On the variability of dynamic functional connectivity assessment methods

Mohammad Torabi[1*], **Georgios D. Mitsis**[2§], **Jean-Baptiste Poline**[3¶*]

[1]Graduate Program in Biological and Biomedical Engineering, McGill University, Montreal, Canada; [2]Department of Bioengineering, McGill University, Montreal, Canada; [3]Montreal Neurological Institute, McGill University, Montreal, Canada

***For correspondence:** mohammad.torabi@mail.mcgill.ca (Mohammad Torabi); jean-baptiste.poline@mcgill.ca (Jean-Baptiste Poline)

Present address: [§]Department of Bioengineering, McGill University, Canada; [¶]Department of Neurology and Neurosurgery, McGill University, Canada

Abstract Dynamic functional connectivity (dFC) has become an important measure for understanding brain function and as a potential biomarker. However, various methodologies have been developed for assessing dFC, and it is unclear how the choice of method affects the results. In this work, we aimed to study the results variability of commonly-used dFC methods. We implemented seven dFC assessment methods in Python and used them to analyze fMRI data of 395 subjects from the Human Connectome Project. We measured the pairwise similarity of dFC results using several similarity metrics in terms of overall, temporal, spatial, and inter-subject similarity. Our results showed a range of weak to strong similarity between the results of different methods, indicating considerable overall variability. Surprisingly, the observed variability in dFC estimates was comparable to the expected natural variation over time, emphasizing the impact of methodological choices on the results. Our findings revealed three distinct groups of methods with significant inter-group variability, each exhibiting distinct assumptions and advantages. These findings highlight the need for multi-analysis approaches to capture the full range of dFC variation. They also emphasize the importance of distinguishing neural-driven dFC variations from physiological confounds, and developing validation frameworks under a known ground truth. To facilitate such investigations, we provide an open-source Python toolbox that enables multi-analysis dFC assessment. This study sheds light on the impact of dFC assessment analytical flexibility, emphasizing the need for careful method selection and validation, and promoting the use of multi-analysis approaches to enhance reliability and interpretability of dFC studies.

Introduction

Functional Connectivity (FC) has emerged as an important measure for understanding brain function, and as a biomarker with significant potential (***Drysdale et al., 2017***; ***Etkin et al., 2019***; ***Reggente et al., 2018***; ***Lurie et al., 2020***). FC is typically assessed using blood-oxygen-level-dependent (BOLD) signals from functional magnetic resonance imaging (fMRI). FC assessment was initially performed under the assumption that it does not change over time (stationarity). However, growing evidence supports that the dynamic variations of FC (dynamic FC – dFC) play a critical role in brain functional organization and may provide a link between neural dynamics and cognition in healthy and diseased conditions (***Allen et al., 2014***; ***Hutchison et al., 2013***; ***Chang and Glover, 2010***). The significance of dFC extends beyond understanding interactions between brain regions. Recent studies have shown that patterns of dFC can reveal differences between healthy and diseased individuals and therefore could be used as a clinical biomarker. In this regard, several studies have reported

alterations in dFC patterns in a number of neurological or psychiatric conditions (autism spectrum disorders (ASD): ***de Lacy et al.*** (***2017***); attention-deficit/hyperactivity disorder (ADHD): ***de Lacy and Calhoun*** (***2018***); depression: ***Kaiser et al.*** (***2016***); post-traumatic stress disorder (PTSD): ***Jin et al.*** (***2017***); schizophrenia (SZ): ***Sakoglu et al.*** (***2010***); Parkinson's disease (PD): ***Díez-Cirarda et al.*** (***2018***); Alzheimer's disease (AD): ***Jones et al.*** (***2012***)).

In recent years, a variety of methodologies for assessing dFC have been developed (***Lurie et al., 2020***; ***Preti et al., 2017***). As the number of available methods continues to grow, there is an increasing need to comprehensively review these methodologies and examine their relative advantages and disadvantages, as well as converge towards recommendations for their application to experimental data (***Du et al., 2018***). To date, several studies have reviewed different aspects of these methods and have highlighted the importance of understanding their limitations and underlying assumptions (***Lurie et al., 2020***; ***Maltbie et al., 2022***; ***Savva et al., 2019***; ***Xie et al., 2019***; ***Du et al., 2018***; ***Keilholz et al., 2017***; ***Preti et al., 2017***; ***Chen et al., 2017***). However, only a few of these studies have compared different methods in practice (***Maltbie et al., 2022***; ***Savva et al., 2019***), and none have provided a comprehensive comparison of the results yielded by commonly-used dFC assessment methods.

The majority of studies that have applied dFC to various applications have not reported a clear justification for adopting a specific methodology for assessing dFC (***Maltbie et al., 2022***; ***Gonzalez-Castillo et al., 2015***; ***Jones et al., 2012***; ***de Lacy and Calhoun, 2018***; ***de Lacy et al., 2017***; ***Damaraju et al., 2014***; ***Engels et al., 2018***; ***Rashid et al., 2016***). For example, in clinical applications, some studies have employed sliding window and clustering methods (e.g. PTSD: ***Jin et al.*** (***2017***); PD: ***Engels et al.*** (***2018***); SZ: ***Damaraju et al.*** (***2014***); ***Du et al.*** (***2018***); ***Abrol et al.*** (***2017***)), while others have preferred Hidden Markov Model (HMM) approaches (e.g. SZ: ***Kottaram et al.*** (***2019***); PTSD: ***Ou et al.*** (***2015***); ***Charquero-Ballester et al.*** (***2022***); mild cognitive impairment (MCI): ***Suk et al.*** (***2016***)), or Time-Frequency methods (e.g. ASD: ***Al-Hiyali et al.*** (***2021***); ***Bernas et al.*** (***2018***); SZ: ***Yaesoubi et al.*** (***2017***)). It's worth noting that very few studies have used multiple methodologies (e.g. Chronic headache: ***Wang et al.*** (***2017***); disorders of consciousness (DOC): ***Cao et al.*** (***2019***)). Similarly, in cognitive and behavioral applications, some studies have used sliding window and clustering methods (e.g. task prediction: ***Gonzalez-Castillo et al.*** (***2015***); cognitive and behavioral flexibility: ***Uddin*** (***2021***)), some have used Co-Activation Patterns (CAP) analysis (e.g. naturalistic stimuli: ***Freitas et al.*** (***2020***); multiple tasks: ***Li et al.*** (***2021***)), HMM (e.g. Sleep stage: ***Stevner et al.*** (***2019***)), and very few have used multiple methods (e.g. working memory task: ***Chen et al.*** (***2015***)). A complete list of the reviewed studies can be found in Tables 1, 2, and 3 in Supplementary Data. Strikingly, out of the 60 dFC studies that we reviewed, only four considered more than one method to assess the robustness of their results. This is somewhat concerning since the variation in the dFC assessment results due to the choice of the assessment methodology, i.e. its analytical flexibility, has been studied on a relatively limited basis until now (***Maltbie et al., 2022***).

Analytical flexibility has recently become a critical issue in neuroimaging, particularly in fMRI data analysis (***Botvinik-Nezer et al., 2020***; ***Bhagwat et al., 2021***; ***Schilling et al., 2021***; ***Bowring et al., 2021***). Several studies have revealed significant variations in the results obtained from multiple independent analyses using the same dataset, highlighting the impact that analytical flexibility may have on scientific findings. For example, a study by ***Botvinik-Nezer et al.*** (***2020***) revealed considerable variation in the results reported by 70 independent teams using a common dataset, emphasizing the need for performing and comparing multiple analyses on the same dataset, as well as identifying the factors that explain the variation in the obtained results.

Similarly, the analytical flexibility of dFC assessment may greatly influence the reliability of the findings about brain functional organization and its alterations in diseased groups. Although several studies have investigated the reproducibility across individuals and scan sites of dFC patterns at rest (***Abrol et al., 2017***; ***Choe et al., 2017***; ***Vidaurre et al., 2018***), no study has fully investigated the analytical flexibility of dFC assessments. It is also unclear how methodological variations compare to the expected biological variations. The need for understanding dFC analytical flexibility in

the context of biological variability is a first critical step for result interpretation and for drawing conclusions about brain functional organization and its alterations in diseased groups.

Resting-state fMRI data lack a reliable biological reference for quantifying the scale of variability of dFC across assessment methods. It is well known that fMRI data are an indirect measure of the neural activity and are also modulated by physiological and motion processes (***Nikolaou et al., 2016***; ***Xifra-Porxas et al., 2021***). This further complicates setting a biological reference. However, dFC variation over time or over subjects can serve as fair biological references. Comparing the variability over methods to dFC temporal variability would yield an interpretable biological scale. In other words, if assessing dFC at one time point across methods results in as much variation as the one found across time points using one method, then variability between methods is as large as the scale of the biological phenomenon that they intend to capture. This approach enabled us to directly compare the magnitude of the effect of methodology choice to the underlying biological phenomenon. An alternative to variability in time is inter-subject variability, which can also help scale the magnitude of variations over method.

In this study we assess the analytical flexibility of dFC estimation by investigating how dFC patterns estimated from the same dataset vary across a large selection of methods. Among the available dFC assessment methods, we have selected seven widely used methods: Co-Activation Patterns (CAP) (***Liu and Duyn, 2013***), the clustering method (***Allen et al., 2014***), continuous Hidden Markov Models (HMM) (***Vidaurre et al., 2017***), discrete HMM (***Ou et al., 2015***), the sliding-window method (***Sakoglu et al., 2010***), the time-frequency (wavelet-based) method (***Chang and Glover, 2010***), and the window-less method (***Yaesoubi et al., 2018***).

The selected dFC assessment methods differ in terms of their basic underlying assumptions, ranging from the window-less method, which places a minimal set of assumptions to the HMM-based methods which place the most rigid assumptions (***Lurie et al., 2020***). Furthermore, some of these methods are state-based methods, meaning that they assume a finite number of group-level FC spatiotemporal patterns or "FC states" recurring over time (CAP, clustering, continuous HMM, discrete HMM, and window-less), while the rest are state-free methods which do not impose any constraints on the estimated FC patterns (sliding window and time-frequency). Moreover, the examined methods differ with regards to whether they assume locality of neighboring time points. Briefly, locality assumption means that the FC at each time point is inferred from the data points that are close in time (sliding window, time-frequency, clustering, continuous HMM, and discrete HMM), or not (CAP and window-less). The methods also differ in whether they consider the temporal ordering of data points (sliding window, time-frequency, clustering, continuous HMM, and discrete HMM) or disregard this information (CAP and window-less). State-based methods may also vary in terms of assuming smooth transitions between different states (continuous HMM, and discrete HMM) or allowing instantaneous reconfiguration (CAP, clustering, and window-less) (***Lurie et al., 2020***). Table1 lists these seven methods along with their number of citations (note that these numbers refer to the particular paper that introduced them, hence they may not be entirely reflective of the overall popularity of each method), year of publication, and assumptions. For the detailed pipeline used for each method, as well as their important hyperparameters, please see Figure4.

