Received: 5 May 2021 | Revised: 19 November 2021 | Accepted: 24 November 2021

DOI: 10.1002/hbm.25736

RESEARCH ARTICLE

WILEY

Test–retest reliability of dynamic functional connectivity in naturalistic paradigm functional magnetic resonance imaging

Xin Zhang[1] | Jiayue Liu[2] | Yang Yang[3] | Shijie Zhao[2] | Lei Guo[2] | Junwei Han[2] | Xintao Hu[2]

[1]Institute of Medical Research, Northwestern Polytechnical University, Xi'an, Shaanxi, China

[2]School of Automation, Northwestern Polytechnical University, Xi'an, Shaanxi, China

[3]School of Life Sciences, Northwestern Polytechnical University, Xi'an, Shaanxi, China

Correspondence
Xintao Hu, School of Automation, Northwestern Polytechnical University, Xi'an, Shaanxi 710072, China.
Email: xhu@nwpu.edu.cn

Funding information
National Key R&D Program of China, Grant/Award Number: 2020AAA0105701; National Natural Science Foundation of China, Grant/Award Numbers: 61603399, 61836006, 61936007, 61806167, 62076205, U1801265; National Natural Science Foundation of Shaanxi Province, Grant/Award Number: 2021JQ-119; Fundamental Research Funds for the Central Universities, Grant/Award Numbers: 31020190QD002, 3102019YX01005

Abstract

Dynamic functional connectivity (dFC) has been increasingly used to characterize the brain transient temporal functional patterns and their alterations in diseased brains. Meanwhile, naturalistic neuroimaging paradigms have been an emerging approach for cognitive neuroscience with high ecological validity. However, the test–retest reliability of dFC in naturalistic paradigm neuroimaging is largely unknown. To address this issue, we examined the test–retest reliability of dFC in functional magnetic resonance imaging (fMRI) under natural viewing condition. The intraclass correlation coefficients (ICC) of four dFC statistics including standard deviation (Std), coefficient of variation (COV), amplitude of low frequency fluctuation (ALFF), and excursion (Excursion) were used to measure the test–retest reliability. The test–retest reliability of dFC in naturalistic viewing condition was then compared with that under resting state. Our experimental results showed that: (a) Global test–retest reliability of dFC was much lower than that of static functional connectivity (sFC) in both resting-state and naturalistic viewing conditions; (b) Both global and local (including visual, limbic and default mode networks) test–retest reliability of dFC could be significantly improved in naturalistic viewing condition compared to that in resting state; (c) There existed strong negative correlation between sFC and dFC, weak negative correlation between dFC and dFC-ICC (i.e., ICC of dFC), as well as weak positive correlation between dFC-ICC and sFC-ICC (i.e., ICC of sFC). The present study provides novel evidence for the promotion of naturalistic paradigm fMRI in functional brain network studies.

KEYWORDS
dynamic functional connectivity (dFC), functional magnetic resonance imaging (fMRI), natural viewing, resting state, test–retest reliability

Xin Zhang and Jiayue Liu contributed equally to this work.

1 | INTRODUCTION

Dynamic functional connectivity (dFC) has become an important tool to probe the transient functional interactions among brain regions and their alterations in disordered brains (Allen et al., 2014; Damaraju et al., 2014; Fiorenzato et al., 2019; Hutchison et al., 2013; Liu et al., 2017). Compared to static functional connectivity (sFC), it enriches the content of functional brain networks and provides a novel way to approach the detailed mechanisms of information processing in the brain. A number of studies have reported that dFC can provide valuable information that is inaccessible with sFC about functional network organizations (Diez-Cirarda et al., 2018; Leonardi & Van De Ville, 2015; X. Zhang et al., 2013). For example, a series of representative functional states have been identified based on dFC and found to be associated with complex brain functions and dysfunctions of neuropsychiatric diseases (Allen et al., 2014; Damaraju et al., 2014; Espinoza et al., 2019; Li et al., 2014; X. Zhang et al., 2013).

Test–retest reliability generally measures the test consistency, that is, the reliability of a test measured over time (J. Wang, Han, et al., 2017; Zuo & Xing, 2014). The test–retest reliability of functional connectivity (FC) in functional magnetic resonance imaging (fMRI) is crucial for its applications in clinic and has been widely discussed in previous studies (Choe et al., 2017; Guo et al., 2012; Noble, Scheinost, & Constable, 2019; Zuo & Xing, 2014). The majority of existing FC studies in basic and translational neuroscience focuses on resting-state fMRI (rs-fMRI) to search for pathophysiological correlates of neurological and psychotic disorders (Tailby, Masterton, Huang, Jackson, & Abbott, 2015; X. Wang et al., 2019; Wehrle et al., 2018; X. Zhang et al., 2020). However, the unconstrained resting state may increase individual variations and make it difficult to separate signals of interest from unwanted behavioral confounds (Sonkusare, Breakspear, & Guo, 2019; Tong, Hocke, & Frederick, 2019). Recently, naturalistic paradigms including dynamic videos, speech and music designed to study real-life sensory experience have been proposed and gained increasing interest (Kauttonen, Hlushchuk, Jaaskelainen, & Tikka, 2018; Kuo et al., 2019; Marshall et al., 2019; Ren, Lv, Guo, Fang, & Guo, 2017; Sonkusare et al., 2019). It has been demonstrated that naturalistic stimuli can achieve improved ecological validity and may be more suitable for challenging populations such as children or cognitively impaired patients (Kim, Wang, Wedell, & Shinkareva, 2016; Kuo et al., 2019; Mandelkow, de Zwart, & Duyn, 2016). To further validate its potential in functional brain network studies, the elucidation of test–retest reliability of FC in naturalistic paradigm fMRI (nfMRI) is of crucial importance.

In previous studies, the test–retest reliability of sFC as well as the topological properties of functional brain networks in nfMRI have been investigated (J. Wang, Ren, et al., 2017). Specifically, the natural movie-viewing paradigm can significantly improve the test–retest reliability of sFC compared to resting-state. This finding encouraged the application of naturalistic paradigms in detecting longitudinal changes of disease progression for neuropsychiatric brains. However, it is still unknown whether the conclusions are still applicable for dFC (Preti, Bolton, & Van De Ville, 2017; Viviano, Raz, Yuan, & Damoiseaux, 2017; C. Zhang, Baum, Adduru, Biswal, & Michael, 2018).

To address this issue, the test–retest reliability of dFC in nfMRI was investigated in the present study. A sliding-window scheme was applied to derive dFC (Li et al., 2014; X. Zhang et al., 2013). Then intraclass correlation coefficient (ICC) analysis of dFC statistics including standard deviation (Std), coefficient of variation (COV), amplitude of low frequency fluctuation (ALFF), and excursion (Excursion) as recommended in previous studies was performed to measure the test–retest reliability of dFC (dFC-ICC) (Choe et al., 2017; Guo et al., 2012; J. Wang, Han, et al., 2017; J. Wang, Ren, et al., 2017; Zuo & Xing, 2014). Based on dFC-ICC, the test–retest reliability of dFC was then evaluated for region of interest (ROI) pairs (ROI-level), sub-networks (network-level) and the whole brain (global-level). It was found that nfMRI significantly increased global-level dFC-ICC compared to rs-fMRI. Several sub-networks including visual network, limbic network and default mode network showed significantly higher intranetwork dFC-ICC in natural viewing condition. In addition, the global dFC-ICC was lower than sFC-ICC. We further discussed the relationships among sFC, dFC, sFC-ICC and dFC-ICC. In general, the present study demonstrated that nfMRI can improve test–retest reliability of dFC compared to rs-fMRI and provides complementary evidence to the potential of nfMRI in exploring functional brain networks.

2 | MATERIALS AND METHODS

2.1 | Participants and experimental paradigms

Twenty healthy right-handed subjects (9 males, 11 females; age from 21 to 31 years) were recruited from the University of Queensland. Written informed consent was obtained for each of them. The study was approved by the human ethics research committee of the University of Queensland and conducted according to National Health and Medical Research Council guidelines. Two scanning sessions (Session I and Session II) with a time interval of three months were acquired for each participant. During each session, two conditions of fMRI data were acquired, that is, 8-min resting-state and 20-min natural movie-viewing stimulus. The resting-state scan was performed prior to the movie stimulus scan. During the movie-viewing scan, the participant was asked to freely view a short movie named "The Butterfly Circus" (Nguyen et al., 2017; J. Wang, Ren, et al., 2017). All participants declared that they had not previously seen this film and were asked not to watch it outside the scan sessions before the experiment. Three subjects were excluded: one with technical problem during data recording and the other two without the second scan. Then, only 17 subjects (10 females and 7 males) were included in the study. More details can be found in previous reports (Nguyen et al., 2017; J. Wang, Ren, et al., 2017).

