.1. Assessing pipeline performance at detecting age-related differences in alpha FC

A total of 36 ANOVA's were conducted to assess which pipeline would best identify age-group differences in alpha frequency FC. A summary of the top performing pipelines (where the ANOVAs provided a partial n^2 over 0.09, indicating a moderate effect size) is provided in Table 2, while the full results for all ANOVAs is provided in the Supplementary Materials 2. The additional paired-samples t-tests conducted to assess if these top performing pipelines significantly differed from one another can also be found in Supplementary Materials 2.

Overall, the effect sizes from the ANOVAs indicated that the top performing pre-processing pipelines for the estimation of power-based FC (using the rMSC metric) included the use of CSD re-referencing, artifact reduction with ICA or wICA, and the use of either 2 or 6 s epochs. Meanwhile, the top performing pipelines for estimating phase-based FC (wPLI and iCOH) included the use of CAR re-referencing, 6 s epochs, and ICA or wICA artifact reduction approaches. Additionally, it is interesting to note that the power-based FC estimate method (rMSC) showed increases in FC from childhood until adulthood, after which FC estimates decreased again in older adults. In contrast, the phase-based FC estimate methods (iCOH and wPLI) showed a decrease in FC estimates from childhood to adulthood, and only minimal change after that. See Fig. 2 for mean FC for each of the four groups for the top performing pipelines.

3.2. SVM classification of age groups according to FC

Fig. 3 depicts the SVM classification performances for the top performing pipelines. SVM results for all pipelines are depicted in Supplementary Materials 2. An ROC curve was calculated for each of the 5 folds, and the average ROC area under the curve across the 5 folds is also reported in Table 3. A value between 0.5 and 1 indicated better model performance.

The pipelines that both demonstrated the best ability to classify age groups according to the results of the SVM and that were significant in the ANOVA tests reported in the previous section included the following pre-processing steps: CSD re-referencing, wICA or ICA cleaning, and the rMSC FC metric. Pipelines that included both 2 s epochs and 6 s epochs were represented in the highest performing set of pipelines, indicating that the choice of epoch length did not substantially impact pipeline performance.

We noted that the youngest age group (<18-year-olds) contained 69.74 % female participants, while other age groups contained ~ 50 % female participants. As such, we assessed if the effects of gender on FC

Table 2
Summary of ANOVA results in detecting age-related changes for the top performing pipelines.

Pipeline	Main Effect	Post-Hoc Group Comparisons (Holm p-values)					
		G1 vs G2	G1 vs G3	G1 vs G4	G2 vs G3	G2 vs G4	G3 vs G4
rMSC_2s_ICA_CSD	p<0.001, n2 = 0.178	<.001	<.001	<.001	<.001	>.05	<.001
rMSC_6s_ICA_CSD	p<0.001, n2 = 0.173	<.001	<.001	<.001	<.001	>.05	<.001
rMSC_2s_wICA_CSD	p<0.001, n2 = 0.171	<.001	<.001	<.001	<.001	.020	<.001
rMSC_6s_wICA_CSD	p<0.001, n2 = 0.161	<.001	<.001	<.001	<.001	>.05	<.001
iCOH_6s_ICA_CAR	p<0.001, n2 = 0.154	<.001	<.001	<.001	>.05	>.05	>.05
iCOH_6s_wICA_CAR	p<0.001, n2 = 0.143	<.001	<.001	<.001	>.05	>.05	>.05
rMSC_2s_wICA_CAR	p<0.001, n2 = 0.139	<.001	<.001	<.001	.004	<.001	<.001
rMSC_2s_MWF_CSD	p<0.001, n2 = 0.132	<.001	<.001	<.001	.017	.018	<.001
rMSC_6s_MWF_CSD	p<0.001, n2 = 0.123	<.001	<.001	<.001	.005	.035	<.001
rMSC_2s_ICA_CAR	p<0.001, n2 = 0.120	<.001	<.001	<.001	.015	<.001	<.001
iCOH_6s_MWF_CAR	p<0.001, n2 = 0.109	<.001	<.001	<.001	>.05	>.05	>.05
rMSC_2s_MWF_CAR	p<0.001, n2 = 0.108	<.001	<.001	.012	.046	<.001	<.001
rMSC_6s_wICA_CAR	p<0.001, n2 = 0.102	<.001	<.001	<.001	.027	.027	<.001
wPLI_6s_ICA_CAR	p<0.001, n2 = 0.091	<.001	<.001	<.001	>.05	>.05	>.05

Note. This table presents only the results for pipelines with a partial n^2 over 0.09. Grey shading denotes significance level for post-hoc tests, with darker shading indicating greater statistical significance. *Abbreviations:* G1 = Group 1 (children ≤ 18); G2 = Group 2 (young adults 19–35); G3 = Group 3 (adults 36–59); G4 = Group 4 (older adults ≥ 60 years of age); The steps included in each pipeline presented in this table are identified by combining the following abbreviations: 2 s = two second epoch length; 6 s = six second epoch length; CAR = common average reference; CSD = current source density; ICA = independent component analysis; ICC = intraclass correlation coefficient; iCOH = imaginary coherence; MWF = multiple Wiener filters; N = number of participants; rMSC = real magnitude squared coherence; wICA = wavelet-enhanced ICA; wPLI = weighted phase lag index.

might have confounded our results pertaining to age. Simple Student's *t*-test were conducted to compare FC estimates between the genders in the youngest age group within the best performing pipelines. All t-tests showed $p > 0.05$, indicating that FC estimates did not differ by gender within the youngest age group.

3.3. Test-retest reliability of FC pipelines

Fig. 4 depicts the results of the ICC the results indicating that all pipelines for estimating alpha FC demonstrated good test–retest reliability (Koo and Li, 2016). Based on ICC results, the pipelines that demonstrated the highest test–retest reliability included the use of rMSC FC metrics, CAR re-referencing and either ICA or MWF artifact reduction approaches. ICC results for all remaining pipelines are summarized in Supplementary Materials 2.