Comparing different methods for assessing dFC poses significant challenges due to several factors. The primary challenge arises from the differences in the underlying assumptions and the mathematical framework of various methods, coupled with the lack of a clear ground truth for measuring the accuracy of the results (***Lurie et al., 2020***; ***Xie et al., 2019***). Additionally, the distribution of estimated dFC values across methods can differ significantly, precluding the use of common similarity metrics such as the Euclidean distance or Pearson's correlation. The distribution of dFC values may not necessarily be Gaussian, and these values may considerably vary with regards to their mean and variance (see Figure14 in Supplementary Data). Furthermore, the output format of different methods varies, making direct comparisons more challenging. For instance, some methods, such as CAP and window-less, rely on first-order statistics and provide outputs in the form of

activity vectors. In contrast, other methods, such as the clustering method, encode dFC as FC matrices, which are second-order statistics. Additionally, state-free methods generate a continuous sequence of FC matrices over time, while state-based methods generate a set of state FC matrices and a state time course.

Comparing dFC methods lacks a general framework. Previous studies that have attempted to compare dFC methods have been limited to specific groups of methods or aspects of dFC. For instance, ***Maltbie et al.*** (***2022***) compared the results obtained by Clustering, CAP, and Phase Synchrony (PS) converted to a discrete set of states, thus limiting the comparisons to state-based methods. To make the results of CAP method comparable to others, they calculated the outer-product of the state mean-activity vectors obtained by this method to construct the corresponding FC matrices. Their comparison criteria were limited to visual comparison of spatial patterns of FC states and a few statistical outcomes derived from FC states, such as fractional occupancy. Similarly, ***Savva et al.*** (***2019***) has compared the results obtained by sliding window analysis using ten different correlation metrics. To the best of our knowledge, no studies have comprehensively compared results across the most common dFC assessment methods(state-based and state-free).

To address these challenges, we developed a comprehensive comparison framework applicable to state-based and state-free methods. This framework compares methods based on spatial, temporal, and overall similarity. We implemented the seven selected dFC assessment methods in a coherent Python package, providing the first open-source implementation of its kind. The seven implemented methods were then used to assess dFC using the resting-state fMRI data of 395 subjects from the Human Connectome Project (HCP) dataset (***Smith et al., 2013***; ***Van Essen et al., 2013***). The outputs were standardized into a common array format to allow for comparison of results from methods with different output formats. The resulting dFC arrays were then compared using the comparison framework. This framework employs different similarity metrics, including non-parametric metrics such as Spearman correlation, to deal with the different distributions that result from different dFC methods. This framework was developed to evaluate the variability of the results using two levels of comparison including subject-level and inter-subject level. Beyond assessing temporal and spatial variability on a subject-level, assessing inter-subject similarity enabled us to evaluate the ability of methods to capture a common biological phenomenon by examining the consistency of their inter-subject correlations. This is particularly relevant for studies focused on identifying population biomarkers. For a complete description of the comparison framework see Section Analytical Flexibility Assessment.

Our findings have important implications for the field of dFC and can guide future studies in several ways. Firstly, by comprehensively comparing seven widely used dFC assessment methods, we shed light on their analytical flexibility and the variability of their results, particularly when compared to the inherent variability of the biological phenomenon they aim to capture. This understanding is crucial for researchers and practitioners when choosing an appropriate method for their specific research question or clinical application. Secondly, our study provides a standardized framework for comparing dFC methods, addressing the challenges posed by their diverse assumptions, mathematical foundations, and output formats. This framework can serve as a basis for future comparative studies and facilitate the development and implementation of new dFC assessment methods. Lastly, our investigation of inter-subject similarity consistency provides insights into the ability of different methods to capture a common biological phenomenon, which is vital for identifying reliable biomarkers of brain function and disorders.

Results

DFC assessment methods can be grouped into 3 categories based on the similarity of their results

Figure1a shows the overall similarity obtained by Spearman correlation of the subject-specific dFC matrices computed using the seven selected methods (an illustration of sample dFC matrices ob-

Table 1. the seven selected dFC assessment methods and their number of citations, year of first publication, and assumptions.

Method	Number of citations	Year of publication	state-based/state-free	Assumes locality of neighboring time points	Imposes smooth transition between states	Considers temporal ordering of data points	Order of statistics used
Sliding Window (***Sakoglu et al., 2010***)	494	2010	state-free	Yes	NA	Yes	second-order
Time-Frequency (***Chang and Glover, 2010***)	1557	2010	state-free	Yes	NA	Yes	second-order
CAP (***Liu and Duyn, 2013***)	505	2013	state-based	No	No	No	first-order
Clustering (***Allen et al., 2014***)	2047	2014	state-based	Yes	No	Yes	second-order
Discrete HMM (***Ou et al., 2015***)	62	2014	state-based	Yes	Yes	Yes	second-order
Continuous HMM (***Vidaurre et al., 2017***)	398	2017	state-based	Yes	Yes	Yes	second-order
Window-less (***Yaesoubi et al., 2018***)	27	2018	state-based	No	No	No	first-order

tained by each method can be seen in Figure 2). The correlation values were averaged over subjects and repeated for each session, as indicated in equation (1) (for results obtained by other metrics see Supplementary Data: Pearson correlation, Figure20; Euclidean distance, Figure21; and mutual information, Figure22; see Figure18 in Supplementary data for the results of two-way ANOVA test on the effect of session (day and direction) on the overall similarity values). Correlation values show a range between weak to strong similarity between dFC methods. The average Spearman similarity of all pairs over all subjects was 0.38, indicating a moderate degree of similarity. However, the corresponding standard deviation was high (0.18; variance: 0.032), about half of the average similarity value. This variability was also significantly larger than the average variance of pairwise similarities over subjects (SD: 0.076, variance: 0.0058). In other words, dFC similarity variability over method pairs was five times higher than over subjects.

We used hierarchical clustering analysis on the pair-wise similarity matrices to investigate and summarize variation between dFC methods. The hierarchical clustering based on these average correlation matrices (Figure1b) suggests that the methods can be classified into three groups: 1) Clustering, Continuous HMM, and Discrete HMM; 2) CAP and Window-less; 3) Sliding Window and Time-Frequency. Although the definition of these groups highly depends on the cutoff value of the clustering, we observed high intra-group similarities and lower inter-group similarities for these 3 groups. Most subjects had the same hierarchical clustering as the one observed on the average across subjects (Figure17 in Supplementary Data).

The distribution of these similarity patterns over subjects (Figure16 in Supplementary Data) indicates that the similarity between certain pairs of methods, notably those involving Time-Frequency, displays greater variability over subjects. This observation suggests that the observed similarity between these methods is more likely influenced by individual differences in BOLD signals than mathematical similarity of their analyses. In other words, the variation in similarity across subjects implies that individual-specific characteristics may play an important role in the similarity between these methods. However, for other pairs of methods, more consistent similarity profiles were observed over subjects.

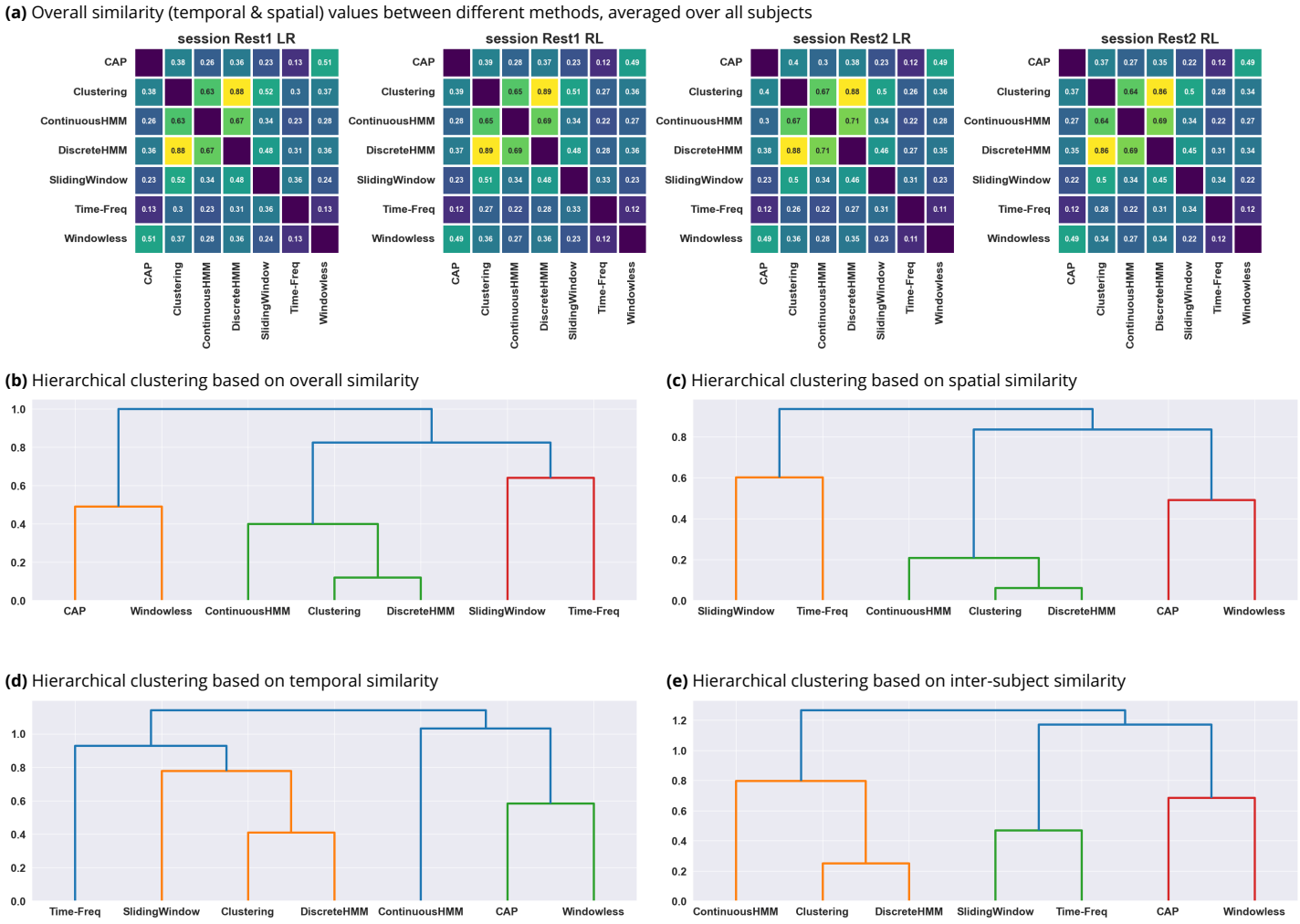


Figure 1. Similarity between dFC patterns obtained from seven methods as assessed by Spearman correlation: (a) The overall (combined spatial and temporal) similarity values averaged over subjects show considerable variation over different method pairs in all sessions; hierarchical clustering based on average similarity (b) overall (comparing dFC matrices), (c) spatial (comparing FC matrices at each time point), (d) temporal (comparing time courses of each functional connection), and (e) inter-subject (comparing inter-subject correlation values) measured by Spearman correlation. The hierarchical clustering results suggest that the methods can be grouped into three groups: 1) Clustering, Continuous HMM, and Discrete HMM; 2) CAP and Window-less; 3) Sliding Window and Time-Frequency. A slightly different pattern was obtained through temporal similarity assessment, which is due to decreased similarity of Continuous HMM and Time-Frequency with the methods in their group. The general similarity between the methods was relatively low, particularly for inter-group comparisons. For results obtained by other metrics see Supplementary Data: Pearson correlation, Figure20; Euclidean distance, Figure21; and mutual information, Figure22

.