2.2 | Image acquisition and preprocessing

Structural and functional MRI images were acquired on a 3T Siemens Trio MRI scanner using a 12-channel heal coil (Siemens Medical System, Germany). The scanning parameters were as follows: (a) high-resolution T1-weighted MPRAGE images with echo time (TE) = 2.89 ms, repetition time (TR) = 4,000 ms, flip angle (FA) = 9°, filed of view (FOV) = 240 × 256 mm^2, and voxel size = 1 × 1 × 1 mm^3; (b) fMRI images using a gradient-echo echo-planar imaging (GE-EPI) sequence with TE = 30 ms, TR = 2,200 ms, FA = 79°, FOV = 192 × 192 mm^2, matrix = 64 × 64, 44 axial slices, voxel size = 3 × 3 × 3 mm^3. Functional images were preprocessed using Statistical Parametric Mapping toolbox (SPM12) and the Data Processing Assistant for Resting-State fMRI software (DPARSF) implemented in Matlab (Mathworks, USA; Chao-Gan & Yu-Feng, 2010). The main steps included: removing the first five volumes, slice timing, motion correction, normalization into the Montreal Neurological Institute (MNI) space, spatial smoothing with a 6-mm full width half maximum (FWHM) Gaussian kernel, regressing out nuisance covariates, band-pass filtering (0.01–0.1 Hz) and detrending. More details about preprocessing can refer to previous reports (Nguyen et al., 2017; J. Wang, Ren, et al., 2017).

2.3 | sFC and dFC calculation

The whole brain was parcellated into 200 region of interests (ROIs) according to the Craddock 200 atlas (Craddock, James, Holtzheimer 3rd, Hu, & Mayberg, 2012). Specially, these 200 ROIs can be grouped into 8 sub-networks according to Yeo 7 network template (see Figure S1): visual network (VN), somatomotor network (SMN), dorsal attention network (DAN), ventral attention network (VAN), limbic network (LN), frontoparietal network (FPN), default mode network (DMN), and other areas (OA) those are not included in the former networks (Yeo et al., 2011). The OA regions mainly cover parts of the cerebellum, brain stem, thalamus, and caudate. For each Craddock ROI, its network-label was determined by one of the Yeo 7 functional networks that the majority of voxels within the ROI falls into. It was attributed to OA if the majority of the voxels within the ROI falls out of all the 7 brain networks. In this way, all of the 200 ROIs were grouped into 8 sub-networks.

2.3.1 | sFC

The fMRI time series of all voxels within each ROI were extracted and averaged as the time series of the corresponding ROI. Then, the Pearson correlation coefficient was calculated between each pair of ROIs based on their time series and a 200 × 200 whole-brain sFC matrix was generated. In this experiment, one resting-state sFC matrix and one natural viewing sFC matrix was separately obtained for each subject during each session.

2.3.2 | dFC

The sliding-window strategy was adopted to assess the dFC time series. Briefly, a rectangle time window with an interval of one TR was applied. Within each sliding window, the Pearson correlation coefficients between each pair of ROIs were calculated. Let the full duration of the fMRI time series be T; $T = TR \times N$ where N is the total number of time points. Let the length of the sliding time window is w; $w = TR \times n$ where n is the number of time points in the window. Then, a three-dimensional $200 \times 200 \times (N - n + 1)$ whole-brain FC matrix sequence, that is, dFC time series, was derived. It should be mentioned that, as suggested by Leonardi and Van De Ville, before dFC calculation a high-pass filtering was previously applied to the fMRI time series that removes frequency component below $1/w$, to avoid spurious fluctuations of FC time series when the window length was short (Leonardi & Van De Ville, 2015). Besides, with respect to the key parameter of the window length in dFC calculation, a sequence of different window sizes ranging from 10 TRs to 100 TRs in increments of 10 TRs were implemented and compared based on the dFC-ICC analysis results to determine the optimal window length for our data.

2.4 | Test–retest reliability metrics

The test–retest reliability of sFC was measured by calculating the ICC of sFC between two sessions for each pair of ROIs and a 200 × 200 whole-brain sFC-ICC matrix can be obtained. As the test–retest reliability of dFC is difficult to be achieved directly from dFC time series, a statistic that can quantify the dynamic fluctuation of the time series will be needed. In this study, four statistics including standard deviation (Std), coefficient of variation (COV), amplitude of low frequency fluctuation (ALFF), and excursion (Excursion) were used to derive the dFC-ICC metrics (C. Zhang et al., 2018). Similarly, one corresponding 200 × 200 whole-brain dFC-ICC matrix can be acquired for each statistic.

Std

For dFC time series, the Std statistic is calculated according to Equation (1).

$$d_{std} = \sqrt{\frac{1}{N-n}\sum_{t=1}^{N-n+1}(r_t - \bar{r})^2} \quad (1)$$

where r_t is the Pearson correlation coefficient during the time window t and $\bar{r}$ is the average of the Pearson correlation coefficients across all of the sliding time windows.

COV

The COV statistic measures the percentage of variation around the arithmetic mean of the FC time series. It is defined as the ratio

between the standard deviation (*Std*) and the mean of the Pearson correlation coefficients ($\bar{r}$).

$$d_{COV} = \frac{d_{std}}{\bar{r}} \tag{2}$$

ALFF

The ALFF statistic was originally proposed to measure the slow fluctuations of resting-state brains and found to be effective in seeking the functional abnormality in disorders (Zang et al., 2007). Specially, through fast Fourier transform, the rs-fMRI time series can be transformed into temporal frequencies and the amplitudes within the generally considered low frequency range were averaged to generate the ALFF metric. Here, as to the dFC time series, the effect of a finite window size is equivalent to applying a low-pass filter to the FC time series fluctuations with a cut-off frequency $1/w$ (Leonardi & Van De Ville, 2015; C. Zhang et al., 2018). Thus, the frequency components from 0 to $1/w$ were summarized to generate the ALFF statistic for dFC time series, as shown in Equation (3).

$$d_{\text{ALFF}} = \sqrt{\frac{1}{n_f}\sum_{i=1}^{n_f}\alpha_i} \tag{3}$$

where n_f is the number of the frequent components and α_i refers to the amplitude of the i-th frequency component.

Excursion

The excursion statistic was provided by Zalesky et al. in 2014, aiming to measure the extent of time-varying fluctuations in the time-resolved dFC (Zalesky, Fornito, Cocchi, Gollo, & Breakspear, 2014). First, a time point t was defined as a median crossing point if the signs of $r_t - median(r)$ and $r_{t-1} - median(r)$ are different, where r_t and r_{t-1} represent the Pearson correlation coefficients during the time window t and $t-1$, and *median(r)* is the median value of the coefficients across all time windows. All of the median crossing points were collected across time. A pair of consecutive crossing points (t_k, t_{k+1}) defines an excursion from the median value. The longer and larger the excursions are from the median, the greater the evidence of non-stationary behaviors. Therefore, the final statistic is designed by summating all excursions as a function of excursion length and height.

$$d_{\text{excursion}} = \sum_{k=1}^{J} l_k^{\alpha} h_k^{\beta} \tag{4}$$

where $l_k = t_{k+1} - t_k$ is the length of Excursion and $h_k = \max\{r_t - median(r), t_k < t < t_{k+1}\}$ indicates the height of the k-th Excursion. J is the total number of all median crossing points. The exponents of α and β are used to adjust the relative weighting between the importance of long versus large excursions; $\alpha = .8$ and $\beta = .3$ were used in this study and the influence of their settings were discussed in Section 4.3.

ICC analysis

Here, a one-way random ANOVA model with subject as random effect was established to calculate the ICC (J. Wang, Han, et al., 2017; C. Zhang et al., 2018; Zuo & Xing, 2014). Through the ANOVA model, the between-subject mean square (MS_b) and within-subject mean square (MS_w) can be derived. Then, the ICC values can be calculated as follows:

$$ICC = \frac{MS_b - MS_w}{MS_b + (d-1)MS_w} \tag{5}$$

where d is the number of observations per subject (i.e., $d = 2$ for our data). Based on sFC and dFC statistics of two scan sessions, the corresponding sFC-ICC and dFC-ICC (Std/COV/ALFF/Excursion) metrics were separately derived for each pair of ROIs (i.e., ROI-level). The ICCs were further investigated in sub-network level. Within- and between-network ICCs were calculated by averaging the ROI-level ICCs of all possible within- and between-network ROI pairs, respectively. The global-level ICC was calculated by averaging all the ROI-level ICCs.