Although pipelines that included the rMSC metric and CAR provided the highest test–retest reliability out of the pipelines we tested, it should be noted that our companion paper (Part 1) showed that the combination of rMSC with CAR re-referencing was associated with high FC estimates when no real FC was simulated (only volume conduction). This is in alignment with previous research showing that rMSC is vulnerable to volume conduction confounds when an appropriate re-referencing montage is not applied (Anastasiadou et al., 2019; Nolte et al., 2004). Thus, the high test–retest reliability for pipelines that included rMSC combined with CAR could reflect the detection of volume conduction of a single common source, rather than high test–retest reliability for true FC. In particular, the contribution of strong alpha activity that may have been volume conducted to two electrodes simultaneously could have artificially inflated the FC estimate. To assess whether alpha activity volume conduction might have affected our test–retest reliability assessment of alpha FC, mean alpha power across the frontal electrodes was calculated in the CAR and CSD re-referenced data after applying the

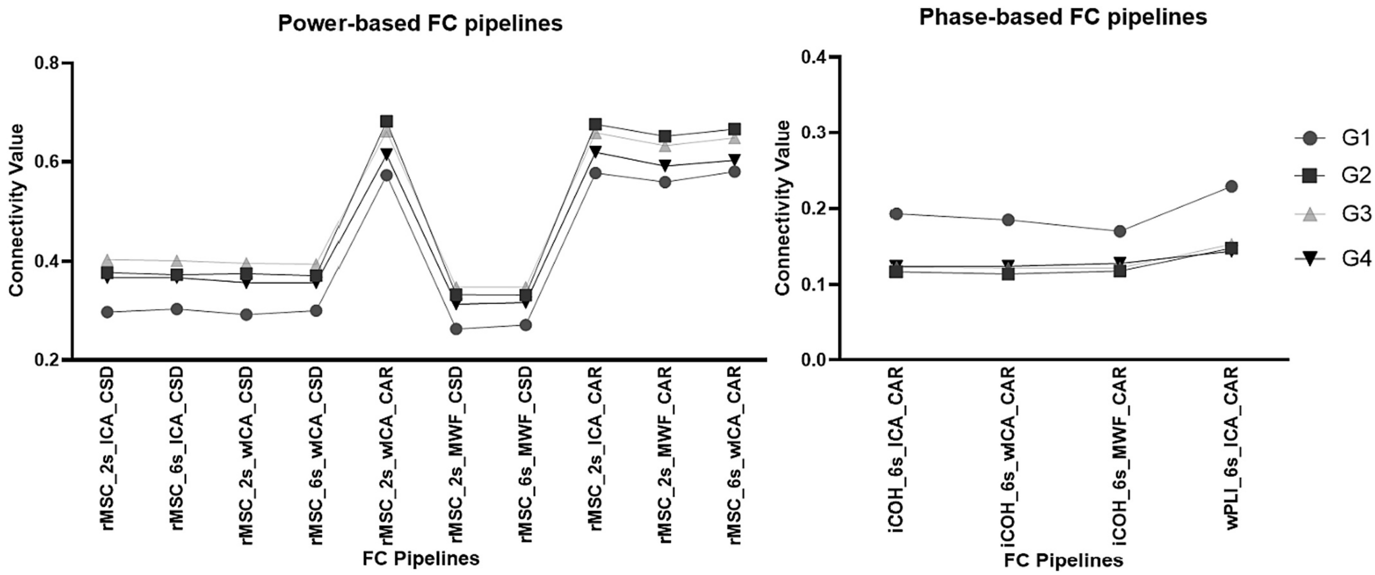

Fig. 2. Mean FC per age-group from the power-based and phase-based FC pipelines for the highest performing pipelines from the ANOVA analyses. *Note.* This figure only depicts means for pipelines that obtained a partial n^2 over 0.09 in ANOVAs, as per Table 2. Please note y-axis range differences for phase versus power FC. *Abbreviations:* G1 = Group 1 (children ≤ 18); G2 = Group 2 (young adults 19–35); G3 = Group 3 (adults 36–59); G4 = Group 4 (older adults ≥ 60 years of age); each of the combined pre-processing pipelines is labelled with the steps included in the pipeline, including the FC metric, the epoch length, the artifact cleaning method, and the re-referencing montage. 2 s = two second epoch length; 6 s = six second epoch length; CAR = common average reference; CSD = current source density; ICA = independent component analysis; ICC = intraclass correlation coefficient; iCOH = imaginary coherence; MWF = multiple Wiener filters; N = number of participants; rMSC = real magnitude squared coherence; wICA = wavelet-enhanced ICA; wPLI = weighted phase lag index.

best performing artifact reduction technique (wICA) and the 6 s epoch conditions. Test-retest reliability was measured on these estimates of alpha power (in contrast to alpha FC). It was observed that CAR re-referencing led to significantly higher test–retest reliability between the two sessions for alpha power estimation compared to CSD re-referencing, where alpha power estimates showed non-significant and poor test-re-rest reliability (see Supplementary Materials 2 for results). This suggests that high the test–retest reliability scores for the rMSC FC metric after CAR re-referencing might have been driven by the volume conduction of a single alpha generating source region rather than capturing true high test–retest reliability for the FC estimates. We note that following the pipelines that used the rMSC metric and CAR re-referencing, pipelines that used the rMSC metric and CSD re-referencing showed the highest test–retest reliability. These “rMSC with CSD re-referencing” FC estimation pipelines also showed the highest performance at distinguishing between age groups based on FC, suggesting they reflect the best all-round performing pipelines out of the approaches we tested.

4. Discussion

This paper assessed the effects of several EEG pre-processing choices on FC estimation. Pre-processing steps that were examined included the selection of artifact cleaning approaches (ICA, wICA, or MWF, applied via the RELAX EEG pre-processing toolbox), the choice of different epoch lengths (2 s versus 6 s epochs), the use of different re-referencing techniques (CAR versus CSD), and the use of different FC metrics (rMSC, iCOH, or wPLI). The performance of these pre-processing options was assessed relative to performance in identifying age-related changes in alpha FC, and demonstrating high test–retest reliability, within a large dataset of EEG data. Within this context, age related-differences in FC reflected the equivalent of a “ground-truth” for evaluating the pipelines as these changes are well-documented and reliably reproduced in the scientific literature. Overall, pipelines that used the CSD re-reference, either ICA or wICA artifact reduction, and the rMSC FC estimation method demonstrated the highest classification accuracy as well as moderate-high test–retest reliability. However, it is worth noting that almost all (33/36) tested pipelines identified significant differences in EEG FC between the age groups we tested.