Similarity in spatial patterns of dFC results follows the overall similarity patterns

The spatial similarity between the dFC matrices obtained from different methods was estimated using Spearman correlation between their corresponding spatial patterns (FC matrices) at each time point for each subject (see equation(2)). The resulting spatial similarity was then averaged over both time and subjects. Hierarchical clustering analysis of the average spatial similarity matrix (see Figure1c) revealed the same hierarchical structure as the one revealed by average overall similarity. The clustering analysis identified the same three groups, suggesting that their spatial similarities follow the same intra- and inter-group pattern.

Across methods dFC assessments reveal lower temporal similarity compared to spatial similarity

The temporal similarity between dFC matrices of each pair of methods was assessed using Spearman correlation between the time courses obtained by the methods for each functional connection (see equation(3)). The resulting temporal similarity matrices were then averaged over functional connections ($ROI \times ROI$) and subjects. While two methods may produce dFC matrices that have similar spatial patterns, the corresponding temporal patterns may differ. Across methods, spatial similarity focuses on the spatial patterns and values of functional connections at a given time, while temporal similarity focuses on the consistency of the time evolution of dFC for each functional connection. The hierarchical clustering based on the average over subjects temporal similarity matrix (Figure1d) reveals an overall decrease in similarity between methods compared to spatial (and overall) similarity. This was particularly noticeable for the similarities between CHMM and TF with other methods in their groups, which yielded considerably lower temporal similarity compared to the corresponding spatial similarity. This suggests that although CHMM and TF exhibit similar spatial patterns to the methods in their groups, they do not exhibit equally similar temporal dynamics (see Figure27 and Figure28 in Supplementary data).

Inter-subject similarities reflect the previously observed overall similarity patterns

To assess an inter-subject similarity between methods, dFC matrices estimated by each method were used to measure inter-subject correlations ($395 \times (395 - 1) / 2$ values) (see equation (4)). The inter-subject values obtained for each method were then compared and the similarity between methods was measured using Spearman correlation (see equation (5)). The resulting correlation values (Figure1e), indicating inter-subject correlation correspondence between methods, revealed the same three groups of methods in terms of their similarity. This implies that if one method identifies two subjects as similar based on their dFC patterns, it is more likely that another method from the same group will also identify those two subjects as similar. This finding suggests that the methods grouped together not only demonstrated similarity in capturing subject-level patterns but also exhibited consistency in capturing inter-subject patterns. For results obtained using Fractional Occupancy (FO) as the feature used for assessing the inter-subject correlations see Figure25 in Supplementary Data.

The variability over method is comparable to the variability over time for most functional connections

The variance of dFC, $dFC(subject, method, time, functionalConnection)$, was calculated over time and method, across functional connections and subjects. To account for the different value distributions of each method, the dFC matrices for each subject, $dFC(subject_n, method_i, time, functionalConnection)$, were normalized by ranking the values, similar to the approach used when calculating Spearman correlation. We calculated variance over time, $var(dFC(subject, method, :, functionalConnection))$, averaging over method, and the variance over method, $var(dFC(subject, :, time, functionalConnection))$, averaging over time. Therefore, both calculations resulted in an array of $subject \times functionalConnection$ values (one 'temporal variance' and one 'methods variance' value per functional connection and per subject). The results, shown in

Figure3a, indicate that the variation over dFC methods was comparable to the variation over time for most functional connections and subjects, leading to an average ratio of $var_{method}/var_{time} = 0.95$, or an average ratio of $SD_{method}/SD_{time} = 0.97$. For the variability of the results obtained across method pairs and method groups see Figure33 and Figure34 in Supplementary data. Moreover, Figure3b highlights the pair of RSNs with functional connections that exhibited higher variability over method than over time. These functional connections mainly include connections between the Default Mode and other networks, such as Parieto-Occipital, FrontoParietal, Salience, Cingulo-Opercular, MedialParietal, Dorsal Attention, and Ventral Attention Networks, as well as most intra-network connections, with the exception of the Cingulo-Opercular Network.

We also compared the variability over method with the variability over subject. After normalizing the dFC matrices by ranking their values, we calculated the variance over method, $var(dFC(subject, :, time, functionalConnection))$, averaging over time and subject, and the variance over subject, $var(dFC(:, method, time, functionalConnection))$, averaging over time and method. Both calculations resulted in an array of $functionalConnection$ values. This yielded an average ratio of $var_{method}/var_{subj} = 0.90$, or an average ratio of $SD_{method}/SD_{subj} = 0.94$ (see Figure35 in Supplementary data).

Discussion

The variability observed in dFC assessment encourages multi-analysis studies

In the present study, our main aim was to assess the analytical flexibility of dFC assessment methods. One of the primary challenges of comparing dFC assessment methods is the lack of a biological reference to determine the significance of the observed variability over methods. While no clear biological reference exists for resting-state fMRI data, we compared the assessed variability over methods with connection variation over time as well as inter-subject variability. The results (Figure3a and Figure35 in Supplementary data) suggest that the variance over methods is on average 0.95 of the variance over time and 0.90 of the variance over subjects. These findings indicate that changing the assessment method may lead to a dFC value which is on average as different as a dFC value obtained at another time point or from another subject using the same method. Moreover, although these biological scales may not be appropriate for every situation, they still can serve as a valid point of reference. Specifically, in the case of comparison with temporal variability, the variation of dFC over time is the main biological phenomenon that the dFC methods aim to measure. These results indicate the variation due to the choice of methodology has an impact as large as the scale of the underlying biological phenomenon of interest.

For further investigation of the analytical flexibility, we conducted a comprehensive comparison of the dFC estimates obtained by different methods using the following approaches: 1) subject-wise comparison of dFC matrices in terms of overall similarity (Figure1b); 2) time point-wise comparison of FC matrices in terms of their spatial similarity (Figure1c); 3) functional connection-wise comparison of time courses in terms of their temporal similarity (Figure1d); 4) comparison in terms of inter-subject correlation similarity (Figure1e). Our similarity analysis identified three groups of methods in terms of the similarity of the obtained dFC patterns. The examined methods, while exhibiting fair intra-group similarities, exhibited considerable inter-group variability.

The identified intra- and inter-group similarity profiles can be largely attributed to the underlying assumptions of the examined methods. For example, both WL and CAP disregard the temporal ordering of the BOLD data, allow for instantaneous reconfiguration of FC states, and do not impose a locality assumption. These assumptions enables them to capture more rapid reconfigurations in FC patterns. Conversely, SWC, CHMM, and DHMM, take into account the temporal ordering of the data, and assume locality of neighboring time points, resulting in smoother changes over time (see Figure9 in Supplementary data). These assumptions allow these methods to accurately quantify the temporal dependencies between FC patterns at different time points and the slower dynamics of FC. Lastly, SW and TF, do not assume the presence of re-occurring group-level FC patterns, which

allows them to capture the individual subject variability to a greater extent. See Table1 for a summary of the assumptions of each method. These assumptions are rarely invoked in the literature applying dFC technique in a specific study.

The patterns of similarity observed in this study do not imply the superiority of any dFC assessment method or group of methods. Instead, they reveal three groups of methods, each characterized by their specific assumptions, advantages, and disadvantages. These results show that a single analysis or group of analyses may not capture the richness of dFC, and that the use of multiple methods from different groups may reveal different features of dFC. In this regard, developing specialized tools to facilitate the conjoint use of multiple dFC methods, such as the one developed in this study, will aid to fully characterize dFC. For instance, by using one method from the SWC, CHMM, and DHMM group, one from the WL and CAP group, and one from the SW and TF group one can capture rapid FC patterns, as well as temporal dependencies of FC patterns, and subject-specific information.

This approach is particularly important when classifying subjects based on features derived from dFC, as shown in our inter-subject similarity results (Figure1e). We found that using methods from two different groups, e.g., CAP and SWC, can lead to different inter-subject similarity patterns, consequently leading to different clusters of subjects. Hence, we recommend that clustering is performed using multiple methods from different groups. This can help uncover meaningful subgroups or relationships that may not be apparent when using a single method, and ensure that the identified clusters are not solely dependent on the choice of a specific method.

Understanding temporal complexity in dynamic functional connectivity assessment: what lies behind the variations in dynamic functional connectivity?

Evaluating the overall similarity among the outcomes of various methods offers a perspective on the inter-method similarity relationships. However, it remains unclear to what extent these similarities can be attributed to the spatial and temporal aspects of dFC. Based on our results, in most cases, while the spatial comparison of dFC patterns obtained by different methods (Figure1c) exhibited strong similarity between methods, the comparison in terms of temporal dynamics exhibited moderate similarity (Figure1d). Even the methods with the highest spatial similarity, SWC and DHMM, showed low similarity in the time domain. This suggests that capturing the temporal variations of FC may be more challenging than capturing the spatial patterns of dFC. Figure27 in Supplementary data shows that although methods having higher spatial similarity are more likely to have higher temporal similarity, the level of temporal similarity is smaller than the level of spatial similarity in most cases. CHMM, as a particular example, exhibited a high spatial similarity with SWC and DHMM, but a low temporal similarity with these methods (also supported by Figure26 in Supplementary data showing the similarity of inter-time correlations).

Our supplementary material results suggest that in several cases, pairs of methods did not exhibit a significantly greater similarity compared to the similarity obtained when the dFC matrices were randomly shuffled with respect to time (Figure28; note that the fractional occupancy of the states was preserved for the shuffled dFC matrices). Pairs of methods exhibiting strong similarity between their time-shuffled dFC matrices indicate that the average similarity between all pairs of their spatial patterns, or FC matrices, even at different time points, is high. For instance, CHMM-SWC exhibited a relatively high overall similarity, however, their similarity did not change significantly when the dFC matrices were time-shuffled. This implies that the dFC matrices of CHMM and SWC did not exhibit more consistency in terms of their patterns of variation over time than the time-shuffled dFC matrices and that the observed strong similarity between them was mostly a result of the strong similarity between their spatial patterns. On the other hand, WL-CAP, despite having a low overall similarity, showed significantly greater similarity when compared to the shuffled-time similarity distribution. This implies that their temporal variations exhibited more consistent patterns than those of the time-shuffled dFC matrices, indicating that their low overall similarity was caused by low similarity of their spatial patterns. These observations highlight that the variability

in the dFC results may spread differently across the temporal and spatial dimensions, and strong similarity in one dimension does not necessarily imply similarity for the other dimension. It also suggests that comparing the overall similarity of dFC results, or comparing only their spatial patterns may not provide a comprehensive understanding of the relation between methods.