Generally, once the ICC values are calculated, they can be categorized into some intervals to indicate different reliability performance. However, Termenon, Jaillard, Delon-Martin, and Achard (2016) suggested that it is more appropriate to assess the reliability by the statistical significance. Thus, the *F*-statistic based on the ANOVA model was used to assess statistical significance of ICC:

$$F = \frac{MS_b}{MS_w} \tag{6}$$

The *p*-value of the *F*-statistic was computed using the degrees of freedom $df1 = 16$ and $df2 = 17$ (for one-way ANOVA). In general, the larger the *F* value is, the smaller the *P*-value will be and the higher ICC (i.e., better reliability) will be achieved.

To assess the statistical differences of ICC between resting-state and natural viewing conditions, a paired nonparametric permutation test was conducted under the null hypothesis that the difference of ICC between two conditions is drawn from a distribution with zero mean. First, two surrogate conditions were created for Session I by concatenating random selected samples from resting-state and natural viewing fMRI data, and the same selection was applied on Session II. Then, the new ICCs and their differences were computed. This process was repeated 5,000 times to generate the null distribution of ICC differences and the true differences in ICC values were compared with this null distribution (one-tailed test, $p < .025$).

3 | RESULTS

After preprocessing, a total of 215 volumes were included in the rs-fMRI data while the natural viewing fMRI data contained 530 volumes. To reduce the influence of volume number on test–retest analysis, the first 215 volumes under natural viewing condition were

segmented for the FC and ICC analysis. The 200 × 200 sFC and dFC statistic matrices were separately derived for each subject in each session of rs-fMRI and nfMRI data. Based on them, one sFC-ICC matrix and four dFC-ICC matrices in size of 200 × 200 were obtained for each condition. After removing duplicates in the symmetric matrix, 19,900 elements covering all the ROI pairs were retained for each matrix.

3.1 | Determination of the window length in dFC evaluation

Window length is a crucial parameter in dFC evaluation based on sliding window (Hutchison et al., 2013). A too short or too long window will result in unreliable dFC results. In this study, the optimal window length is determined in conjunction with dFC-ICC analysis results. As mentioned in Section 2.3.2, a set of different window sizes ranging from 10 TRs to 100 TRs with a step of 10 TRs were applied to assess dFC time series. The dFC statistics of Std, COV, ALFF, and Excursion were derived and then the corresponding dFC-ICC values across each pair of ROIs, that is, ROI-level dFC-ICC, were calculated for rs-fMRI and nfMRI. Furthermore, the whole-brain average dFC-ICC, that is, global-level dFC-ICC, based on the four dFC statistics were computed by averaging the dFC-ICC values of all the 19,900 ROI pairs and visualized in Figure 1.

It can be observed in Figure 1 that the global-level dFC-ICC based on Std, ALFF and Excursion exhibited similar variations in both conditions as the time window length increased. For resting state, the ICC was maximized at the 30-TR-length window for Std. As to the ALFF and Excursion statistics, the second highest dFC-ICC was achieved at 30-TR-length window with a subtle difference from the peak (20-TR-length window). For natural viewing condition, the peak values of the ICC curves emerged at window size of 10 TRs and the second highest dFC-ICC values were achieved at window size of 30 TRs for the three statistics of Std, ALFF, and Excursion. However, the window length of 10 TRs is not an optimal option for dFC evaluation, as a too short window may introduce more fake dynamics that are not belong to the real dFC content and then further influence the accuracy of ICC analysis (Hutchison et al., 2013; Zalesky & Breakspear, 2015). Moreover, the peak dFC-ICC value was obtained at window size of 30 TRs for both conditions in COV statistic. Thus, the length of the window was set as 30 TRs in further experiments.

3.2 | The dFC statistic results for resting state and natural viewing

The dFC-ICC matrices and their differences between natural viewing and resting state using the four dFC statistics were compared, as shown in Figure S2. Similar spatial patterns of dFC-ICC were observed in all the four dFC statistics. The Excursion statistic was with the highest test–retest reliability measure for both resting-state and natural viewing conditions. This observation is in line with a previous study, which demonstrates that the Excursion statistic performs better in maximizing reliability as well as maintaining individual differences in resting-state fMRI (C. Zhang et al., 2018). Therefore, the Excursion statistic was selected as a single metric to derive ICC measures for the following dFC-ICC analysis.

For each session of rs-fMRI and nfMRI, the dFC time series and the corresponding dFC Excursion results were calculated using the 30-TR-length sliding window and those of one randomly selected subject were shown in Figure S3. The group average Excursion matrices of each session for both conditions were shown in Figure 2a. As mentioned in Section 2.3, the 200 Craddock ROIs were grouped into 8 sub-networks, that is, VN (1), SMN (2), DAN (3), VAN (4), LN (5), FPN (6), DMN (7), and OA (8) (Yeo et al., 2011). It can be seen that there exists similar excursion pattern in rs-fMRI and nfMRI, that is, low intranetwork and high internetwork Excursion, indicating less FC dynamics within each sub-network than between sub-networks. The distribution of Excursion across all of the ROI pairs for each session of

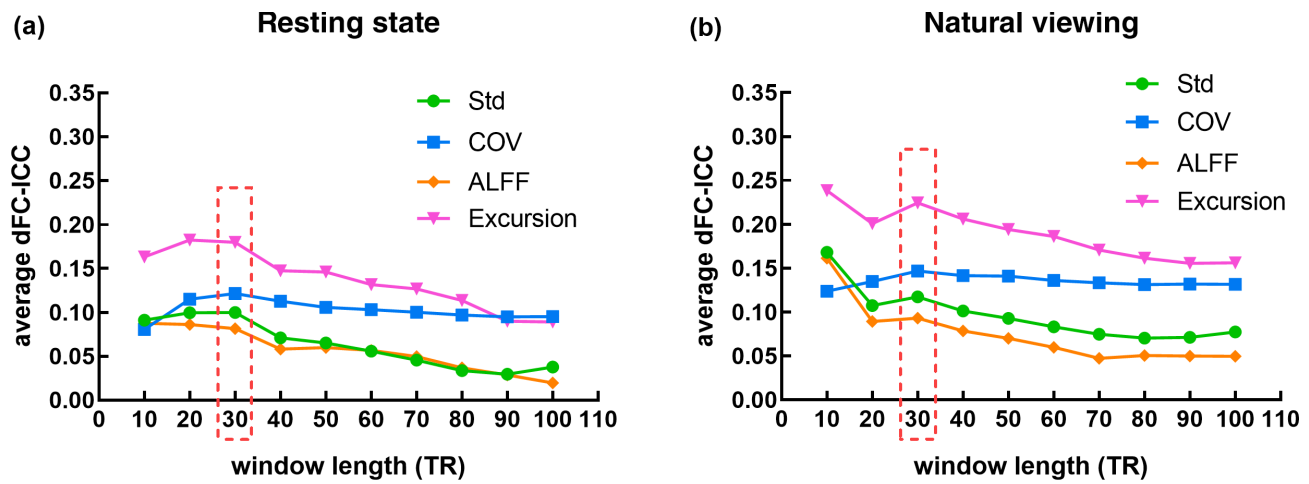

FIGURE 1 The variations of the global-level dFC-ICC along with different time windows (10 TRs, 20 TRs, 30 TRs, …, 100 TRs) based on the four dFC statistics of Std, COV, ALFF, and Excursion: (a) Resting state; (b) Natural viewing