4.1. *Top performing artifact cleaning techniques*

Overall, the use of ICA and wICA (implemented via the RELAX toolbox) to reduce artifacts in the EEG data prior to FC estimation appeared to enable better performance in FC identification for the current dataset, demonstrating superiority over the MWF cleaning approach. However, this result was not consistent across all of the pipeline combinations we tested and depended on the other pre-processing steps used in conjunction with the cleaning technique. Nonetheless, our results suggest that the use of wICA cleaning was most effective when applied to the current dataset, and demonstrated higher test–retest reliability when paired with the rMSC metric and CSD re-referencing approaches (representing the high performing pipeline settings from the other conditions) compared to both the ICA and MWF approaches. However, this recommendation must be caveated by the note that analyses in the present study were limited to alpha FC and may not generalize equally to FC within other oscillatory bands. In addition, all artifact reduction approaches were implemented within the RELAX toolbox, so other implementations of these cleaning methods may show different results (although RELAX was designed to implement high performing artifact reduction approaches synthesized from previous literature, and to clean data in a manner that matches manual data cleaning by experts; Bailey et al., 2023a).

This study also provides the first assessment of the novel MWF technique in FC assessment. Recent evidence suggests that pipelines employing MWF cleaning perform well at reducing artifacts in EEG data (see Bailey et al., 2023). While we noted high test–retest for pipelines employing MWF, the MWF cleaning method did not perform as well as the wICA and ICA approaches at enabling the detection of differences in alpha FC between age groups.

4.2. *Top performing re-referencing techniques*

As highlighted in our companion article (Milejivic et al., in

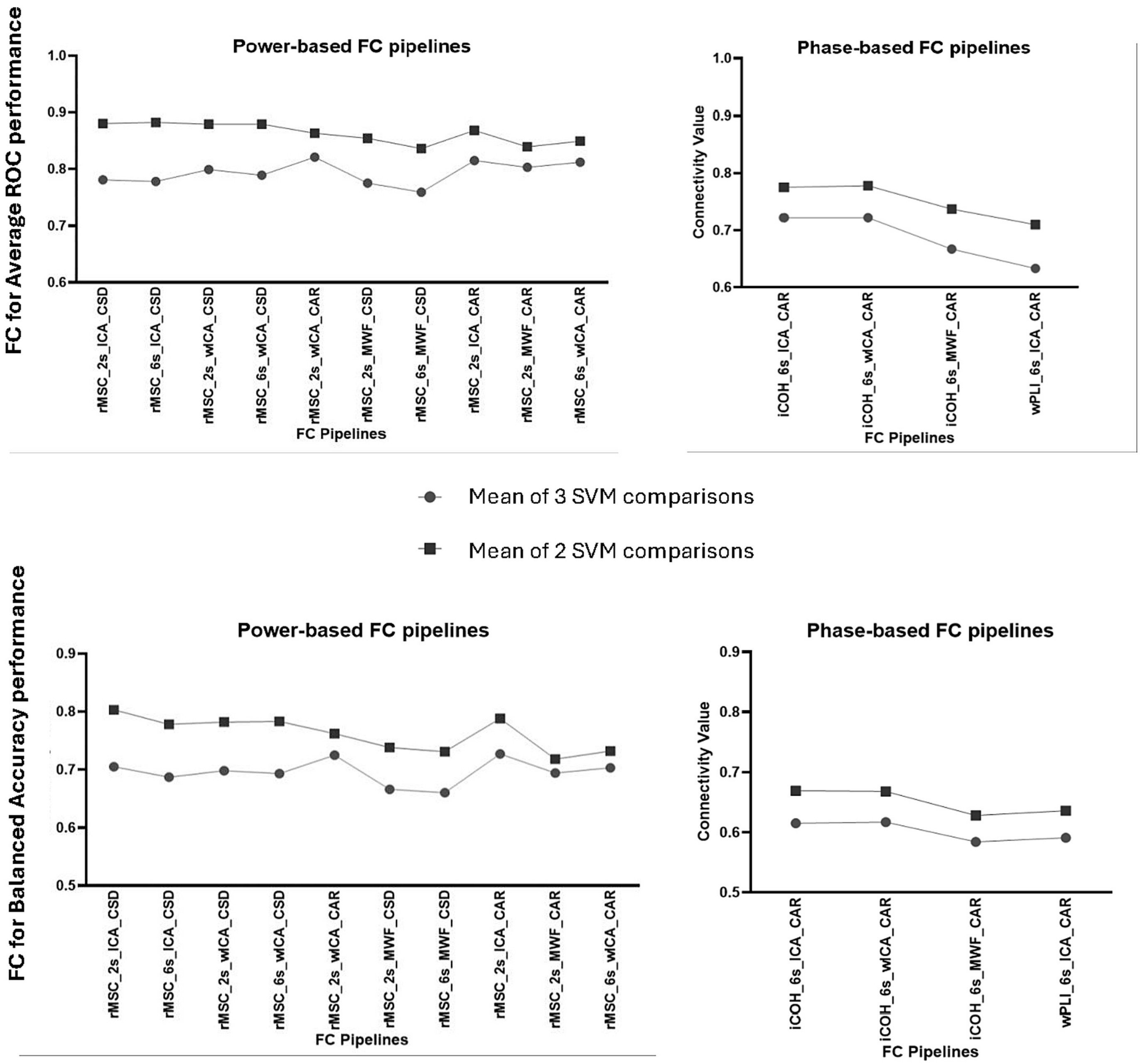

Fig. 3. Mean SVM outputs depicts the average ROC and Balanced Accuracy for the highest performing pipelines. *Note.* Only depicts means for topmost performing pipelines. *Abbreviations:* 2 s = two second epoch length; 6 s = six second epoch length; average ROC = average Receiver Operating Characteristic across 5 kernel folds; CAR = common average reference; CSD = current source density; ICA = Independent Component Analysis; iCOH = imaginary coherence; MWF = multiple Wiener filters; rMSC = real magnitude squared coherence; wICA = wavelet-enhanced ICA; wPLI = weighted phase lag index. The “mean of 3 SVM comparisons” depicts the classification performance across simple support vector machines (SVM) trained to classify G1 from G2, G1 from G3, and G1 from G4. The “mean of 2 SVM comparisons” depicts the classification accuracies across SVMs trained to classify G1 from G3 and G1 from G4 (which were the classifiers that provided the best performance across most of the pipelines).

submission), it is important to consider re-referencing techniques in combination with the selected FC metric. Within this study, the CAR re-referencing technique combined with rMSC provided higher test–retest reliability when compared to rMSC with CSD. However, in our companion article (Part 1, Miljevic et al., 2025), our results showed that rMSC with CAR sometimes identified higher FC in conditions where only volume conducted signals were simulated compared to conditions where physiologically plausible (time-lagged) FC was simulated. This is likely due to rMSC’s vulnerability to volume conduction and CAR’s lack of control for this potential confound in contrast to CSD (Miljevic et al., 2024). Our current results also suggest that the high test–retest reliability for the pipeline that combined rMSC with the CAR re-referencing approach may have been due to the volume conduction of a single alpha source. As such, we recommend the use of the CSD re-reference in combination with rMSC to control for the effects of volume conduction.