Future research should conduct more in-depth comparisons of dFC assessment methods, focusing on the temporal dynamics of their outputs, to shed light on the reason for the observed variability over methods. This is crucial, as temporal dynamics play a significant role in dFC and have already been a source of controversy regarding their origins. Our results suggest that a part of these temporal variations may originate from the methodology used for assessing them.

Previous studies have demonstrated that physiological factors, such as respiration, cardiac activity, and motion artifacts, also modulate the temporal variations of dFC (***Nikolaou et al., 2016***; ***Xifra-Porxas et al., 2021***). These non-neural sources of variability can confound the interpretation of dFC patterns and potentially lead to erroneous conclusions. Future research could develop advanced techniques, including denoising algorithms, physiological signal regression, and data-driven approaches, to effectively isolate the neural-driven part of dFC variations from the physiological sources. By disentangling these sources, researchers could gain a more precise understanding of the genuine neural dynamics underlying dFC, enhancing the reliability of dFC studies and their potential applications in clinical diagnostics and treatment evaluations.

Study limitations and future directions

Potential future work should review external validation techniques and develop a validation framework to assess the accuracy of dFC assessments in more controlled paradigms, such as predicting cognitive state of the subjects, using simulated data, and assessing the consistency of the result with the neural activation patterns derived from simultaneous electrophysiological recordings. Evaluating the dFC assessment methods using such paradigms will lead to a better understanding of the reasons behind the observed variability between the results yielded by different dFC methods. Furthermore, it would shed light on the varying levels of method variability observed across different RSNs by identifying the factors that contribute to higher variability in certain RSN pairs compared to others. Additionally, this would also provide a more reliable biological reference for quantifying the magnitude of the observed methods' variability. One of the limitations of the current study is the lack of a clear biological reference. The nature of resting-state data makes it challenging to establish expected patterns or variations in dFC over time or across subjects. This is primarily because subjects are free to engage in any mental activity during the resting state (***Cohen, 2018***; ***Lurie et al., 2020***). In contrast, controlled paradigms, such as the ones mentioned earlier, involve specific instructions or tasks, resulting in more deterministic dFC patterns. For instance, in task-based paradigms, subjects perform specific tasks, leading to more comparable dFC patterns and enhanced interpretability of inter-subject variability. Similarly, simultaneous fMRI and electrophysiological recordings can provide a more reliable biological reference through the integration of electrophysiological activity. Therefore, future work should prioritize the use of such paradigms to facilitate the utilization of more interpretable biological references.

It is worth noting that in our study, we adopted the hyperparameter values recommended by the original paper or consensus among the community for each method, which may contribute to the observed variability in our results. Different hyperparameter choices could lead to different similarity patterns among methods and varying levels of variability. Therefore, future work should also investigate the sensitivity of dFC assessment results to the choice of hyperparameters and their effect on the observed variability.

Assessing the robustness of neuroscience results with respect to dFC analytical flexibility should be strongly recommended. We propose the utilization of multi-analysis approaches, preferably using methods that show less similar results, for assessing dFC and reporting a more comprehensive view of the data, made easier with the python toolbox released with this article. While the use of several methods may cause interpretation and statistical challenges, assessing the variability of the

results is important. As this may not always be feasible, it is also reasonable to select a single dFC method based on the specific research objectives, method strengths and limitations, compatibility with the study design and data characteristics.

We have developed a Python toolbox, termed `multi_analysis_dfc`, that simplifies the implementation of multi-analysis dFC assessment. The toolbox enables users to apply various dFC assessment methods to their data and experiment with different hyperparameters, allowing for a comprehensive analysis of dFC variability over methods and hyperparameters. Users can compare the results of different methods within a single package, making it easy to explore and evaluate dFC assessment techniques. This toolbox is unique in providing an open-source implementation of commonly used dFC methods and should be a valuable resource for researchers seeking to explore and compare different dFC assessment techniques. The toolbox is available on GitHub and is designed to be a collaborative package where other methods implementation can be integrated.

Conclusions

In conclusion, the present study aimed to evaluate the analytical flexibility of dFC measurement methods and establish how analytical flexibility compares to biological variability. The comparison results show a wide range of similarity patterns across methods. The variability found in the results highlights the importance of carefully motivating and validating the choice of a specific dFC assessment method. The use of multiple methods would mitigate the issues arising from analytical flexibility and better characterize the richness of dFC measurements. Future work should aim to develop a validation framework and evaluate the accuracy of dFC assessments through external validation techniques.

Methods and Materials

Data

The data used in this study are Blood-oxygen-level-dependent (BOLD) time series of resting-state fMRI data of 395 young, and healthy subjects (age range: 22-36 years) from the S1200 release of the 3T Human Connectome Project (HCP) dataset (***Smith et al., 2013***; ***Van Essen et al., 2013***), preprocessed as described in ***Xifra-Porxas et al.*** (***2021***). Each subject was scanned during four sessions on two different days. Each day included two 15-min scans, one acquired with left-to-right (LR) phase encoding direction and the other one with right-to-left (RL). In this study, except Figure1a, where we show results from all four sessions, all other main results are shown using data from the first session `Rest1_LR`. Each scan consists of 1200 time points with TR=0.72 sec. We used the HCP FIX-denoised data, which includes additional denoising steps such as de-trending, head motion correction, and denoising via spatial ICA compared to the minimally preprocessed data available in the HCP database. The data were then parcellated into 333 regions of interest (ROIs) using the Gordon parcellation (***Gordon et al., 2016***), an atlas widely used in the literature. Each of the ROIs belongs to a specific Resting State Network (RSN). Among these 333 ROIs, 47 did not belong to any brain network and therefore were excluded, resulting in 286 ROIs. For computational reasons, we have uniformly downsampled these 286 ROIs to 96 ROIs. Each ROI time series was z-standardized before the analysis.

State-free Methods

These methods are model free and therefore do not need to be implemented at the group level before dFC assessment. They were applied to each subject separately. See Figure 4 for an illustration of state-free methods and their dFC matrices.

Sliding Window (SW) (***Sakoglu et al., 2010***)

A sliding tapered window of length 44s, or 60 time points (following ***Allen et al.*** (***2014***)), and 50 percent overlap was used to find the FC of ROI pairs over time. The tapered window was generated by convolving a rectangle of length 44s with a Gaussian curve with $\sigma = 3$ TRs following (***Allen***

et al., 2014) and then moved along 1200 time points with a 30 time points overlap. This procedure resulted in 38 windows and hence a $38 \times ROI \times ROI$ dFC matrix. The main hyperparameters of this method are window length, overlap ratio, and tapered or rectangular window.

Time-Frequency (TF) (***Chang and Glover, 2010***)

The wavelet transform was performed by applying $k = 101$ wavelet kernels on each ROI time course, resulting in a wavelet time series with k frequency scales. Next, the Wavelet Transform Coherence (WTC) was computed between the time courses of each ROI pair for each frequency scale and averaged across all frequency scales following ***Savva et al.*** (***2021***). The resulting pairwise WTC time courses formed the corresponding dFC matrices of shape $1200 \times ROI \times ROI$. Note that following ***Chang and Glover*** (***2010***), only the WTC values outside the cone of influence were considered for the final dFC matrix. The hyperparameters of this method include number of frequency scales, wavelet type, significance level of coherence magnitude, and considering the cone of influence or not.

State-based Methods

Since state-based methods assume a finite number of FC states re-occurring over time and all subjects, they initially need to be applied on a group level to identify the group-level FC states. Each state is represented by a state FC matrix of shape $ROI \times ROI$ which corresponds to the FC pattern at the time points that the state occurs. Next, the fitted models are applied to each individual subject, and individual dFC matrices are obtained. Applying the fitted models on subject-level data assigns each time point to one of the group-level FC states, resulting in a state time course. The state FC matrix of each state is then placed at the time points that belong to that state to build the dFC matrix of that subject. There is still no consensus on the number of brain states. In this study, to make the results more comparable, the number of FC states was assumed to be 12 for all state-based methods following ***Vidaurre et al.*** (***2017***). Figures 1 through 5 in Supplementary Data shows the state FC matrices obtained by each state-based method. See Figure 4 for an illustration of state-based methods and their state FC matrices and state time courses.

Co-Activation Patterns (CAP) (***Liu and Duyn, 2013***)

The Co-Activation Patterns (CAP) method is a point-process analysis with individual time point time resolution and very few assumptions. The original CAP method used activation thresholding within a seed region and subsequent averaging of the BOLD time series at significant time points (***Liu and Duyn, 2013***). The original method does not directly yield dFC matrices. The extended version of CAP analysis (***Liu et al., 2013***) applies k-means clustering directly to BOLD time series. In the present study, for computational reasons, the clustering was performed in two stages following a process similar to ***Ou et al.*** (***2015***). First, we clustered the time points of each subject (vectors of $length = ROI$) to find subject-level cluster centroids. Next, we clustered the resulting cluster centroids from all subjects to identify 12 group-level cluster centroids, corresponding to 12 FC states. Finally, we obtained the state FC matrices following the implementation proposed by ***Maltbie et al.*** (***2022***), by calculating the $ROI \times ROI$ outer-product matrix of centroids with themselves. The clustering labels of time points were considered as the state time courses. The shape of the final dFC matrix was $1200 \times ROI \times ROI$. The hyperparameters of this method are the number of subject-level clusters, and the number of FC states.

Sliding Window + Clustering (SWC) (***Allen et al., 2014***)

This method relies on the dFC matrix assessed using the Sliding Window method. Therefore, we initially assessed the dFC matrix using the Sliding Window method as described above, which had a shape of $38 \times ROI \times ROI$. Next, we considered each time point of the dFC matrix as a sample and vectorized it to a feature vector of length $ROI \times (ROI - 1)/2$. Then, using a two-level k-means clustering procedure similar to the one described for the CAP method, we clustered feature vectors

into 12 clusters corresponding to 12 FC states. The FC matrix of each state was then obtained by reshaping the centroid vectors to their original shape of $ROI \times ROI$. The resulting clustering label sequences were considered as the state time courses. The shape of the final dFC matrix was $38 \times ROI \times ROI$. The hyperparameters of this method are those of sliding window method, number of subject-level clusters, and number of FC states.

Continuous Hidden Markov Model (CHMM) (***Vidaurre et al., 2017***)

In this method, a continuous Hidden Markov Model (continuous HMM or CHMM) with a Gaussian observation model was used. BOLD time series from resting-state fMRI data were directly used as the continuous observation sequences of the HMM. Subsequently, 12 hidden states were identified, each corresponding to a FC state. Each hidden state was represented by a multivariate Gaussian Model. The mean and covariance of each hidden state were inferred by fitting the model on the concatenated BOLD time series of all subjects. Next, the fitted model was applied on the subject-level data to infer the sequence of hidden states for each subject. The inferred sequence was then used as the state time course and the covariance matrix of each hidden state was considered as the corresponding state FC matrix. The shape of the final dFC matrix was $1200 \times ROI \times ROI$. The only hyperparameters of this method are the observation model type (e.g. Gaussian, or Autoregressive), and number of FC states.