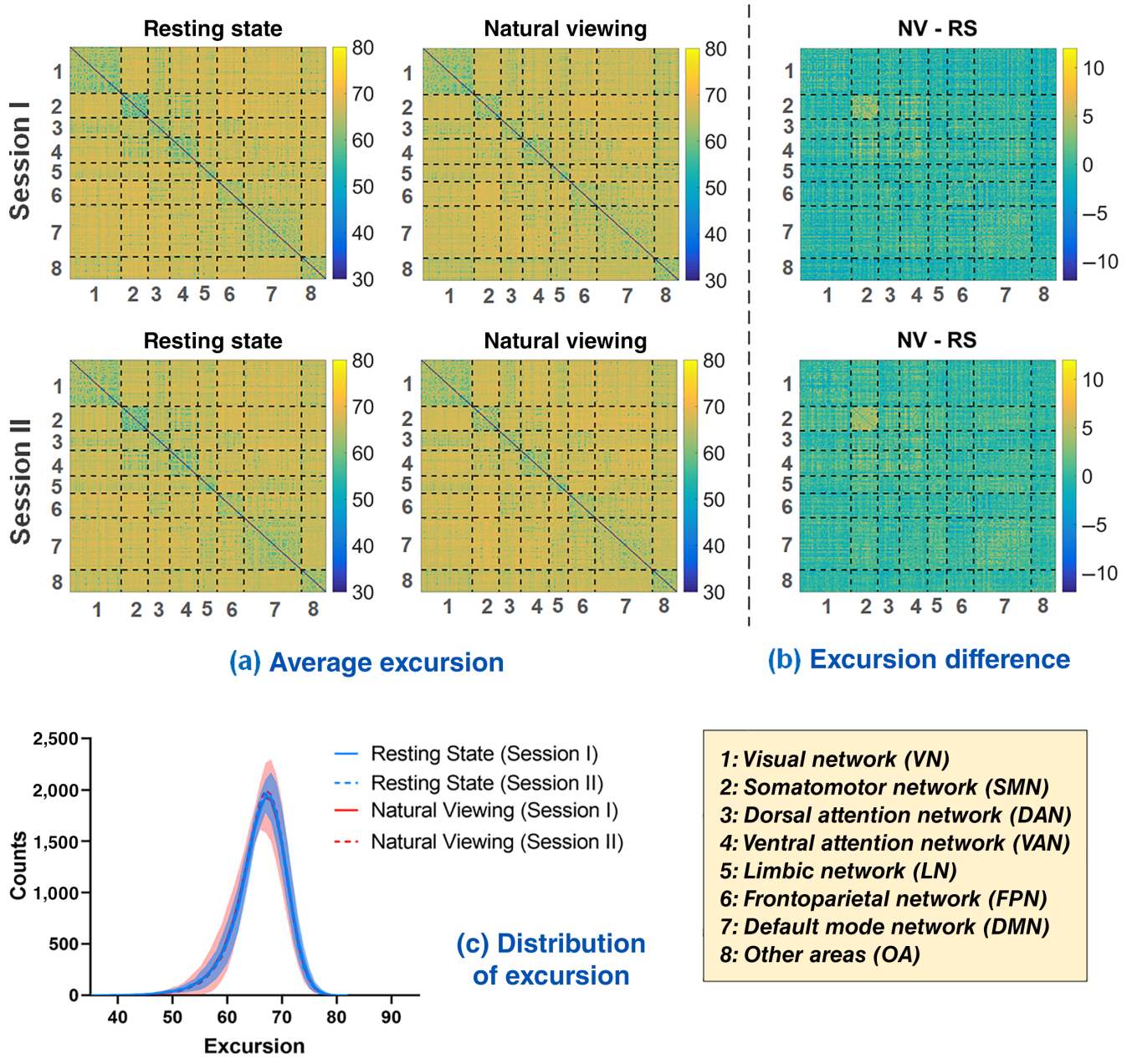

FIGURE 2 The Excursion of dFC. (a) The group average Excursion matrix of each session for resting state and natural viewing, respectively. (b) The group average Excursion differences between two conditions: natural viewing (NV) minus resting state (RS). (c) The distribution of Excursion across all the ROI pairs from each session of resting state (in cyan) and natural viewing (in magenta). Shade signifies standard error of the mean (SEM) across subjects. For visual clarity, only the SEM of Session I is displayed for each condition

resting state (in cyan) and natural viewing (in magenta) was drawn in Figure 2c. As to the group difference between resting state and natural viewing in Figure 2b, paired *t*-test was performed separately for each session (one-tailed test, $p < .025$). For Session I, the intranetwork Excursion tend to be higher in natural viewing than resting state in SMN ($p = .023$, uncorrected) and significant decreased excursion between OA and other subnetworks including VN ($p = .012$, uncorrected), SMN ($p = .016$, uncorrected), DAN ($p = .004$, uncorrected), LN ($p = .013$, uncorrected), and DMN ($p = .016$, uncorrected); while for Session II, significant higher Excursion were found within SMN ($p = .009$, uncorrected) and DMN ($p = .007$, uncorrected), and between LN and DMN ($p = .005$, uncorrected) for natural viewing. However, after multiple comparison correction across all the sub-network pairs using false discovery rate (FDR) method, none of the above differences were statistically significant (corrected $p > .025$). Furthermore, the whole-brain average Excursion for each subject was calculated and also compared between two conditions; no significant differences were found between them for both sessions (two-tailed *t*-test, $p = .635$).

3.3 | Statistical analysis of dFC-ICC for resting state and natural viewing

3.3.1 | Statistical results of significant dFC-ICC for resting state and natural viewing

The ICC results for each pair of ROIs based on the Excursion statistic, that is, ROI-level dFC-ICC, were calculated for resting-state and natural viewing conditions respectively (Figure 3a). The average dFC-ICC results of each sub-network (i.e., intranetwork ICC, corresponding to the diagonal boxes in the former two matrices) were also shown in the third column of Figure 3a. The distribution of dFC-ICC values across all of the ROI pairs for each condition was provided in Figure S4a. For each pair of ROIs, the *F*-statistic was computed during ICC analysis based on one-way ANOVA model. As shown in Figure 3b, a total of 5,380 out of 19,900 natural viewing ROI pairs had *p*-values lower than .05 and 539 remained significant after multiple comparison correction across all the ROI pairs using FDR correction method (corrected p <.05). As a contrast, for resting state, 4,372 out of 19,900 ROI pairs showed significant *p*-values (p <.05) and only 264 pairs were retained after FDR correction (corrected p <.05).

As to the intranetwork ICC, it can be observed from Figure 3b that the natural viewing condition had more significant ICC in VN, LN, DMN, and OA (in magenta dashed circle) while resting state exhibited more significant ICC occupations in SMN and DAN (in navy blue dashed circle). Their difference was small in VAN and FPN. Besides, natural viewing showed a relatively balanced percentage of significant ICC across each sub-network compared with resting state (the third column of Figure 3b).

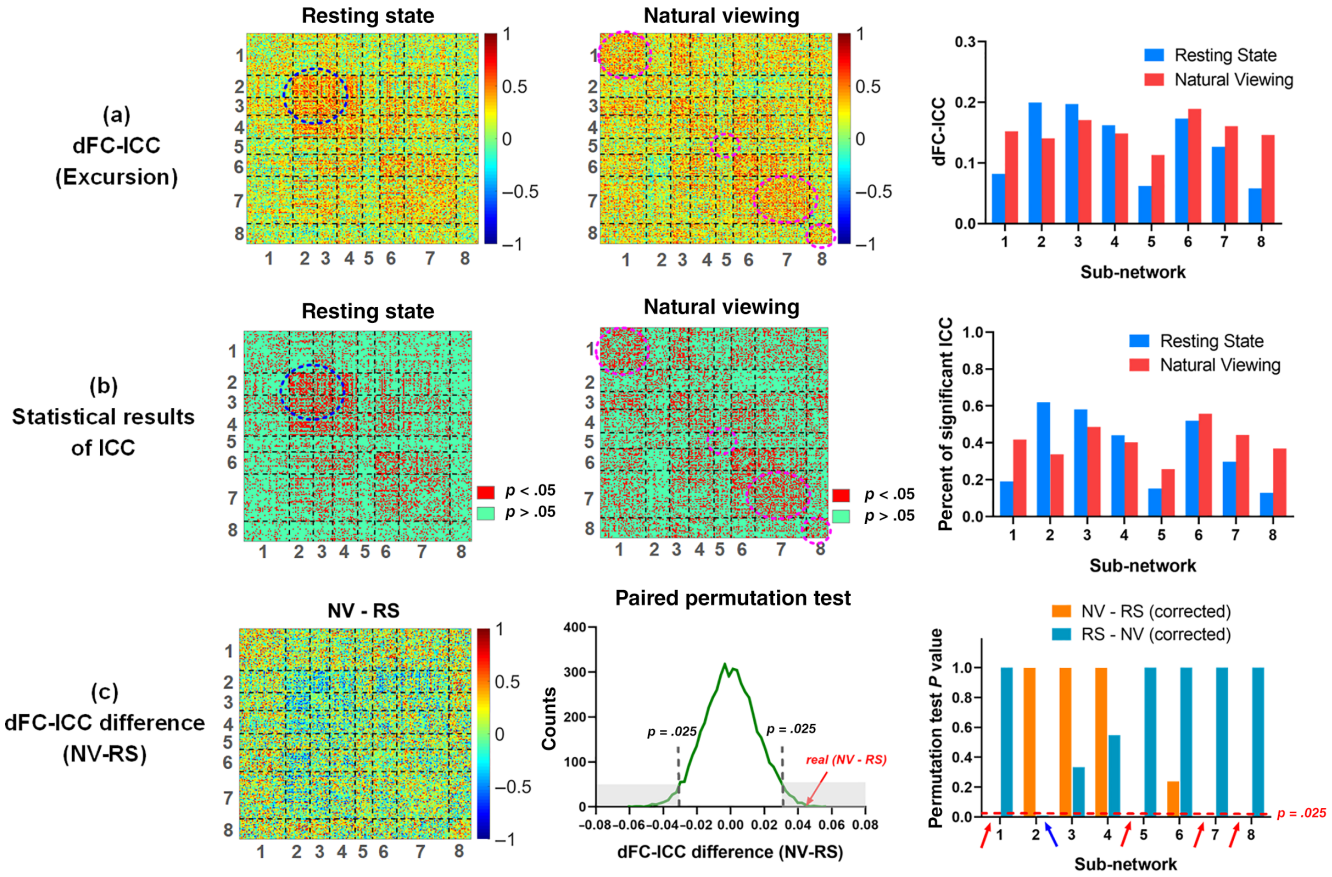

FIGURE 3 The dFC-ICC and statistical results. (a) The first and second columns were the Excursion-based dFC-ICC matrices of resting-state and natural viewing conditions. The last column showed the average dFC-ICC results of 8 sub-networks (i.e., intranetwork ICC) for both conditions. (b) The statistical results of dFC-ICC based on one-way ANOVA model. The first and second columns showed the ROI pairs with significant ICC results in red color (p <.05). The last column showed the percentages of significant ICC values in each sub-network for both conditions. (c) The first column was the dFC-ICC difference matrix, that is, natural viewing (NV) dFC-ICC minus resting state (RS) dFC-ICC. The second column showed the paired permutation test of the dFC-ICC difference (NV − RS) with gray dashed lines indicating 95% CIs. The last column showed the *p* values of each intranetwork average dFC-ICC difference between two conditions after FDR correction across the sub-networks