Finally, our results suggest that when combined with the CSD re-referencing montage, the rMSC method provided the best ability to detect differences between the different age groups and enabled the best performance at classifying participants into their age groups based on their alpha FC out of the pipelines we tested. However, we cannot draw general conclusions about the best FC metric for application across

Table 3
Summary of machine learning results for the top performing pipelines.

Pipeline	Mean of all 3 machine learning comparisons		Mean of 2 best machine learning comparisons	
	Average ROC	Balanced Accuracy	Average ROC	Balanced Accuracy
rMSC_2s_ICA_CSD	0.781	0.705	0.880	0.803
rMSC_6s_ICA_CSD	0.778	0.687	0.882	0.778
rMSC_2s_wICA_CSD	0.799	0.698	0.879	0.782
rMSC_6s_wICA_CSD	0.789	0.693	0.879	0.783
iCOH_6s_ICA_CAR	0.722	0.615	0.775	0.669
iCOH_6s_wICA_CAR	0.722	0.617	0.778	0.668
rMSC_2s_wICA_CAR	0.821	0.725	0.863	0.762
rMSC_2s_MWF_CSD	0.775	0.666	0.854	0.738
rMSC_6s_MWF_CSD	0.759	0.660	0.836	0.731
rMSC_2s_ICA_CAR	0.815	0.727	0.868	0.788
iCOH_6s_MWF_CAR	0.667	0.584	0.737	0.628
rMSC_2s_MWF_CAR	0.803	0.694	0.839	0.718
rMSC_6s_wICA_CAR	0.812	0.703	0.849	0.732
wPLI_6s_ICA_CAR	0.633	0.591	0.710	0.636

Note. Only results for top performing pipelines displayed. *Abbreviations:* 2 s = two second epoch length; 6 s = six second epoch length; Average ROC = Average Receiver Operating Characteristic across 5 kernel folds; CAR = common average reference; CSD = current source density; ICA = independent component analysis; ICC = intraclass correlation coefficient; iCOH = imaginary coherence; MWF = multiple Wiener filters; N = number of participants; rMSC = real magnitude squared coherence; wICA = wavelet-enhanced ICA; wPLI = weighted phase lag index. Columns 2–3 depict summaries for the mean classification performance across simple support vector machines (SVM) trained to classify G1 from G2, G1 from G3, and G1 from G4. Columns 4–5 depict the means for only classification accuracies across SVMs trained to classify G1 from G3 and G1 from G4 (which were the classifiers that provided the best performance across most of the pipelines).

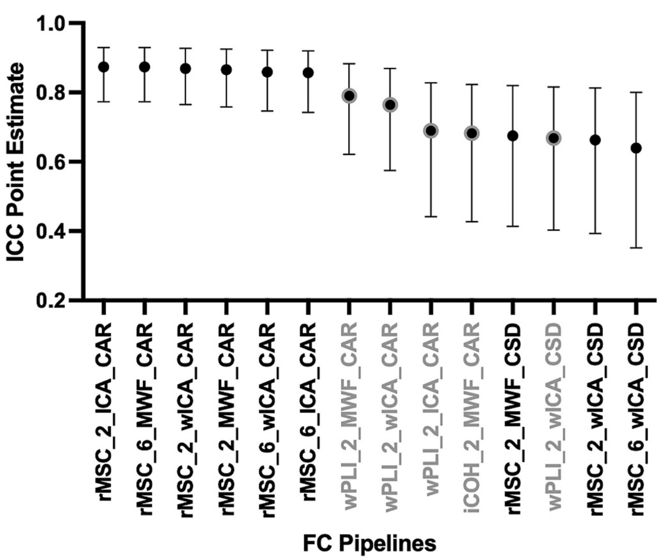

Fig. 4. Intraclass correlation coefficient results for the test-re-test reliability of the highest performing pipelines. *Note.* Phase-based measures are in grey, power-based measures in black, circles depict means and the bars depict upper/lower 95 % confidence intervals. ICC Point Estimate values < 0.5 = poor, values between 0.5 and 0.75 = moderate, values between 0.75 and 0.9 = good, and values > 0.9 = excellent reliability (Koo and Li, 2016). *Abbreviations:* 2 s = two second epoch length; 6 s = six second epoch length; CAR = common average reference; CSD = current source density; ICA = independent component analysis; ICC = intraclass correlation coefficient; iCOH = imaginary coherence; MWF = multiple Wiener filters; N = number of participants; rMSC = real magnitude squared coherence; wICA = wavelet-enhanced ICA; wPLI = weighted phase lag index.

studies, as FC metric use is dependent on data and research question specific factors, for example whether phase or power FC might be expected to differ between groups or conditions.

4.3. Top performing epoch lengths

Epoch length was not found to substantially alter FC estimates, with both 2 and 6 s epochs appearing in the top performing pipelines. This lack of significant differences in epoch lengths may be due to the inclusion of fewer epochs than is recommended as optimal, where for alpha frequency FC, the use of more than 40 epochs is suggested to increase performance of the FC pipeline (Miljevic et al., 2025). However, due to data length limitations we were not able to include this number of epochs. As such, we can make no clear recommendations can be made from our results as to the best epoch length for estimating FC.

4.4. Power and phase alpha FC pipelines demonstrate good test–retest reliability

In addition to assessing the ability of each FC estimation pipeline to detect differences between age groups, we investigated the test–retest reliability of the pipelines on a subset of the data where two EEG recordings at least a week apart were available. Most of the pipelines we tested demonstrated good test–retest reliability. The highest test–retest reliability was provided by the pipeline where data were cleaned of artifacts by ICA, CAR re-referencing was applied, and rMSC was used to estimate FC. Other pipelines employing rMSC and CAR also provided good test–retest reliability. However, as previously noted, pipelines combining rMSC and CAR were also found to demonstrate lower effect sizes and discriminatory power for the different age groups. Furthermore, both our companion article (Miljevic et al., 2025) and other literature (Bastos, and Schoffelen, 2016; Marzetti et al., 2007) has suggested that methods which apply CAR re-referencing with FC estimation methods that do not control for volume conduction are vulnerable to the detection of spurious connectivity (e.g. due to volume conduction).