Discrete Hidden Markov Model (DHMM) (***Ou et al., 2015***)

For this method, a discrete Hidden Markov Model (discrete HMM or DHMM) with a binomial observation model was used. Since this method is based on the results of the clustering method, we initially assessed dFC matrices using the clustering method, which had a shape of $38 \times ROI \times ROI$. The number of FC states assumed for this initial clustering analysis was equal to the number of observations assumed for the discrete HMM method, and not the number of final FC states. Furthermore, the DHMM used here requires a discrete observation sequence. Therefore, the state time courses obtained by the clustering method were used as the discrete observation sequences of the discrete HMM. The HMM was then fitted to the observation sequences to identify 12 hidden states corresponding to the 12 FC states. The resulting hidden state sequences were used as the state time courses and the FC matrix of each FC state was obtained by averaging all FC matrices assigned to that FC state in the dFC matrices obtained by the clustering method. The shape of the final dFC matrix was $38 \times ROI \times ROI$. The hyperparameters of this method include hyperparameters of the clustering method, number of observations, and number of FC states.

Window-less (WL) (***Yaesoubi et al., 2018***)

This method relies on estimating the dominant linear patterns in the sample space of all BOLD time series values. Each dominant linear pattern corresponds to one of the FC states. To estimate the dominant linear patterns, a sparse dictionary learning algorithm, the $k-SVD$ (***Aharon et al., 2006***) algorithm, was applied to the group-level data. Dominant linear patterns are each represented by a dictionary element. The 12 dictionary elements corresponding to the 12 FC states were stored in the dictionary matrix D. Each time point was approximated by a linear combination of these dictionary elements. The coefficients of the linear combinations for each time point were stored in rows of the mixing matrix M. A hard sparsity constrain was applied to the algorithm so that it assigns one and only one dictionary to each time point, suggesting that rows of M were sparse. The state FC matrices were obtained by calculating the outer-product matrices of the dictionary elements in the columns of D. The state time course was calculated using the mixing matrix M. The shape of the final dFC matrix was $1200 \times ROI \times ROI$. The only hyperparameter of this method is the number of FC states.

Analytical Flexibility Assessment

We used Python to implement all seven methods and estimate the dFC matrix for each of the 395 preprocessed fMRI subjects' data, resulting in a total of 2765 dFC matrices. Figure 2 shows

representative dFC matrices obtained by different methods using the BOLD data of one subject. To ensure comparability, as sliding window based methods (SW, SWC, and DHMM) downsample the time samples, the dFC matrices of other methods were uniformly downsampled prior to further analysis to match the resolution of the sliding window based methods. The output of each method was converted to a common format of a $38 \times 96 \times 96$ dFC matrix, resulting in 7 dFC matrices of the same size per subject. This resulted in a dFC results array of dimensions $dFC(subject, method, time, ROI, ROI) \in (395, 7, 38, 96, 96)$. The dFC results array was also reshaped in some cases to $dFC(subject, method, time, functionalConnection) \in (395, 7, 38, 4560)$ by vectorizing the lower triangle of 2D FC matrices of shape $ROI \times ROI$ into $functionalConnection \times 1$. To assess the analytical flexibility of the dFC results across methods, we calculated the $7 \times (7-1)/2$ pairwise similarity using several similarity metrics, including Spearman correlation, Pearson correlation, Euclidean distance, and mutual information. We evaluated the similarity between each pair of methods in terms of correlation of their: dFC matrices $dFC(subject, method, :, :, :)$ (overall similarity), time courses of functional connections $dFC(subject, method, :, ROI, ROI)$ (temporal similarity), FC patterns at each time point $dFC(subject, method, time, :, :)$ (spatial similarity), and graph properties such as degree and clustering coefficient (see Figure 23 and Figure 24 in Supplementary Data). The equations used to calculate each of these similarity metrics are as follows:

$$similarity_{overall}(subj_n, method_i, method_j) = corr(dFC(subj_n, method_i, :, :, :), dFC(subj_n, method_j, :, :, :)) \quad (1)$$

$$similarity_{spatial}(subj_n, method_i, method_j, time_t) = corr(dFC(subj_n, method_i, time_t, :, :), dFC(subj_n, method_j, time_t, :, :)) \quad (2)$$

$$similarity_{temporal}(subj_n, method_i, method_j, ROI_k, ROI_l) = corr(dFC(subj_n, method_i, :, ROI_k, ROI_l), dFC(subj_n, method_j, :, ROI_k, ROI_l)) \quad (3)$$

Additionally, we compared the correspondence between the inter-subject correlations of different methods to obtain the inter-subject similarity:

$$interSubjCorr(subj_a, subj_b, method_i) = corr(dFC(subj_a, method_i, :, :, :), dFC(subj_b, method_i, :, :, :)) \quad (4)$$

$$similarity_{inter-subject}(method_i, method_j) = corr(interSubjCorr(:, :, method_i), interSubjCorr(:, :, method_j)) \quad (5)$$

To identify groups of methods that produce similar results, we used hierarchical clustering with the Ward method to analyze the similarity values calculated using the aforementioned metrics (Spearman correlation, Pearson correlation, Euclidean distance, and mutual information). For correlation-based metrics, $1 - correlation$ was used as the distance between the methods. The hierarchical clustering allowed us to summarize the measured similarity matrices and, moreover, group together methods that exhibited a higher degree of similarity.

References

Abrol A, Damaraju E, Miller RL, Stephen JM, Claus ED, Mayer AR, Calhoun VD. Replicability of time-varying connectivity patterns in large resting state fMRI samples. NeuroImage. 2017 Dec; 163:160–176. https://www.sciencedirect.com/science/article/pii/S1053811917307681, doi: 10.1016/j.neuroimage.2017.09.020.

Aharon M, Elad M, Bruckstein A. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. IEEE Transactions on Signal Processing. 2006 Nov; 54(11):4311–4322. doi: 10.1109/TSP.2006.881199, conference Name: IEEE Transactions on Signal Processing.

Al-Hiyali MI, Yahya N, Faye I, Hussein AF. Identification of Autism Subtypes Based on Wavelet Coherence of BOLD FMRI Signals Using Convolutional Neural Network. Sensors (Basel, Switzerland). 2021 Aug; 21(16):5256. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8398492/, doi: 10.3390/s21165256.

Allen EA, Damaraju E, Plis SM, Erhardt EB, Eichele T, Calhoun VD. Tracking Whole-Brain Connectivity Dynamics in the Resting State. Cerebral Cortex. 2014 Mar; 24(3):663–676. https://doi.org/10.1093/cercor/bhs352, doi: 10.1093/cercor/bhs352.

Bernas A, Aldenkamp AP, Zinger S. Wavelet coherence-based classifier: A resting-state functional MRI study on neurodynamics in adolescents with high-functioning autism. Computer Methods and Programs in Biomedicine. 2018 Feb; 154:143–151. https://www.sciencedirect.com/science/article/pii/S0169260717303097, doi: 10.1016/j.cmpb.2017.11.017.

Bhagwat N, Barry A, Dickie EW, Brown ST, Devenyi GA, Hatano K, DuPre E, Dagher A, Chakravarty M, Greenwood CMT, Misic B, Kennedy DN, Poline JB. Understanding the impact of preprocessing pipelines on neuroimaging cortical surface analyses. GigaScience. 2021 Jan; 10(1):giaa155. doi: 10.1093/gigascience/giaa155.

Botvinik-Nezer R, Holzmeister F, Camerer CF, Dreber A, Huber J, Johannesson M, Kirchler M, Iwanir R, Mumford JA, Adcock RA, Avesani P, Baczkowski BM, Bajracharya A, Bakst L, Ball S, Barilari M, Bault N, Beaton D, Beitner J, Benoit RG, et al. Variability in the analysis of a single neuroimaging dataset by many teams. Nature. 2020 Jun; 582(7810):84–88. https://www.nature.com/articles/s41586-020-2314-9, doi: 10.1038/s41586-020-2314-9, number: 7810 Publisher: Nature Publishing Group.

Bowring A, Maumet C, Nichols TE. Erratum: Exploring the impact of analysis software on task fMRI results. Human Brain Mapping. 2021; 42(5):1564–1578. https://onlinelibrary.wiley.com/doi/abs/10.1002/hbm.25302, doi: 10.1002/hbm.25302, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/hbm.25302.

Cao B, Chen Y, Yu R, Chen L, Chen P, Weng Y, Chen Q, Song J, Xie Q, Huang R. Abnormal dynamic properties of functional connectivity in disorders of consciousness. NeuroImage: Clinical. 2019 Jan; 24:102071. https://www.sciencedirect.com/science/article/pii/S2213158219304188, doi: 10.1016/j.nicl.2019.102071.

Chang C, Glover GH. Time–frequency dynamics of resting-state brain connectivity measured with fMRI. NeuroImage. 2010 Mar; 50(1):81–98. https://www.sciencedirect.com/science/article/pii/S1053811909012981, doi: 10.1016/j.neuroimage.2009.12.011.

Charquero-Ballester M, Kleim B, Vidaurre D, Ruff C, Stark E, Tuulari JJ, McManners H, Bar-Haim Y, Bouquillon L, Moseley A, Williams SCR, Woolrich MW, Kringelbach ML, Ehlers A. Effective psychological therapy for PTSD changes the dynamics of specific large-scale brain networks. Human Brain Mapping. 2022; 43(10):3207–3220. https://onlinelibrary.wiley.com/doi/abs/10.1002/hbm.25846, doi: 10.1002/hbm.25846, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/hbm.25846.

Chen JE, Chang C, Greicius MD, Glover GH. Introducing co-activation pattern metrics to quantify spontaneous brain network dynamics. NeuroImage. 2015 May; 111:476–488. https://www.sciencedirect.com/science/article/pii/S105381191500083X, doi: 10.1016/j.neuroimage.2015.01.057.

Chen JE, Rubinov M, Chang C. Methods and Considerations for Dynamic Analysis of Functional MR Imaging Data. Neuroimaging Clinics. 2017 Nov; 27(4):547–560. https://www.neuroimaging.theclinics.com/article/S1052-5149(17)30067-9/fulltext, doi: 10.1016/j.nic.2017.06.009, publisher: Elsevier.

Choe AS, Nebel MB, Barber AD, Cohen JR, Xu Y, Pekar JJ, Caffo B, Lindquist MA. Comparing test-retest reliability of dynamic functional connectivity methods. NeuroImage. 2017 Sep; 158:155–175. https://www.sciencedirect.com/science/article/pii/S1053811917305736, doi: 10.1016/j.neuroimage.2017.07.005.

Cohen JR. The behavioral and cognitive relevance of time-varying, dynamic changes in functional connectivity. NeuroImage. 2018 Oct; 180:515–525. https://linkinghub.elsevier.com/retrieve/pii/S105381191730784X, doi: 10.1016/j.neuroimage.2017.09.036.