3.3.2 | Statistical comparison of dFC-ICC between resting state and natural viewing

As shown in Figure 1, the global-level dFC-ICC values of resting state and natural viewing were .180 and .225, respectively, that is, an increment of .045 for natural viewing compared with resting state. As for network-level, the natural viewing condition showed higher intranetwork dFC-ICC values in VN, LN, DMN, and OA, but decreased dFC-ICC values in SMN and DAN relative to resting state, as visualized in the last column of Figure 3a. The resting-state dFC-ICC differed sharply among sub-networks, ranging from the highest value of .200 (SMN) to the lowest value of .058 (OA). By contrast, the natural viewing condition had a relatively balanced ICC distribution across each sub-network, varying from the highest value of .189 (FPN) to the lowest value of .113 (LN). The dFC-ICC values of the other six sub-networks for natural viewing were all larger than .140. Furthermore, the dFC-ICC differences of each ROI pair (i.e., ROI-level) between natural viewing and resting state were also visualized in the first column of Figure 3c. The distributions of dFC-ICC for each condition and the dFC-ICC differences between them across all of the ROI pairs were visualized in Figure S4.

To validate the statistical significance of between-condition dFC-ICC difference, paired permutation tests were conducted along ROI-level, network-level and global-level. The global-level dFC-ICC of natural viewing was significantly larger than resting state ($p = .001$). When comparing the intranetwork dFC-ICC differences, natural viewing showed significant increased dFC-ICC in VN (p <.001, uncorrected), LN (p <.001, uncorrected), DMN (p <.001, uncorrected), and OA ($p = .010$, uncorrected) and decreased dFC-ICC in SMN ($p = .001$, uncorrected) than resting state. After FDR correction, all of the above differences remained significant (Figure 3c). As to the ROI-level, a total of 3,104 ROI pairs showed significantly increased dFC-ICC and 2,042 pairs showed significantly decreased dFC-ICC values for natural viewing than resting state (paired permutation test, uncorrected p <.025). After FDR correction, 1,187 pairs out of 3,104 and 641 pairs out of 2,042 remained significant (corrected p <.025, see Figure S5 for more details).

3.4 | Comparison between sFC-ICC and dFC-ICC for resting state and natural viewing

A previous study has reported that natural viewing condition increased the test–retest reliability of topological properties of sFC compared to resting state (J. Wang, Ren, et al., 2017). In this study, the difference between sFC and dFC was further investigated. Both conditions exhibited relatively higher sFC-ICC than dFC-ICC. Specially, the difference results of sFC-ICC minus dFC-ICC across all the ROI pairs and sub-network pairs were shown in Figure 4. It can be seen that for most of ROI pairs, the sFC-ICC exceeded the dFC-ICC in each condition; their differences were larger for natural viewing, that is, more red points in the difference matrix. Moreover, for network-level, all of the sFC-ICC surpassed the dFC-ICC (Figure 4b). The difference value of sFC-ICC minus dFC-ICC across ROI pairs ranged from −0.914 to 1.324 for resting state and from −0.770 to 1.339 for natural viewing. According to the sub-network difference results in Figure 4b, the intranetwork differences between sFC-ICC and dFC-ICC were much lower than internetwork differences for both conditions.

3.5 | Relationships between sFC, dFC, and their ICCs for resting state and natural viewing

In this part, the relationships between sFC, sFC-ICC, dFC (i.e., dFC Excursion statistic) and dFC-ICC were explored for both conditions. First, the sFC and dFC Excursion matrices were averaged along subjects and sessions. Then the Pearson correlation coefficients between sFC and sFC-ICC, dFC and dFC-ICC, sFC and dFC, sFC-ICC and dFC-ICC for each condition were computed and summarized in Figure 5.

For both conditions, significant strong negative correlations were observed between sFC and dFC (resting state: $r = -.898$, p <.001, 95% CI = [−0.900, −0.895]; natural viewing: $r = -.896$, p <.001, 95% CI = [−0.898, −0.893]) while significant weak positive correlations existed between sFC-ICC and dFC-ICC (resting state: $r = .377$, p <.001, 95% CI = [0.365, 0.389]; natural viewing $r = .291$, p <.001, 95% CI = [0.278, 0.303]). Meanwhile, significant weak negative correlations were found between dFC and dFC-ICC for each condition (resting state: $r = -.281$, p <.001, 95% CI = [−0.294, −0.268]; natural viewing: $r = -.320$, p <.001, 95% CI = [−0.332, −0.307]). As to the relationship between sFC and sFC-ICC, the correlation was very low for resting state ($r = .142$, p <.001, 95% CI = [0.128, 0.155]) while no significant correlation was found in natural viewing condition ($r = .006$, $p = .406$).

4 | DISCUSSION

4.1 | DFC-ICC as a function of the number of fMRI volumes

As mentioned above, the first 215 volumes of nfMRI data were segmented for dFC and dFC-ICC analysis to match with rs-fMRI. When the whole scan data with 530 volumes was selected for dFC evaluation and ICC analysis, the global-level dFC-ICC values is 0.330 for 30-TR-length time window, larger than the 215 volumes of both conditions. Furthermore, the influence of the involved volumes in dFC-ICC analysis was investigated based on rs-fMRI and nfMRI, respectively. With an increment of 5 volumes from 100 volumes to 215 volumes for rs-fMRI and to 530 volumes for nfMRI involved in dFC analysis, the global-level dFC-ICC results were summarized in Figure 6. With the number of volumes as the independent factor and dFC-ICC as the dependent factor, a linear regression model was estimated for rs-fMRI ($R^2 = .985$, p <.001) as well as for nfMRI ($R^2 = .928$, p <.001), suggesting that dFC-ICC was in strong positive relation with the volume number for both conditions. In other words,