In contrast, the CSD re-reference reduces the influence of volume conduction from single sources on FC estimates, so the use of CSD re-referencing prior to rMSC can be recommended. Alternatively, we note that iCOH and wPLI control for volume conduction within their computation method and thus may not require the use of CSD re-referencing. Nevertheless, we did find that CSD re-referencing improved the test–retest reliability of the iCOH FC metric (but not the wPLI metric).

Given our overall results, and taking into account existing literature and the results of our companion study (Miljevic et al., 2025), we recommend an FC estimation pipeline that cleans data with wICA, re-references data with CSD, and uses the rMSC FC estimation metric to obtain best detection of meaningful effects, highest test–retest reliability, and protection against spurious volume conduction-connectivity in power-based measures. Meanwhile, if researchers are specifically interested in applying phase-based FC measures, then our results suggest that pipelines utilising the iCOH FC metric appeared to discern between age groups best when combined with 6 s epochs (as the iCOH metric only performed highly when 6 s epochs were used), the MWF artifact cleaning technique was used, and the CAR re-referencing montage was applied. This pipeline consistently appeared in the top 10 best performing pipelines for both ANOVA, machine learning, and test–retest examination, unlike other phase-based FC pipelines which did not show the same consistency of high performance across our tests.

4.5. Phase-based FC metrics and power-based FC metrics show an opposite pattern in their age-related effects

Interestingly, this study also detected a different pattern of age-related changes between phase-based and power-based FC. Specifically, adults showed stronger FC than children when using the power-based metric rMSC in alignment with our expectations. In contrast, for the phase-based metrics (iCOH and wPLI), we observed higher FC in children compared to the adulthood groups.

These results may suggest that age-related changes in alpha FC are driven by specific neurophysiological properties that are captured differently by the different FC metrics. It seems that the increase in alpha FC from childhood to adulthood is driven by increases in frequency power FC with increasing age, whereas the decrease is driven by reductions in phase FC with increasing age. This is an interesting distinction that, while supported by some literature using different metrics (see Miljevic et al., 2023), has not been explicitly reported or interpreted by previous research.

Higher values for phase-based connectivity metrics (e.g. iCOH and wPLI) are suggested to reflect more precise neuron-to-neuron patterns of synchrony, with closely timed voltage shifts thought to relate to information processing in the brain (Bush and Sejnowski, 1996; Ng et al., 2013; Schyns et al., 2011; Varela et al., 2001). In contrast, higher values for power-based connectivity metrics (e.g. rMSC) imply that connected brain regions show commonalities in their dynamic neural activity, but the connectivity pattern is less direct and more reflective of activation of wider neuronal populations, with the potential for high values even if oscillations occur at slightly different frequencies or with the oscillatory frequency shifting over time (Hanslmayr et al., 2011; Varela et al., 2001). In interpreting this result, we note that alpha oscillations are associated with the top-down inhibition of brain regions that are not relevant for the current focus of attention task, and play a central role in supporting working memory (Klimesch, 2012; Klimesch et al., 2007). Within this framework, the decrease in phase FC from early to late adulthood in the current study broadly aligns with prior behavioural evidence that working memory performance, and inhibitory functions, also typically decline from early to late adulthood (Alloway and Alloway, 2013; Andrés et al., 2008; Hale et al., 2011).

The age-related differences in phase-based alpha FC observed in our study may suggest that younger brains require tightly coupled connections to engage inhibitory processes across multiple brain regions simultaneously. In contrast, older brains might engage the same inhibitory mechanisms, but with the alpha inhibitory mechanisms more independently engaged across brain regions, without the need for the highly synchronized timing that is detected with phase-based FC metrics. This developmental change might reflect the increased neural system maturity that we know occurs from childhood to adulthood (Edde et al., 2021; Johnson et al., 2009). In EEG specifically, this possibility is further supported by studies like Chung et al. (2022) where the authors note the most significant age-related differences occur in the alpha frequency. Specifically, Chung et al. found that resting-state EEG FC networks in the healthy brain show significant increases in the degree of segregation and integration with age, speculating that normal brain maturation during is characterised by functional specialization and efficiency enhancement. Smit et al. (2016) notes a similar result further showing that FC plateaus in adulthood and then begins to decrease significantly in later adulthood (~55 years of age). However, further work is needed to interrogate the underlying neural mechanisms contributing to phase-based and power-based EEG FC estimation to better understand these age-related changes.

4.6. Improvements, considerations, and future directions

While our study had a number of strengths, including the testing of multiple settings in several EEG pre-processing steps for FC estimation in a very large dataset, it also has several limitations. An important consideration for our tests of age-related differences in FC was that the sample of children (<18-year-olds) had a higher proportion of females than other age groups (69.74 % female in the children group vs an average of ~50 % female across the other groups). However, direct comparison of males and females in the <18 years age group revealed no significant differences in the FC estimates, suggesting that the higher proportion of females in this group was unlikely to have affected the results of this study for these top performing pipelines.

In addition, the EEG dataset used in the present study included a relatively low number of 26 recording electrodes. There is evidence that EEG artifact removal with ICA is less effective with fewer channels as this may limit the number of components that can be reliably isolated (Onton et al., 2006; Pontifex et al., 2017). The number of electrodes may also provide a reduced ability to detect FC due to a reduced probability of coverage across areas where experimental effects might be detected. However, no studies to date have directly investigated how the number of EEG channels may affect FC estimation and therefore further research in this area is needed. Similarly, based on past literature, there is some indication that ICA decomposition benefits from longer EEG recording lengths (Iriarte et al., 2003; Stone, 2002). This may need to be further investigated, as the EEG data we tested contained shorter recording lengths (2 min). Additionally, while our results allow for inferences about the application of ICA, wICA or MWF to reduce artifacts prior to FC estimation, the application of these approaches was restricted to within the RELAX EEG pre-processing toolbox, and to the default settings (Bailey et al., 2023a). Although the RELAX default settings were based on both standard approaches within the literature and designed to match the manual cleaning of EEG data by experts (Bailey et al., 2023a), there may be variation between different EEG artifact reduction parameter settings, so we note that it is possible alternative applications of these cleaning approaches might lead to different results. It should also be noted that we only tested one variation of the CSD re-reference technique, using CSD parameters that are commonly applied in existing literature and set as the defaults within EEGLAB. It is important to note that the algorithm can vary substantially based on user inputs, which might alter study outcomes.