Damaraju E, Allen EA, Belger A, Ford JM, McEwen S, Mathalon DH, Mueller BA, Pearlson GD, Potkin SG, Preda A, Turner JA, Vaidya JG, van Erp TG, Calhoun VD. Dynamic functional connectivity analysis reveals transient states of dysconnectivity in schizophrenia. NeuroImage: Clinical. 2014 Jan; 5:298–308. https://www.sciencedirect.com/science/article/pii/S2213158214000953, doi: 10.1016/j.nicl.2014.07.003.

Drysdale AT, Grosenick L, Downar J, Dunlop K, Mansouri F, Meng Y, Fetcho RN, Zebley B, Oathes DJ, Etkin A, Schatzberg AF, Sudheimer K, Keller J, Mayberg HS, Gunning FM, Alexopoulos GS, Fox MD, Pascual-Leone A, Voss HU, Casey BJ, et al. Resting-state connectivity biomarkers define neurophysiological subtypes

of depression. Nature Medicine. 2017 Jan; 23(1):28–38. https://www.nature.com/articles/nm.4246, doi: 10.1038/nm.4246, bandiera_abtest: a Cg_type: Nature Research Journals Number: 1 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Depression;Diagnostic markers;Predictive markers Subject_term_id: depression;diagnostic-markers;predictive-markers.

Du Y, Fryer SL, Fu Z, Lin D, Sui J, Chen J, Damaraju E, Mennigen E, Stuart B, Loewy RL, Mathalon DH, Calhoun VD. Dynamic functional connectivity impairments in early schizophrenia and clinical high-risk for psychosis. NeuroImage. 2018 Oct; 180:632–645. https://www.sciencedirect.com/science/article/pii/S1053811917308492, doi: 10.1016/j.neuroimage.2017.10.022.

Díez-Cirarda M, Strafella AP, Kim J, Peña J, Ojeda N, Cabrera-Zubizarreta A, Ibarretxe-Bilbao N. Dynamic functional connectivity in Parkinson's disease patients with mild cognitive impairment and normal cognition. NeuroImage: Clinical. 2018 Jan; 17:847–855. https://www.sciencedirect.com/science/article/pii/S2213158217303182, doi: 10.1016/j.nicl.2017.12.013.

Engels G, Vlaar A, McCoy B, Scherder E, Douw L. Dynamic Functional Connectivity and Symptoms of Parkinson's Disease: A Resting-State fMRI Study. Frontiers in Aging Neuroscience. 2018 Nov; 10:388. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6266764/, doi: 10.3389/fnagi.2018.00388.

Etkin A, Maron-Katz A, Wu W, Fonzo GA, Huemer J, Vértes PE, Patenaude B, Richiardi J, Goodkind MS, Keller CJ, Ramos-Cejudo J, Zaiko YV, Peng KK, Shpigel E, Longwell P, Toll RT, Thompson A, Zack S, Gonzalez B, Edelstein R, et al. Using fMRI connectivity to define a treatment-resistant form of post-traumatic stress disorder. Science Translational Medicine. 2019 Apr; https://www.science.org/doi/abs/10.1126/scitranslmed.aal3236, doi: 10.1126/scitranslmed.aal3236, publisher: American Association for the Advancement of Science.

Freitas LGA, Bolton TAW, Krikler BE, Jochaut D, Giraud AL, Hüppi PS, Van De Ville D. Time-resolved effective connectivity in task fMRI: Psychophysiological interactions of Co-Activation patterns. NeuroImage. 2020 May; 212:116635. https://www.sciencedirect.com/science/article/pii/S1053811920301221, doi: 10.1016/j.neuroimage.2020.116635.

Gonzalez-Castillo J, Hoy CW, Handwerker DA, Robinson ME, Buchanan LC, Saad ZS, Bandettini PA. Tracking ongoing cognition in individuals using brief, whole-brain functional connectivity patterns. Proceedings of the National Academy of Sciences. 2015 Jul; 112(28):8762–8767. https://www.pnas.org/doi/full/10.1073/pnas.1501242112, doi: 10.1073/pnas.1501242112, publisher: Proceedings of the National Academy of Sciences.

Gordon EM, Laumann TO, Adeyemo B, Huckins JF, Kelley WM, Petersen SE. Generation and Evaluation of a Cortical Area Parcellation from Resting-State Correlations. Cerebral Cortex. 2016 Jan; 26(1):288–303. https://doi.org/10.1093/cercor/bhu239, doi: 10.1093/cercor/bhu239.

Hutchison RM, Womelsdorf T, Allen EA, Bandettini PA, Calhoun VD, Corbetta M, Della Penna S, Duyn JH, Glover GH, Gonzalez-Castillo J, Handwerker DA, Keilholz S, Kiviniemi V, Leopold DA, de Pasquale F, Sporns O, Walter M, Chang C. Dynamic functional connectivity: promise, issues, and interpretations. NeuroImage. 2013 Oct; 80:360–378. doi: 10.1016/j.neuroimage.2013.05.079.

Jin C, Jia H, Lanka P, Rangaprakash D, Li L, Liu T, Hu X, Deshpande G. Dynamic brain connectivity is a better predictor of PTSD than static connectivity. Human Brain Mapping. 2017; 38(9):4479–4496. https://onlinelibrary.wiley.com/doi/abs/10.1002/hbm.23676, doi: 10.1002/hbm.23676, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/hbm.23676.

Jones DT, Vemuri P, Murphy MC, Gunter JL, Senjem ML, Machulda MM, Przybelski SA, Gregg BE, Kantarci K, Knopman DS, Boeve BF, Petersen RC, Jr CRJ. Non-Stationarity in the "Resting Brain's" Modular Architecture. PLOS ONE. 2012 Jun; 7(6):e39731. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0039731, doi: 10.1371/journal.pone.0039731, publisher: Public Library of Science.

Kaiser RH, Whitfield-Gabrieli S, Dillon DG, Goer F, Beltzer M, Minkel J, Smoski M, Dichter G, Pizzagalli DA. Dynamic Resting-State Functional Connectivity in Major Depression. Neuropsychopharmacology. 2016 Jun; 41(7):1822–1830. https://www.nature.com/articles/npp2015352, doi: 10.1038/npp.2015.352, bandiera_abtest: a Cg_type: Nature Research Journals Number: 7 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Depression Subject_term_id: depression.

Keilholz S, Caballero-Gaudes C, Bandettini P, Deco G, Calhoun V. Time-Resolved Resting-State Functional Magnetic Resonance Imaging Analysis: Current Status, Challenges, and New Directions. Brain Connectivity. 2017 Oct; 7(8):465–481. doi: 10.1089/brain.2017.0543.

Kottaram A, Johnston LA, Cocchi L, Ganella EP, Everall I, Pantelis C, Kotagiri R, Zalesky A. Brain network dynamics in schizophrenia: Reduced dynamism of the default mode network. Human Brain Mapping. 2019; 40(7):2212–2228. https://onlinelibrary.wiley.com/doi/abs/10.1002/hbm.24519, doi: 10.1002/hbm.24519, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/hbm.24519.

de Lacy N, Doherty D, King BH, Rachakonda S, Calhoun VD. Disruption to control network function correlates with altered dynamic connectivity in the wider autism spectrum. NeuroImage: Clinical. 2017 Jan; 15:513–524. https://www.sciencedirect.com/science/article/pii/S2213158217301298, doi: 10.1016/j.nicl.2017.05.024.

de Lacy N, Calhoun VD. Dynamic connectivity and the effects of maturation in youth with attention deficit hyperactivity disorder. Network Neuroscience. 2018 Dec; 3(1):195–216. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6372020/, doi: 10.1162/netn_a_00063.

Li M, Dahmani L, Wang D, Ren J, Stocklein S, Lin Y, Luan G, Zhang Z, Lu G, Galiè F, Han Y, Pascual-Leone A, Wang M, Fox MD, Liu H. Co-activation patterns across multiple tasks reveal robust anti-correlated functional networks. NeuroImage. 2021 Feb; 227:117680. https://www.sciencedirect.com/science/article/pii/S1053811920311654, doi: 10.1016/j.neuroimage.2020.117680.

Liu X, Chang C, Duyn JH. Decomposition of Spontaneous Brain Activity into Distinct fMRI Co-activation Patterns. Frontiers in Systems Neuroscience. 2013; 0. https://www.frontiersin.org/articles/10.3389/fnsys.2013.00101/full, doi: 10.3389/fnsys.2013.00101, publisher: Frontiers.

Liu X, Duyn JH. Time-varying functional network information extracted from brief instances of spontaneous brain activity. Proceedings of the National Academy of Sciences. 2013 Mar; 110(11):4392–4397. http://www.pnas.org/lookup/doi/10.1073/pnas.1216856110, doi: 10.1073/pnas.1216856110.

Lurie DJ, Kessler D, Bassett DS, Betzel RF, Breakspear M, Kheilholz S, Kucyi A, Liégeois R, Lindquist MA, McIntosh AR, Poldrack RA, Shine JM, Thompson WH, Bielczyk NZ, Douw L, Kraft D, Miller RL, Muthuraman M, Pasquini L, Razi A, et al. Questions and controversies in the study of time-varying functional connectivity in resting fMRI. Network Neuroscience. 2020 Jan; 4(1):30–69. https://direct.mit.edu/netn/article/4/1/30-69/95807, doi: 10.1162/netn_a_00116.

Maltbie E, Yousefi B, Zhang X, Kashyap A, Keilholz S. Comparison of Resting-State Functional MRI Methods for Characterizing Brain Dynamics. Frontiers in Neural Circuits. 2022 Apr; https://www.proquest.com/docview/2646830978/abstract/204382497DC347F4PQ/1, doi: http://dx.doi.org/10.3389/fncir.2022.681544, place: Lausanne, Switzerland Publisher: Frontiers Research Foundation Section: ORIGINAL RESEARCH article.

Nikolaou F, Orphanidou C, Papakyriakou P, Murphy K, Wise RG, Mitsis GD. Spontaneous physiological variability modulates dynamic functional connectivity in resting-state functional magnetic resonance imaging. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences. 2016 May; 374(2067):20150183. https://royalsocietypublishing.org/doi/10.1098/rsta.2015.0183, doi: 10.1098/rsta.2015.0183, publisher: Royal Society.

Ou J, Xie L, Jin C, Li X, Zhu D, Jiang R, Chen Y, Zhang J, Li L, Liu T. Characterizing and Differentiating Brain State Dynamics via Hidden Markov Models. Brain Topography. 2015 Sep; 28(5):666–679. https://doi.org/10.1007/s10548-014-0406-2, doi: 10.1007/s10548-014-0406-2.

Preti MG, Bolton TA, Van De Ville D. The dynamic functional connectome: State-of-the-art and perspectives. NeuroImage. 2017 Oct; 160:41–54. https://www.sciencedirect.com/science/article/pii/S1053811916307881, doi: 10.1016/j.neuroimage.2016.12.061.