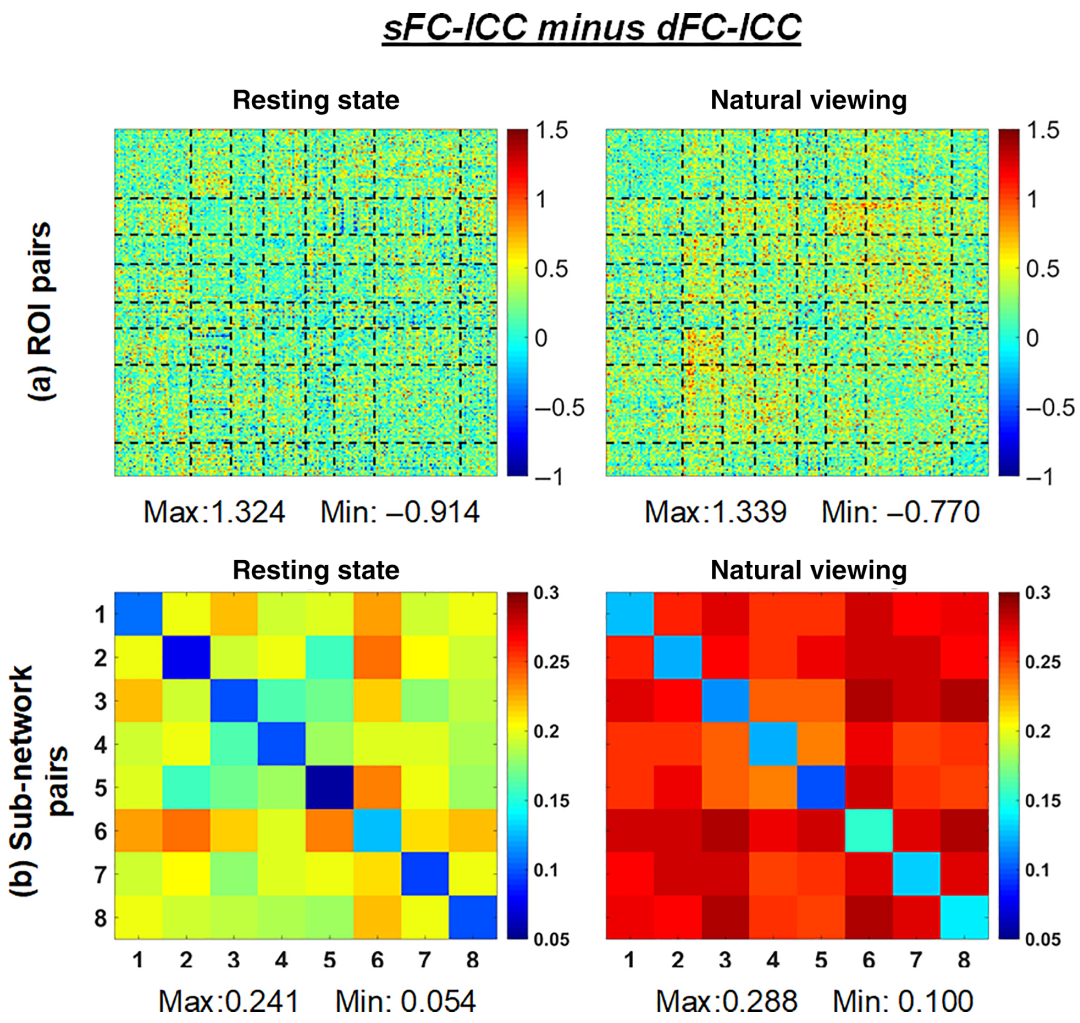

FIGURE 4 The visualization of sFC-ICC minus dFC-ICC for resting state and natural viewing: (a) across all the ROI pairs; (b) across all the sub-network pairs

a longer scan period can achieve relatively better test–retest performance for dFC analysis.

4.2 | The impact of window type in dFC evaluation

In the present study, a rectangle time window was applied in sliding window dFC estimation. However, various types of windows have been proposed to alleviate the effect of sudden changes associated with the edges of the rectangle window (Allen et al., 2014; Mokhtari, Akhlaghi, Simpson, Wu, & Laurienti, 2019). Here, two classical windows, that is, Gaussian window and hamming window, were applied to derive dFC and dFC-ICC based on the Excursion statistic. The interval of sliding window is 1 TR and the window length was set as 45 TR, that is, 1.5 times of the length of rectangle window (Mokhtari et al., 2019). The dFC-ICC matrices and their differences between two conditions using the three types of windows were shown in Figure S6. All the three types of windows resulted in similar spatial patterns of dFC-ICC for both resting state and naturalistic viewing conditions. Furthermore, natural viewing condition increases both global and local dFC-ICC compared to resting state no matter what type of window is applied.

4.3 | The influence of excursion parameters on dFC-ICC

According to Equation 4, two parameters α and β are involved in dFC Excursion calculation. In this section, the influence of the two parameter values on dFC-ICC results are summarized in Figure 7. Using different parameter combinations of α in (.1, .2, .3, ..., 1) and β in (.1, .2, .3, ..., 1), the corresponding global-level dFC-ICC results were recorded and visualized in Figure 7a,b. It can be found that the highest dFC-ICC was achieved using the combination of $\alpha = .8$ and $\beta = .3$ for both conditions. Thus, this parameter combination was adopted in our study. Besides, for the majority of parameter combinations, the natural viewing condition had relatively higher dFC-ICC values than resting state (Figure 7c).

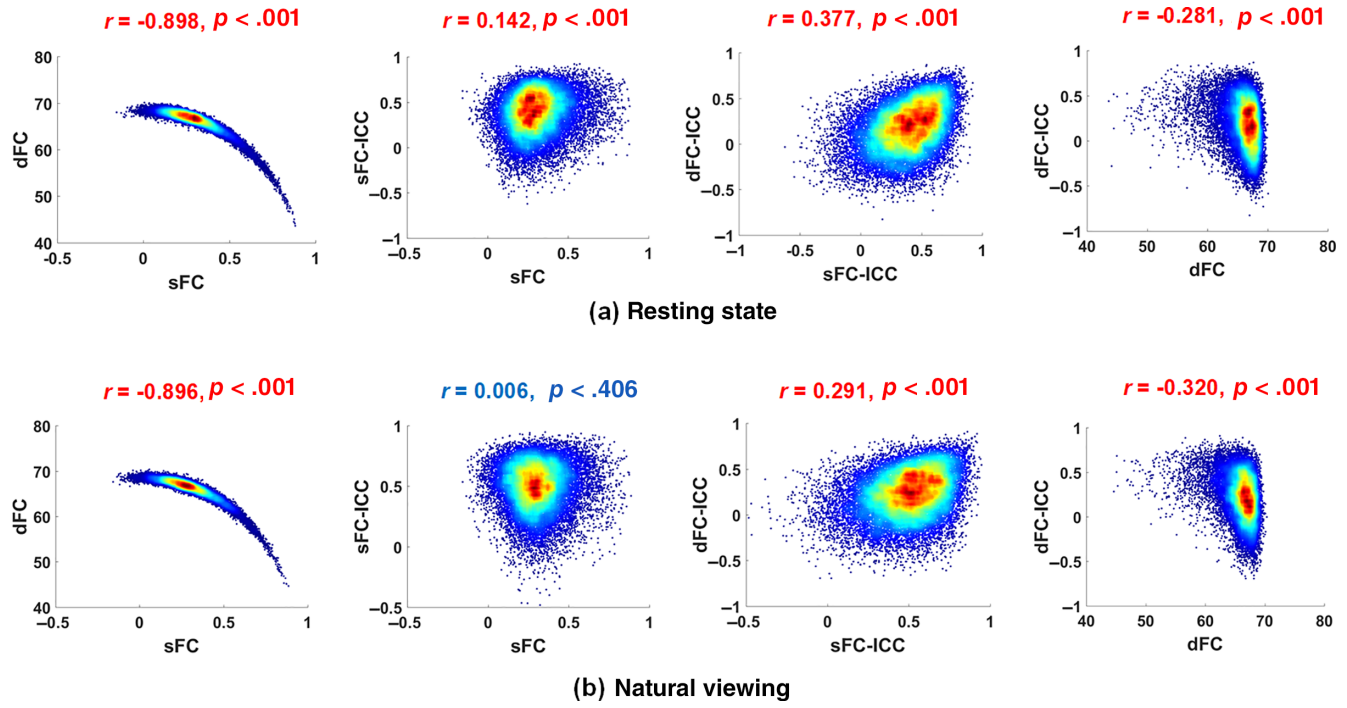

FIGURE 5 The associations between sFC, sFC-ICC, dFC (Excursion) and dFC-ICC matrices as heat maps for all ROI pairs with Pearson correlation coefficients and corresponding *p* values: (a) Resting state (b) Natural viewing

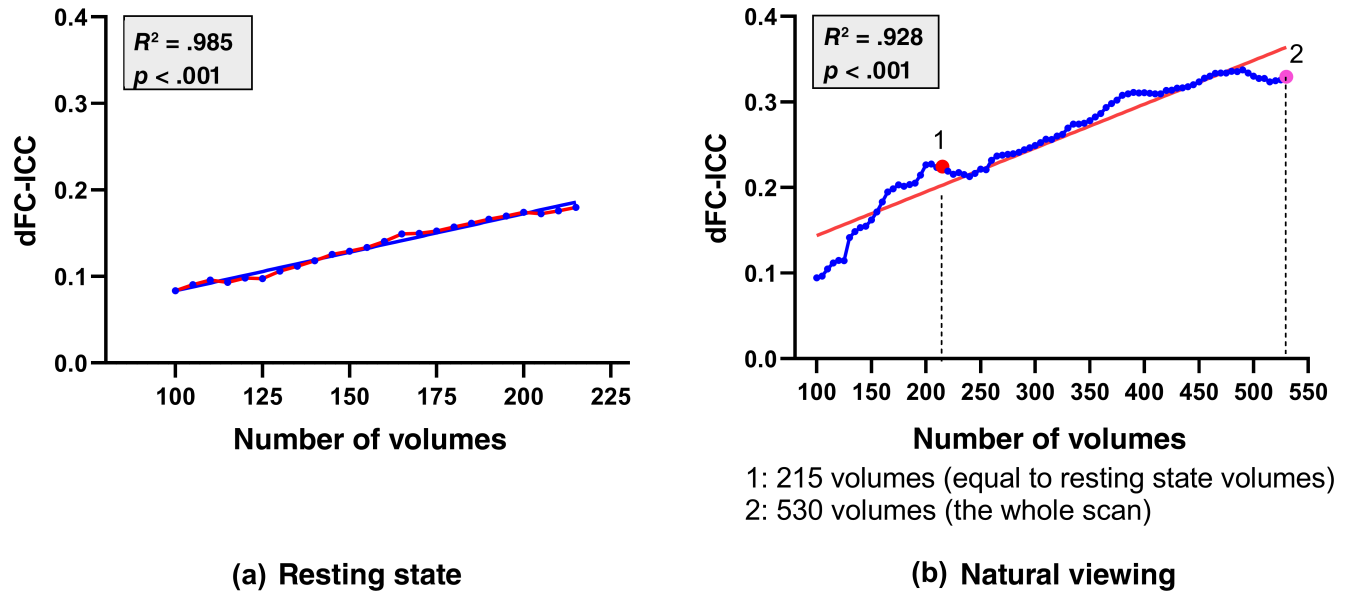

FIGURE 6 The dFC-ICC variations along with the involved number of volumes in dFC analysis with an increment of 5 volumes (in blue). A linear regression model was estimated with dFC-ICC as the dependent factor and the number of volumes as the independent factor (red line) for resting-state (a) and natural viewing (b) conditions, respectively