Additionally, the present study focused solely on scalp-space EEG FC. Scalp-space recordings are known to be influenced by several potential confounds when estimating FC, including volume conduction (Brunner et al., 2016). Scalp estimates of FC are also not spatially precise in comparison to source-space FC measures where the underlying neural sources or origins of the recorded electrical signals are estimated (Hassan et al., 2014). Despite these limitations, we were able to identify an FC estimation pipeline that provided a good ability to classify age group from scalp space alpha FC and demonstrated high test–retest reliability. In part, this may be because our top performing pipelines used approaches that controlled for volume conduction (e.g. CSD). However, while source space measures are better able to localize activity, scalp space recordings make fewer assumptions about the nature of the data than source space FC, due to the requirement for modelling to solve the inverse problem when determining source generators of EEG data (Burle et al., 2015; Liu et al., 2023; Michel and Brunet, 2019). Scalp space EEG analyses can provide a global or 'global-regions' overview of brain activity and the simplicity and non-invasive approach of scalp-based FC estimation means the method may be more easily translated to clinical application. Therefore, the presented pipeline employing scalp-space EEG FC provides viable and reliable recommendations for future research, and while we recommend further testing and replication, our results suggest that scalp space FC measures retain utility and should not be ignored.

While we tested a reasonable range of EEG pre-processing approaches that are common in the literature, these comparisons were not exhaustive and other methods are available. Heterogeneity in pre-processing and metric selection approaches is particularly noteworthy for FC estimation algorithm, with >20 algorithms reported in the literature (for recent reviews see Chiarion et al., 2023; Miljevic et al., 2023; Sharma and Meena, 2024). Future research could aim to more comprehensively explore the EEG pre-processing steps and different FC computation metrics to determine how FC estimates are influenced by variations in pre-processing and the choice of FC metric. We would also encourage the development and investigation of how different pipelines relate to the identification of FC in frequencies outside of the alpha range. Nonetheless, given the evidence provided in the current study, we encourage future researchers to use the more robust pre-processing steps reported here in their EEG FC estimation pipelines. Use of these high performing FC estimation pipelines will also ensure standardization across research and allow for more meaningful and clear comparison of research findings across the literature in meta-analytical works.

While it was not the primary aim of this study, it is worth noting that future research evaluating EEG FC should take into consideration (and control for) the effects of age. We note that age-related FC change (both phase- and power-based) was present despite a rather heterogeneous sample of participants with various noted clinical conditions. Moreover, our study also indicates a different pattern of developmental shift in brain connectivity from childhood to adulthood depending on the FC metric used, whereby power-based connectivity increases while phase-based connectivity decreases from childhood to adulthood. It is evident here that no matter the pre-processing steps and FC metrics used, age-related differences in FC are present and prominent, particularly from childhood to young adulthood. Therefore, no matter the setting or population, researchers and clinicians should be aware that age-related FC changes may pose confounds to their non-age-related analyses. Exploration of these differences and the underlying mechanisms that they may reflect is required so we can more precisely understand their potential functional impacts.

5. Conclusion

The present study provided evidence for several FC estimation pipelines which showed robust performance across a number of conditions, including the detection of age-related differences in alpha FC, the classification of participants to age groups based on their alpha FC, and demonstrating high test–retest reliability. The best performing pipelines enabled assessment of power-based FC, and involved the cleaning of artifacts via wICA, re-referencing data using the CSD approach, and using rMSC as the FC metric. Alternatively, if researchers are interested in phase-based FC estimation, the highest performing phase-based pipeline involved artifacts being cleaned with MWF, the use of the CAR re-referencing approach, and FC estimation with iCOH. Nonetheless, we recognize that more work is needed consolidate these findings and expand our understanding of the complex interactions between EEG pre-processing steps and FC analyses, particularly for frequencies beyond the alpha band. Several pipelines for evaluating EEG FC have been published recently including high-density EEG analysis and graph theory (Iandolo et al., 2019), pipelines for evaluating FC in cortical source space (Xie et al., 2022), and a 'biomarker' pipeline for various EEG feature extraction (Gil Ávila et al., 2023). However, no studies that we could find have directly compared various EEG pre-processing steps or combinations of steps to ascertain the quality of FC estimations, enabling validation of a robust and powerful pipeline for EEG FC identification. Here, we present an account of how this may be achieved, furthering understanding of the effects of different EEG (pre-) processing steps on FC estimation.

Authorship confirmation & contribution statement

All who meet authorship criteria are listed as authors, and all certify that they have participated sufficiently in the work to take public responsibility for the content. With AM contributing to conceptualization, design, literature search and study inclusion/exclusion, writing – preparation, creation, writing – editing, and revision; OWM and PBF contribution to the writing – editing, and revision; and NWB to conceptualization, design, writing – editing, and revision.

Funding

In the last 3 years PBF has received equipment for research from Neurosoft, Nexstim and Brainsway Ltd. He has served on scientific advisory boards for Magstim and LivaNova and received speaker fees from Otsuka. He has also acted as a founder and board member for TMS Clinics Australia and Resonance Therapeutics. PBF is supported by a National Health and Medical Research Council of Australia Investigator grant (1193596). AM was funded by an Epworth Medical Foundation grant.

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.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.clinph.2025.03.042.