Rashid B, Arbabshirani MR, Damaraju E, Cetin MS, Miller R, Pearlson GD, Calhoun VD. Classification of schizophrenia and bipolar patients using static and dynamic resting-state fMRI brain connectivity. NeuroImage. 2016 Jul; 134:645–657. https://www.sciencedirect.com/science/article/pii/S105381191630091X, doi: 10.1016/j.neuroimage.2016.04.051.

Reggente N, Moody TD, Morfini F, Sheen C, Rissman J, O'Neill J, Feusner JD. Multivariate resting-state functional connectivity predicts response to cognitive behavioral therapy in obsessive–compulsive disorder. Proceedings of the National Academy of Sciences. 2018 Feb; 115(9):2222–2227. https://www.pnas.org/content/115/9/2222, doi: 10.1073/pnas.1716686115, publisher: National Academy of Sciences Section: Biological Sciences.

Sakoglu U, Pearlson GD, Kiehl KA, Wang YM, Michael AM, Calhoun VD. A method for evaluating dynamic functional network connectivity and task-modulation: application to schizophrenia. Magnetic Resonance Materials in Physics, Biology and Medicine. 2010 Dec; 23(5):351–366. https://doi.org/10.1007/s10334-010-0197-8, doi: 10.1007/s10334-010-0197-8.

Savva AD, Matsopoulos GK, Mitsis GD. A Wavelet-Based Approach for Estimating Time-Varying Connectivity in Resting-State fMRI. Brain Connectivity. 2021 Jun; p. brain.2021.0015. https://www.liebertpub.com/doi/10.1089/brain.2021.0015, doi: 10.1089/brain.2021.0015.

Savva AD, Mitsis GD, Matsopoulos GK. Assessment of dynamic functional connectivity in resting-state fMRI using the sliding window technique. Brain and Behavior. 2019 Apr; 9(4):e01255. https://onlinelibrary.wiley.com/doi/abs/10.1002/brb3.1255, doi: 10.1002/brb3.1255.

Schilling KG, Rheault F, Petit L, Hansen CB, Nath V, Yeh FC, Girard G, Barakovic M, Rafael-Patino J, Yu T, Fischi-Gomez E, Pizzolato M, Ocampo-Pineda M, Schiavi S, Canales-Rodríguez EJ, Daducci A, Granziera C, Innocenti G, Thiran JP, Mancini L, et al. Tractography dissection variability: What happens when 42 groups dissect 14 white matter bundles on the same dataset? NeuroImage. 2021 Nov; 243:118502. doi: 10.1016/j.neuroimage.2021.118502.

Smith SM, Beckmann CF, Andersson J, Auerbach EJ, Bijsterbosch J, Douaud G, Duff E, Feinberg DA, Griffanti L, Harms MP, Kelly M, Laumann T, Miller KL, Moeller S, Petersen S, Power J, Salimi-Khorshidi G, Snyder AZ, Vu AT, Woolrich MW, et al. Resting-state fMRI in the Human Connectome Project. NeuroImage. 2013 Oct; 80:144–168. https://www.sciencedirect.com/science/article/pii/S1053811913005338, doi: 10.1016/j.neuroimage.2013.05.039.

Stevner ABA, Vidaurre D, Cabral J, Rapuano K, Nielsen SFV, Tagliazucchi E, Laufs H, Vuust P, Deco G, Woolrich MW, Van Someren E, Kringelbach ML. Discovery of key whole-brain transitions and dynamics during human wakefulness and non-REM sleep. Nature Communications. 2019 Mar; 10(1):1035. https://www.nature.com/articles/s41467-019-08934-3, doi: 10.1038/s41467-019-08934-3, number: 1 Publisher: Nature Publishing Group.

Suk HI, Wee CY, Lee SW, Shen D. State-space model with deep learning for functional dynamics estimation in resting-state fMRI. NeuroImage. 2016 Apr; 129:292–307. doi: 10.1016/j.neuroimage.2016.01.005.

Uddin LQ. Cognitive and behavioural flexibility: neural mechanisms and clinical considerations. Nature Reviews Neuroscience. 2021 Mar; 22(3):167–179. https://www.nature.com/articles/s41583-021-00428-w, doi: 10.1038/s41583-021-00428-w, number: 3 Publisher: Nature Publishing Group.

Van Essen DC, Smith SM, Barch DM, Behrens TEJ, Yacoub E, Ugurbil K. The WU-Minn Human Connectome Project: An overview. NeuroImage. 2013 Oct; 80:62–79. https://www.sciencedirect.com/science/article/pii/S1053811913005351, doi: 10.1016/j.neuroimage.2013.05.041.

Vidaurre D, Abeysuriya R, Becker R, Quinn AJ, Alfaro-Almagro F, Smith SM, Woolrich MW. Discovering dynamic brain networks from big data in rest and task. NeuroImage. 2018 Oct; 180:646–656. https://www.sciencedirect.com/science/article/pii/S1053811917305487, doi: 10.1016/j.neuroimage.2017.06.077.

Vidaurre D, Smith SM, Woolrich MW. Brain network dynamics are hierarchically organized in time. Proceedings of the National Academy of Sciences. 2017 Nov; 114(48):12827–12832. https://www.pnas.org/content/114/48/12827, doi: 10.1073/pnas.1705120114, publisher: National Academy of Sciences Section: Biological Sciences.

Wang Z, Yang Q, Chen LM. Abnormal dynamics of cortical resting state functional connectivity in chronic headache patients. Magnetic Resonance Imaging. 2017 Feb; 36:56–67. https://www.sciencedirect.com/science/article/pii/S0730725X1630162X, doi: 10.1016/j.mri.2016.10.015.

Xie H, Zheng CY, Handwerker DA, Bandettini PA, Calhoun VD, Mitra S, Gonzalez-Castillo J. Efficacy of different dynamic functional connectivity methods to capture cognitively relevant information. NeuroImage. 2019 Mar; 188:502–514. https://www.sciencedirect.com/science/article/pii/S1053811918321815, doi: 10.1016/j.neuroimage.2018.12.037.

Xifra-Porxas A, Kassinopoulos M, Mitsis GD. Physiological and motion signatures in static and time-varying functional connectivity and their subject identifiability. eLife. 2021 Aug; 10:e62324. https://doi.org/10.7554/eLife.62324, doi: 10.7554/eLife.62324, publisher: eLife Sciences Publications, Ltd.

Yaesoubi M, Adalı T, Calhoun VD. A window-less approach for capturing time-varying connectivity in fMRI data reveals the presence of states with variable rates of change. Human Brain Mapping. 2018; 39(4):1626–1636. https://onlinelibrary.wiley.com/doi/abs/10.1002/hbm.23939, doi: https://doi.org/10.1002/hbm.23939, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/hbm.23939.

Yaesoubi M, Miller RL, Bustillo J, Lim KO, Vaidya J, Calhoun VD. A joint time-frequency analysis of resting-state functional connectivity reveals novel patterns of connectivity shared between or unique to schizophrenia patients and healthy controls. NeuroImage: Clinical. 2017 Jan; 15:761–768. https://www.sciencedirect.com/science/article/pii/S2213158217301547, doi: 10.1016/j.nicl.2017.06.023.

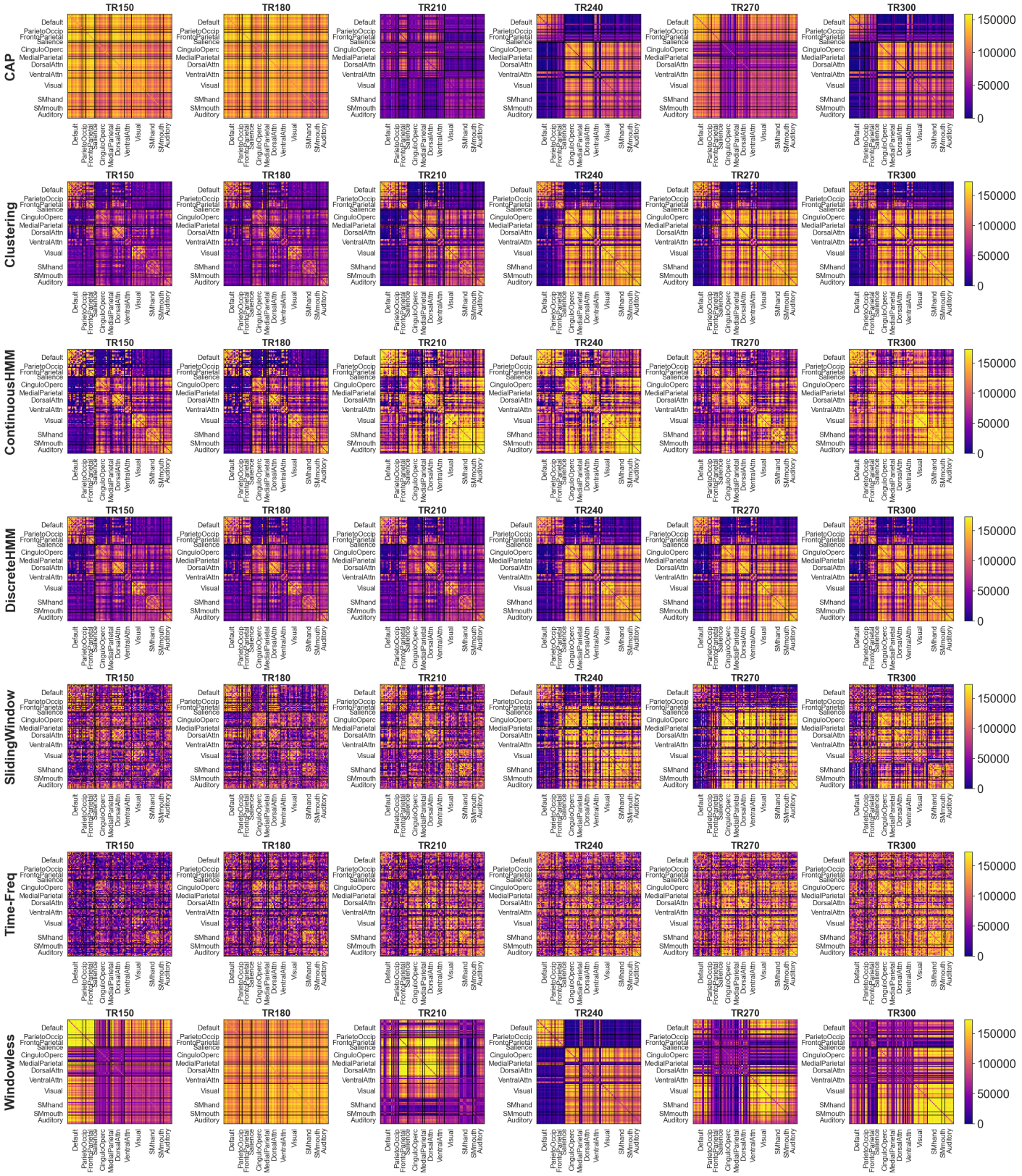
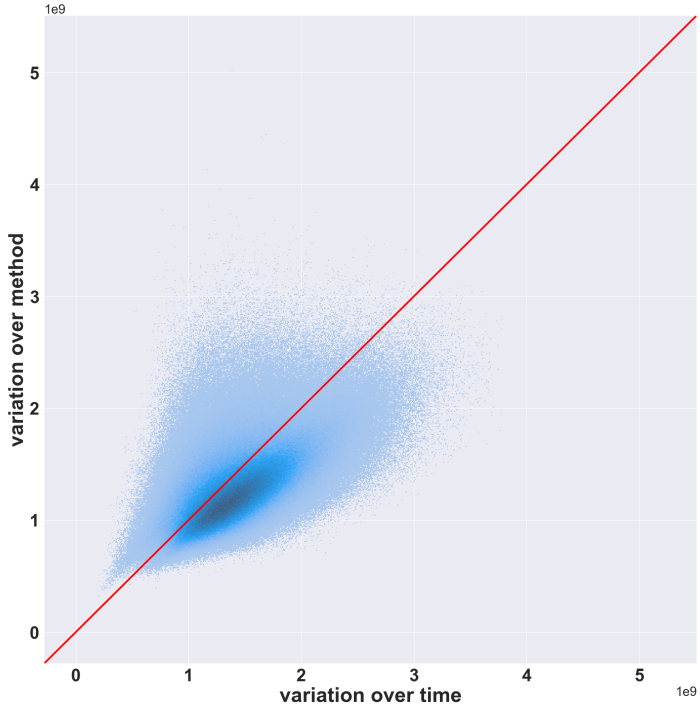
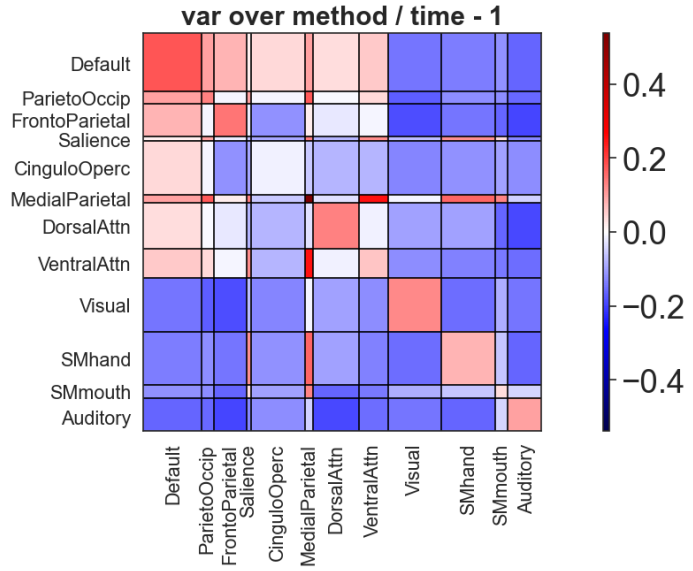