4.4 | The difference between dFC-ICC and sFC-ICC

According to Figure 4, most of the ROI pairs showed higher sFC-ICC than dFC-ICC and the global-level sFC-ICC is much higher than the global-level dFC-ICC for both resting-state and natural viewing conditions (J. Wang, Ren, et al., 2017). Specially, a total of 2,744 (13.8%) natural viewing ROI pairs showed high dFC-ICC (larger than .5) and 2,193 (11.0%) for resting state, while the number of high sFC-ICC was 10,353 (52.0%) for natural viewing and 6,522 (32.8%) for resting sate.

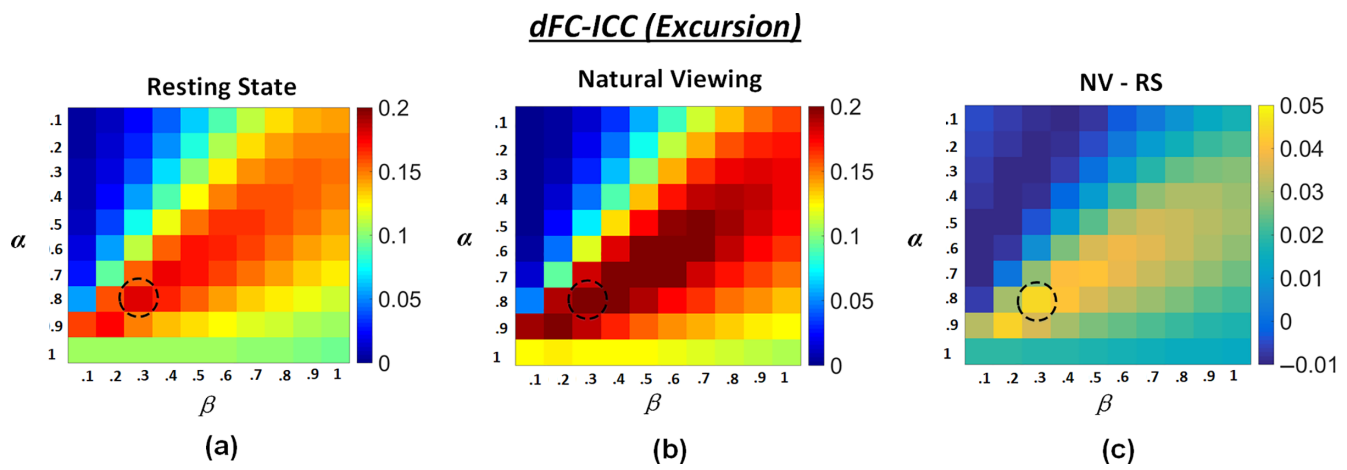

FIGURE 7 The global-level dFC-ICC results using different parameter combinations of α in (.1, .2, .3, ..., 1) and β in (.1, .2, .3, ..., 1) in Excursion estimation: (a) Resting state; (b) Natural viewing; (c) Natural viewing (NV) minus resting state (RS)

These results were consistent with the findings based on high-quality human connectome project (HCP) data and other brain atlas (C. Zhang et al., 2018).

The causes of this phenomenon may be various. First, sFC-ICC was derived directly from sFC matrix while dFC-ICC was derived from the dFC statistic, which was an indirect evaluation of dFC variations (C. Zhang et al., 2018). The statistic may not fully represent or describe the dynamic information included in FC variations. Second, dFC analysis is more challenged than sFC analysis. The introduction of temporal dimension largely increased the difficulty and complexity of dFC analysis. Though sliding window estimation method was commonly used, the exploration of accurate dFC evaluation is still ongoing due to the limitations in current approaches (Hutchison et al., 2013). From this view, the accuracy of dFC evaluation method may influence the dFC-ICC result. Third, high sFC-ICC means high overall consistence between two scans, while dFC-ICC focuses on the consistence of temporal segments during the scan; the latter may be more dynamic, complex and uncertain due to the dynamic characteristic of our brain functional activity, no matter at rest or under naturalistic stimuli. This uncertainty may reduce the test–retest reliability when considering the temporal consistency of FC segments between two scans.

4.5 | Associations between sFC, dFC, and their ICCs

As shown in Figure 5, the associations between sFC, dFC, sFC-ICC, and dFC-ICC were similar for resting-state and natural viewing conditions (C. Zhang et al., 2018). First, strong negative correlation was reported between sFC and dFC statistic, in consistent with Zhang et al.'s finding based on high-quality and large-number resting-state HCP data. Given that sFC characterizes an overall FC level of the whole scan, this negative correlation suggests that when the average FC strength is high, the dFC fluctuation is reduced and vice versa. As to the other three correlations, the strengths were obviously lower than Zhang et al.'s results. The negative correlations between dFC and dFC-ICC suggest that when dFC fluctuates more, the dFC-ICC will be weakened. Then, for a specific fMRI scan, high temporal FC variations will lead to a low average sFC and degrade the dFC-ICC performance. The weak positive correlation between sFC-ICC and dFC-ICC suggests that dFC-ICC is consistent with sFC-ICC to some degree when evaluating the test–retest performance of fMRI paradigms. With respect to the relationship between sFC and sFC-ICC, different from Zhang et al.'s finding, the correlation was much lower for resting state while no correlation was found for natural viewing paradigm. The above results may suggest that the imaging quality and number of samples may influence the ICC evaluation. A larger number of populations and improved imaging quality may be adopted for further validation in future.

4.6 | Impact of brain atlas

To further examine the impact of brain atlas on our findings, the Anatomical Automatic Labeling (AAL) atlas was applied to evaluate whole-brain dFC and dFC-ICC. Using 30-TR-length time window, the global-level dFC-ICC was .255 for natural viewing and .183 for resting state, similar to the results of using Craddock atlas. The relationships between sFC, sFC-ICC, dFC and dFC-ICC were also similar: significant strong negative correlation between sFC and dFC (resting state: $r = -.897$, p <.001, 95% CI = [−0.902, −0.892]; natural viewing: $r = -.894$, p <.001, 95% CI = [−0.899, −0.889]), weak negative correlation between dFC and dFC-ICC (resting state: $r = -.264$, p <.001, 95% CI = [−0.285, −0.242]; natural viewing: $r = -.296$, p <.001, 95% CI = [−0.318, −0.274]) as well as positive correlation between sFC-ICC and dFC-ICC (resting state: $r = .336$, p <.001, 95% CI = [0.315, 0.357]; natural viewing: $r = .311$, p <.001, 95% CI = [0.289, 0.333]). The positive correlation between sFC and sFC-ICC was also very weak ($r = .177$, p <.001, 95% CI = [0.154, 0.200]) for resting state

and natural viewing condition ($r = .035$, $p = .004$, 95% CI = [0.011, 0.059]).

4.7 | Effect of scan order

The effect of scan order on test–retest reliability has been discussed in previous studies. For example, Kristo et al. reported higher test–retest reliability of task-fMRI compared to resting-state fMRI in motor network detection. They performed the motor-task before resting-state scan, but found no scan order effects on the reliability (Kristo et al., 2014). Yang et al. observed significant higher test–retest reliability of task-based cerebral blood follow (CBF) during psychomotor vigilance test (PVT) than resting-state CBF in both pre- and post-task resting-state scans (Yang et al., 2019).