References

Alloway, T.P., Alloway, R.G., 2013. Working memory across the lifespan: a cross-sectional approach. J. Cogn. Psychol. 25 (1), 84–93.
Anastasiadou, M.N., Christodoulakis, M., Papathanasiou, E.S., Papacostas, S.S., Hadjipapas, A., Mitsis, G.D., 2019. Graph theoretical characteristics of EEG-based functional brain networks in patients with epilepsy: the effect of reference choice and volume conduction. Front. Neurosci. 13, 221.
Andrés, P., Guerrini, C., Phillips, L.H., Perfect, T.J., 2008. Differential effects of aging on executive and automatic inhibition. Dev. Neuropsychol. 33 (2), 101–123.
Bailey, N.W., Biabani, M., Hill, A.T., Miljevic, A., Rogasch, N.C., McQueen, B., Fitzgerald, P.B., 2023a. Introducing RELAX: an automated pre-processing pipeline for cleaning EEG data-Part 1: algorithm and application to oscillations. Clin. Neurophysiol. 149, 178–201.
Bailey, N.W., Hill, A.T., Biabani, M., Murphy, O.W., Rogasch, N.C., McQueen, B., Fitzgerald, P.B., 2023b. RELAX part 2: a fully automated EEG data cleaning algorithm that is applicable to event-related-potentials. Clin. Neurophysiol. 149, 202–222.
Bailey, N.W., Hill, A.T., Godfrey, K., Perera, M.P.N., Rogasch, N.C., Fitzgibbon, B.M., Fitzgerald, P.B., 2024. EEG is better when cleaning effectively targets artifacts. bioRxiv, 2024-06.
Bastos, A.M., Schoffelen, J.M., 2016. A tutorial review of functional connectivity analysis methods and their interpretational pitfalls. Front. Syst. Neurosci. 9 (175).
Brunner, C., Billinger, M., Seeber, M., Mullen, T.R., Makeig, S., 2016. Volume conduction influences scalp-based connectivity estimates. Front. Comput. Neurosci. 10, 121.
Burle, B., Spieser, L., Roger, C., Casini, L., Hasbroucq, T., Vidal, F., 2015. Spatial and temporal resolutions of EEG: is it really black and white? A scalp current density view. Int. J. Psychophysiol. 97 (3), 210–220.
Bush, P., Sejnowski, T., 1996. Inhibition synchronizes sparsely connected cortical neurons within and between columns in realistic network models. J. Comput. Neurosci. 3, 91–110.
Castellanos, N.P., Makarov, V.A., 2006. Recovering EEG brain signals: artifact suppression with wavelet enhanced independent component analysis. J. Neurosci. Methods 158 (2), 300–312.
Cohen, M.X., 2014. Analyzing Neural Time Series Data: Theory and Practice. MIT Press.
Chiarion, G., Sparacino, L., Antonacci, Y., Faes, L., Mesin, L., 2023. Connectivity analysis in EEG data: a tutorial review of the state of the art and emerging trends. Bioengineering 10 (3), 372.
Chung, Y.G., Jeon, Y., Kim, R.G., Cho, A., Kim, H., Hwang, H., Kim, K.J., 2022. Variations of resting-state EEG-based functional networks in brain maturation from early childhood to adolescence. J. Clin. Neurol. 18 (5), 581.
Da Silva, F.L., 1999. EEG analysis: theory and practice. Electroencephalography: basic principles, clinical applications and related fields, pp. 1125–1159.
Duffy, F.H., Mcanulty, G.B., Albert, M.S., 1996. Effects of age upon interhemispheric EEG coherence in normal adults. Neurobiol. Aging 17 (4), 587–599.
Edde, M., Leroux, G., Altena, E., Chanraud, S., 2021. Functional brain connectivity changes across the human life span: from fetal development to old age. J. Neurosci. Res. 99 (1), 236–262.
Field, A.P., Miles, J., Field, Z., 2012. Discovering Statistics Using R. Sage.
Gil Ávila, C., Bott, F. S., Tiemann, L., Hohn, V. D., May, E. S., Nickel, M. M., Ploner, M., 2023. DISCOVER-EEG: an open, fully automated EEG pipeline for biomarker discovery in clinical neuroscience. bioRxiv, 2023-01.
Hagemann, D., Naumann, E., Thayer, J.F., 2001. The quest for the EEG reference revisited: A glance from brain asymmetry research. Psychophysiology 38 (5), 847–857.
Hale, S., Rose, N.S., Myerson, J., Strube, M.J., Sommers, M., Tye-Murray, N., Spehar, B., 2011. The structure of working memory abilities across the adult life span. Psychol. Aging 26 (1), 92.
Hanslmayr, S., Gross, J., Klimesch, W., Shapiro, K.L., 2011. The role of alpha oscillations in temporal attention. Brain Res. Rev. 67 (1–2), 331–343.
Hassan, M., Dufor, O., Merlet, I., Berrou, C., Wendling, F., 2014. EEG source connectivity analysis: from dense array recordings to brain networks. PLoS One 9 (8), e105041.