Figure 2. A sample segment of dFC matrices of one subject obtained by the implemented methods: each row shows the dFC matrix obtained by each method and each column shows the corresponding time point, or TR. Each FC matrix has a size of $ROI \times ROI$ and ROIs are divided into 12 RSNs. To account for the different value distributions of each method and better comparability, the dFC matrices were normalized by ranking the values, similar to the approach used when calculating Spearman correlation.

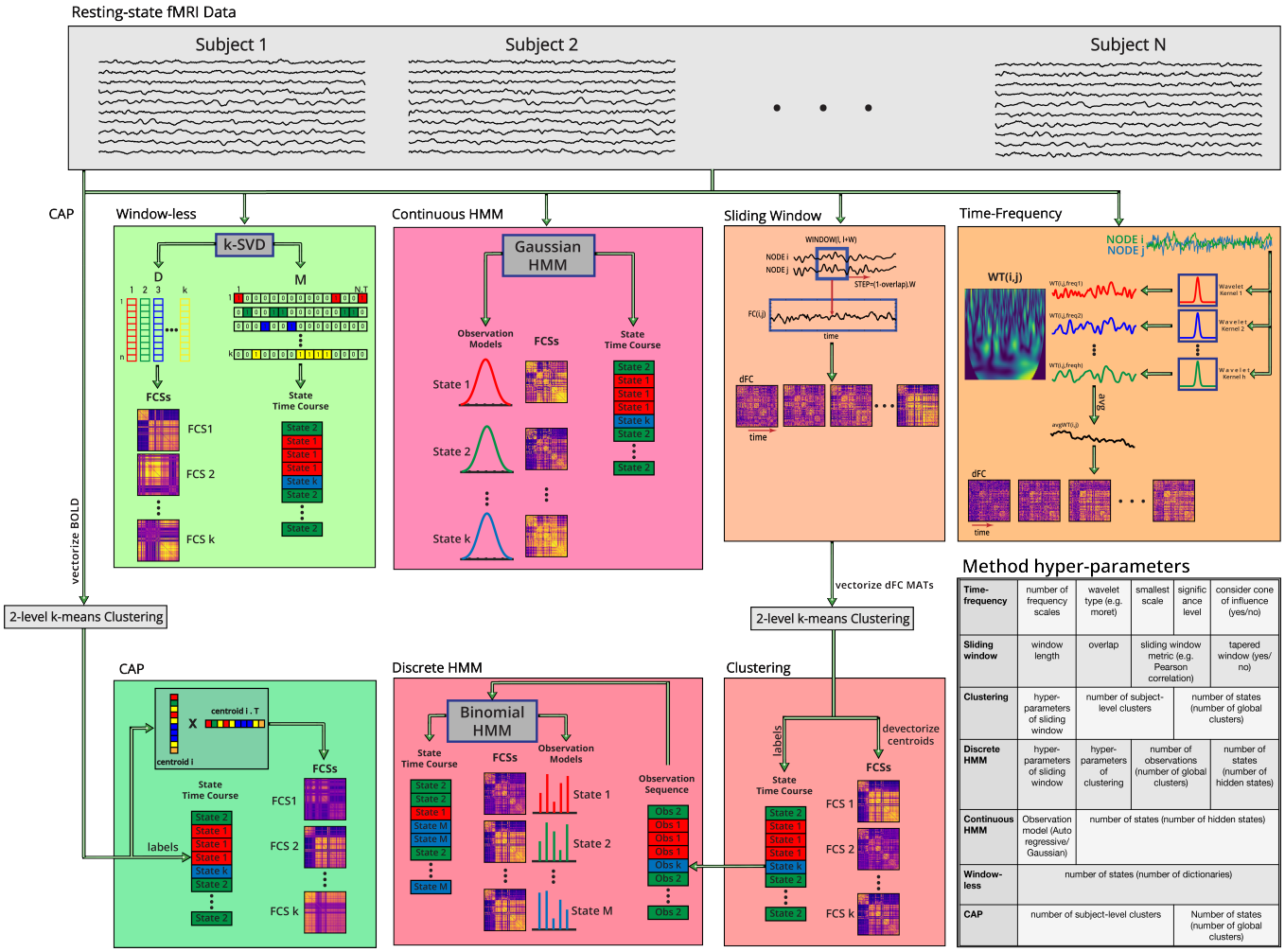


(a) variance over method vs time across functional connections and subjects



(b) variance over method vs time across networks averaged over subjects

Figure 3. a) A scatter plot of the variance over methods versus the variance over time for each functional connection and each subject ($subj \times (ROI \times (ROI - 1)/2)$ points in total) indicates that the variance over dFC methods is comparable to the variance over time, with an average ratio of $var_{method}/var_{time} = 0.95$ ($SD_{method}/SD_{time} = 0.97$). b) Ratios of variance over method divided by variance over time across functional connections averaged over subjects. The ratios were subtracted by one ($ratio - 1$) so that zero values (white) correspond to equal variance over method and time. The current figure highlights Resting-State Networks (RSNs) pairs with functional connections that are more variable over method than time (red) and those more variable over time than method (blue). The variance values corresponding to functional connections belonging to each pair of RSNs were averaged to yield a single variance value prior to computing the ratios. The plot shows that the functional connections between RSNs such as the FrontoParietal and Ventral Attention networks exhibited equal variance values over time and method, while the functional connections between RSNs such as the Default Mode and Parieto-occipital, and Default Mode and FrontoParietal Networks, and most intra-network connections exhibited higher variation over method than over time.



Time-frequency	number of frequency scales	wavelet type (e.g. morlet)	smallest scale	significance level	consider cone of influence (yes/no)
Sliding window	window length	overlap	sliding window metric (e.g. Pearson correlation)	tapered window (yes/no)	
Clustering	hyper-parameters of sliding window	number of subject-level clusters	number of states (number of global clusters)		
Discrete HMM	hyper-parameters of sliding window	hyper-parameters of clustering	number of observations (number of global clusters)	number of states (number of hidden states)	
Continuous HMM	Observation model (Auto regressive/ Gaussian)	number of states (number of hidden states)			
Window-less	number of states (number of dictionaries)				
CAP	number of subject-level clusters	Number of states (number of global clusters)			

Figure 4. pipeline of the seven dFC assessment methods (time-series of all subjects are concatenated in time dimension for state-based methods prior to finding FC states): **Time-frequency** (***Chang and Glover, 2010***): Wavelet Transform Coherence (WTC) between each pair of regions is computed after applying the wavelet transform on region time courses. Finally, dFC assessment is performed by averaging WTC over all frequency scales. **Sliding window** (***Sakoglu et al., 2010***): dFC is assessed for each time window using a sliding window which moves over data time points and typically uses Pearson correlation as the metric to quantify FC for each pair of regions. **Clustering** (***Allen et al., 2014***): applies 2-level k-means clustering on the sequence of measured dFC matrices from sliding window method to find the clusters that correspond to different FC states. Cluster centroids are considered as the state FC matrices, and the sequence of clustering labels is considered as the state time course. **Discrete Hidden Markov Model (HMM)** (***Ou et al., 2015***): the state FC matrices and the state time course inferred by clustering are used as the observation sequence of a binomial HMM. The model then finds the hidden states and infers the corresponding state FC matrices and the state time course. **Continuous HMM** (***Vidaurre et al., 2017***): the BOLD data is fed to a Gaussian HMM. The model finds the hidden states and their continuous observation models. It then infers the state FC matrices and the state time course. **Window-less** (***Yaesoubi et al., 2018***): a sparse dictionary learning is applied on the BOLD data using a k-SVD model to infer the dictionary matrix (D) and the mixing matrix (M). Next, state FC matrices and the state time course are obtained from D and M respectively. **Co-Activation Pattern (CAP)** (***Liu and Duyn, 2013***): clustering is performed on vectors of BOLD activity at each time point. Next, outer product of each centroid vector and its transpose is calculated and forms the state FC matrix corresponding to that cluster. The sequence of clustering labels of time points is considered as the state time course. Table A shows a list of all implemented dFC assessment methods and their important hyperparameters. Note that the state FC matrices in this figure are denoted using the term "FCS."