Hill, A.T., Bailey, N.W., Zomorrodi, R., Hadas, I., Kirkovski, M., Das, S., Lum, J.A.G., Enticott, P.G., 2023. EEG microstates in early-to-middle childhood show associations with age, biological sex, and alpha power. Hum. Brain Mapp. 44 (18), 6484–6498. https://doi.org/10.1002/hbm.26525.
Hill, A.T., Clark, G.M., Bigelow, F.J., Lum, J.A., Enticott, P.G., 2022. Periodic and aperiodic neural activity displays age-dependent changes across early-to-middle childhood. Dev. Cogn. Neurosci. 54, 101076.
Iandolo, R., Samogin, J., Barban, F., Buccelli, S., Taberna, G., Semprini, M., Chiappalone, M., 2019, March. A pipeline integrating high-density EEG analysis and graph theory: a feasibility study on resting state functional connectivity. In: 2019 9th International IEEE/EMBS Conference on Neural Engineering (NER). IEEE, pp. 271–274.
Inuso, G., La Foresta, F., Mammone, N., Morabito, F.C., 2007, August. Wavelet-ICA methodology for efficient artifact removal from Electroencephalographic recordings. In: 2007 International Joint Conference on Neural Networks. IEEE, pp. 1524–1529.
Iriarte, J., Urrestarazu, E., Valencia, M., Alegre, M., Malanda, A., Viteri, C., Artieda, J., 2003. Independent component analysis as a tool to eliminate artifacts in EEG: a quantitative study. J. Clin. Neurophysiol. 20 (4), 249–257.
Johnson, S.B., Blum, R.W., Giedd, J.N., 2009. Adolescent maturity and the brain: the promise and pitfalls of neuroscience research in adolescent health policy. J. Adolesc. Health 45 (3), 216–221.
Kavčič, A., Demšar, J., Georgiev, D., Bon, J., Soltirovska-Šalamon, A., 2023. Age related changes and sex related differences of functional brain networks in childhood: a high-density EEG study. Clin. Neurophysiol. 150, 216–226.
Klimesch, W., 2012. Alpha-band oscillations, attention, and controlled access to stored information. Trends Cogn. Sci. 16 (12), 606–617.
Klimesch, W., Sauseng, P., Hanslmayr, S., 2007. EEG alpha oscillations: the inhibition–timing hypothesis. Brain Res. Rev. 53 (1), 63–88.
Koo, T.K., Li, M.Y., 2016. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J. Chiropr. Med. 15 (2), 155–163.
Liu, C., Downey, R.J., Mu, Y., Richer, N., Hwang, J., Shah, V.A., Ferris, D.P., 2023. Comparison of EEG source localization using simplified and anatomically accurate head models in younger and older adults. IEEE Trans. Neural Syst. Rehabil. Eng. 31, 2591–2602.
Marzetti, L., Nolte, G., Perrucci, M.G., Romani, G.L., Del Gratta, C., 2007. The use of standardized infinity reference in EEG coherency studies. Neuroimage 36 (1), 48–63.
Meier, T.B., Desphande, A.S., Vergun, S., Nair, V.A., Song, J., Biswal, B.B., Prabhakaran, V., 2012. Support vector machine classification and characterization of age-related reorganization of functional brain networks. Neuroimage 60 (1), 601–613.
Michel, C.M., Brunet, D., 2019. EEG source imaging: a practical review of the analysis steps. Front. Neurol. 10, 325.
Michels, L., Muthuraman, M., Lüchinger, R., Martin, E., Anwar, A.R., Raethjen, J., Siniatchkin, M., 2013. Developmental changes of functional and directed resting-state connectivities associated with neuronal oscillations in EEG. Neuroimage 81, 231–242.
Miljevic, A., Bailey, N.W., Murphy, O.W., Perera, M.P.N., Fitzgerald, P.B., 2024. Estimating sensor-space EEG connectivity PART 1: identifying optimal choices for re-referencing and epoching for EEG functional connectivity in simulated data. (submitted to Clinical Neurophysiology).
Miljevic, A., Bailey, N.W., Murphy, O.W., Perera, M.P.N., Fitzgerald, P.B., 2023. Alterations in EEG functional connectivity in individuals with depression: a systematic review. J. Affect. Disord.
Miljevic, A., Bailey, N.W., Vila-Rodriguez, F., Herring, S.E., Fitzgerald, P.B., 2022. Electroencephalographic connectivity: a fundamental guide and checklist for optimal study design and evaluation. Biol. Psychiatry: Cognit. Neurosci. Neuroimaging 7 (6), 546–554.
Miljevic, A., Murphy, O.M., Fitzgerald, P.B., & Bailey, N.W. (2025). Estimating sensor-space EEG connectivity PART 1: Identifying optimal methods for functional connectivity in simulated data. Clinical Neurophysiology (in press).
Ng, B.S.W., Logothetis, N.K., Kayser, C., 2013. EEG phase patterns reflect the selectivity of neural firing. Cereb. Cortex 23 (2), 389–398.
Nolte, G., Bai, U., Wheaton, L., Mari, Z., Vorbach, S., Hallet, M., 2004. Identifying true brain interaction from EEG data using the imaginary part of coherency. Clin. Neurophysiol. 115, 2292–2307.
Onton, J., Westerfield, M., Townsend, J., Makeig, S., 2006. Imaging human EEG dynamics using independent component analysis. Neurosci. Biobehav. Rev. 30 (6), 808–822.
Panteliadis, C.P., 2021. Historical Overview of Electroencephalography: from Antiquity to the Beginning of the 21st Century. J. Brain Neurol. Disord 3, 1–10.
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Duchesnay, É., 2011. Scikit-learn: machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830.
Pion-Tonachini, L., Kreutz-Delgado, K., Makeig, S., 2019. ICLabel: an automated electroencephalographic independent component classifier, dataset, and website. Neuroimage 198, 181–197.
Pontifex, M.B., Gwizdala, K.L., Parks, A.C., Billinger, M., Brunner, C., 2017. Variability of ICA decomposition may impact EEG signals when used to remove eyeblink artifacts. Psychophysiology 54 (3), 386–398.
Popov, T., Tröndle, M., Baranczuk-Turska, Z., Pfeiffer, C., Haufe, S., Langer, N., 2023. Test–retest reliability of resting-state EEG in young and older adults. Psychophysiology, e14268.
Rolle, C.E., Narayan, M., Wu, W., Toll, R., Johnson, N., Caudle, T., Etkin, A., 2022. Functional connectivity using high density EEG shows competitive reliability and agreement across test/retest sessions. J. Neurosci. Methods 367, 109424.
Sabatini, B.L., Regehr, W.G., 1999. Timing of synaptic transmission. Annual review of physiology. Psychophysiology 61 (1), 521–542.
Scally, B., Burke, M.R., Bunce, D., Delvenne, J.F., 2018. Resting-state EEG power and connectivity are associated with alpha peak frequency slowing in healthy aging. Neurobiol. Aging 71, 149–155.
Schyns, P.G., Thut, G., Gross, J., 2011. Cracking the code of oscillatory activity. PLoS Biol. 9 (5), e1001064.
Sharma, R., Meena, H.K., 2024. Emerging trends in EEG signal processing: a systematic review. SN Comput. Sci. 5 (4), 1–14.
Shrout, P.E., Fleiss, J.L., 1979. Intraclass correlation: uses in assessing rater reliability. Psychol. Bull. 80 (2), 420.
Smit, D.J., de Geus, E.J., Boersma, M., Boomsma, D.I., Stam, C.J., 2016. Life-span development of brain network integration assessed with phase lag index connectivity and minimum spanning tree graphs. Brain Connect. 6 (4), 312–325.
Somers, B., Francart, T., Bertrand, A., 2018. A generic EEG artifact removal algorithm based on the multi-channel Wiener filter. J. Neural Eng. 15 (3), 036007.
Stone, J.V., 2002. Independent component analysis: an introduction. Trends Cogn. Sci. 6 (2), 59–64.
van Diessen, E, Numan, T, van Dellen, E, van der Kooi, AW, Boersma, M, Hofman, D, et al., 2015. Opportunities and methodological challenges in EEG and MEG resting state functional brain network research. Clin Neurophysiol 126, 1468–1481.
van Dijk, H., van Wingen, G., Denys, D., Olbrich, S., van Ruth, R., Arns, M., 2022. The two decades brainclinics research archive for insights in neurophysiology (TDBRAIN) database. Sci. Data 9 (1), 333.
Varela, F., Lachaux, J.P., Rodriguez, E., Martinerie, J., 2001. The brainweb: phase synchronization and large-scale integration. Nat. Rev. Neurosci. 2 (4), 229–239.
Vysata, O., Kukal, J., Prochazka, A., Pazdera, L., Simko, J., Valis, M., 2014. Age-related changes in EEG coherence. Neurol. Neurochir. Pol. 48 (1), 35–38.
Xie, W., Toll, R.T., Nelson, C.A., 2022. EEG functional connectivity analysis in the source space. Dev. Cogn. Neurosci. 56, 101119.
Yao, D., 2001. A method to standardize a reference of scalp EEG recordings to a point at infinity. Physiol. Meas. 22 (4), 693.