In the current study, the resting-state scan was performed prior to the natural viewing scan as it has been reported in previous studies that prior cognitive states may influence resting-state functional connectivities (Cecchetto, Fischmeister, Reichert, Bagga, & Schopf, 2019; Tambini, Ketz, & Davachi, 2010; Waites, Stanislavsky, Abbott, & Jackson, 2005). We then further inspected the effect of scan order on the test–retest reliability of dFC. Specifically, both the resting-state and natural viewing sessions were divided into two equal and unoverlapped subsessions (the 1st–108th volumes and 109th–215th volumes, respectively). Then the excursion-based global-level dFC-ICC results were compared between subsession-pair using different parameter combinations of α and β (Figure S7). The second subsession was with strongly increased dFC-ICC compared to the first subsession in both resting-state and natural viewing conditions (Figure S7a,b), indicating that scan order could be a confounding factor in the evaluation of test–retest reliability of dFC. However, the dFC-ICC increasement in the second subsession was much higher in natural viewing condition (Figure S7b) compared to that in the resting-state condition (Figure S7a), suggesting that the viewers' better engagement in the storyline of the movie stimuli may help to improve the test–retest reliability of dFC. Furthermore, relatively strong increasement of dFC-ICC was observed in the second subsession of natural viewing condition compared to that in the resting-state condition (Figure S7d), though the dFC-ICC difference between the first subsessions of the two conditions was not significant (Figure S7c). Taken together, these results suggested that natural viewing condition improved dFC-ICC compared to resting state though the effect of scan order may exist.

5 | CONCLUSION

Utilizing two scan sessions of resting-state and natural viewing fMRI data from the same group of subjects, the test–retest reliability of dFC statistics were investigated and compared between the above two conditions. For both conditions, the global dFC-ICC was much lower than the sFC-ICC. Specially, significantly elevated global dFC-ICC and local dFC-ICC in visual, limbic and default mode networks were found in natural viewing paradigm compared with resting state. The association between sFC and dFC demonstrated that the more fluctuations of dFC, the lower sFC will be achieved; weak positive correlation between dFC-ICC and sFC-ICC as well as weak negative correlation between dFC and dFC-ICC were also found. The current study demonstrates that naturalistic fMRI paradigm can improve the test–retest reliability of dFC, providing novel evidence for its promotion in functional brain network studies.

CONFLICT OF INTEREST

The authors declare no potential conflict of interest.

DATA AVAILABILITY STATEMENT

The data used to support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Xin Zhang https://orcid.org/0000-0002-5927-9283
Junwei Han https://orcid.org/0000-0001-5545-7217

REFERENCES

Allen, E. A., Damaraju, E., Plis, S. M., Erhardt, E. B., Eichele, T., & Calhoun, V. D. (2014). Tracking whole-brain connectivity dynamics in the resting state. *Cerebral Cortex*, *24*(3), 663–676. https://doi.org/10.1093/cercor/bhs352

Cecchetto, C., Fischmeister, F. P. S., Reichert, J. L., Bagga, D., & Schopf, V. (2019). When to collect resting-state data: The influence of odor on post-task resting-state connectivity. *NeuroImage*, *191*, 361–366. https://doi.org/10.1016/j.neuroimage.2019.02.050

Chao-Gan, Y., & Yu-Feng, Z. (2010). DPARSF: A MATLAB toolbox for "Pipeline" data analysis of resting-state fMRI. *Frontiers in Systems Neuroscience*, *4*, 13. https://doi.org/10.3389/fnsys.2010.00013

Choe, A. S., Nebel, M. B., Barber, A. D., Cohen, J. R., Xu, Y., Pekar, J. J., … Lindquist, M. A. (2017). Comparing test-retest reliability of dynamic functional connectivity methods. *NeuroImage*, *158*, 155–175. https://doi.org/10.1016/j.neuroimage.2017.07.005

Craddock, R. C., James, G. A., Holtzheimer, P. E., 3rd, Hu, X. P., & Mayberg, H. S. (2012). A whole brain fMRI atlas generated via spatially constrained spectral clustering. *Human Brain Mapping*, *33*(8), 1914–1928. https://doi.org/10.1002/hbm.21333

Damaraju, E., Allen, E. A., Belger, A., Ford, J. M., McEwen, S., Mathalon, D. H., … Calhoun, V. D. (2014). Dynamic functional connectivity analysis reveals transient states of dysconnectivity in schizophrenia. *NeuroImage: Clinical*, *5*, 298–308. https://doi.org/10.1016/j.nicl.2014.07.003

Diez-Cirarda, M., Strafella, A. P., Kim, J., Pena, J., Ojeda, N., Cabrera-Zubizarreta, A., & Ibarretxe-Bilbao, N. (2018). Dynamic functional connectivity in Parkinson's disease patients with mild cognitive impairment and normal cognition. *NeuroImage: Clinical*, *17*, 847–855. https://doi.org/10.1016/j.nicl.2017.12.013

Espinoza, F. A., Liu, J., Ciarochi, J., Turner, J. A., Vergara, V. M., Caprihan, A., … Calhoun, V. D. (2019). Dynamic functional network connectivity in Huntington's disease and its associations with motor and cognitive measures. *Human Brain Mapping*, *40*(6), 1955–1968. https://doi.org/10.1002/hbm.24504

Fiorenzato, E., Strafella, A. P., Kim, J., Schifano, R., Weis, L., Antonini, A., & Biundo, R. (2019). Dynamic functional connectivity changes associated with dementia in Parkinson's disease. *Brain*, *142*(9), 2860–2872. https://doi.org/10.1093/brain/awz192

Guo, C. C., Kurth, F., Zhou, J., Mayer, E. A., Eickhoff, S. B., Kramer, J. H., & Seeley, W. W. (2012). One-year test-retest reliability of intrinsic connectivity network fMRI in older adults. *NeuroImage*, *61*(4), 1471–1483. https://doi.org/10.1016/j.neuroimage.2012.03.027

Hutchison, R. M., Womelsdorf, T., Allen, E. A., Bandettini, P. A., Calhoun, V. D., Corbetta, M., … Chang, C. (2013). Dynamic functional connectivity: promise, issues, and interpretations. *NeuroImage*, *80*, 360–378. https://doi.org/10.1016/j.neuroimage.2013.05.079

Kauttonen, J., Hlushchuk, Y., Jaaskelainen, I. P., & Tikka, P. (2018). Brain mechanisms underlying cue-based memorizing during free viewing of movie Memento. *NeuroImage*, *172*, 313–325. https://doi.org/10.1016/j.neuroimage.2018.01.068

Kim, J., Wang, J., Wedell, D. H., & Shinkareva, S. V. (2016). Identifying core affect in individuals from fMRI responses to dynamic naturalistic audiovisual stimuli. *PLoS One*, *11*(9), e0161589. https://doi.org/10.1371/journal.pone.0161589

Kristo, G., Rutten, G. J., Raemaekers, M., de Gelder, B., Rombouts, S. A., & Ramsey, N. F. (2014). Task and task-free FMRI reproducibility comparison for motor network identification. *Human Brain Mapping*, *35*(1), 340–352. https://doi.org/10.1002/hbm.22180

Kuo, P. C., Tseng, Y. L., Zilles, K., Suen, S., Eickhoff, S. B., Lee, J. D., … Liou, M. (2019). Brain dynamics and connectivity networks under natural auditory stimulation. *NeuroImage*, *202*, 116042. https://doi.org/10.1016/j.neuroimage.2019.116042

Leonardi, N., & Van De Ville, D. (2015). On spurious and real fluctuations of dynamic functional connectivity during rest. *NeuroImage*, *104*, 430–436. https://doi.org/10.1016/j.neuroimage.2014.09.007

Li, X., Zhu, D., Jiang, X., Jin, C., Zhang, X., Guo, L., … Liu, T. (2014). Dynamic functional connectomics signatures for characterization and differentiation of PTSD patients. *Human Brain Mapping*, *35*(4), 1761–1778. https://doi.org/10.1002/hbm.22290

Liu, F., Wang, Y., Li, M., Wang, W., Li, R., Zhang, Z., … Chen, H. (2017). Dynamic functional network connectivity in idiopathic generalized epilepsy with generalized tonic-clonic seizure. *Human Brain Mapping*, *38*(2), 957–973. https://doi.org/10.1002/hbm.23430

Mandelkow, H., de Zwart, J. A., & Duyn, J. H. (2016). Linear Discriminant Analysis Achieves High Classification Accuracy for the BOLD fMRI Response to Naturalistic Movie Stimuli. *Frontiers in Human Neuroscience*, *10*, 128. https://doi.org/10.3389/fnhum.2016.00128

Marshall, C. R., Hardy, C. J. D., Russell, L. L., Bond, R. L., Sivasathiaseelan, H., Greaves, C., … Warren, J. D. (2019). The functional neuroanatomy of emotion processing in frontotemporal dementias. *Brain*, *142*(9), 2873–2887. https://doi.org/10.1093/brain/awz204

Mokhtari, F., Akhlaghi, M. I., Simpson, S. L., Wu, G., & Laurienti, P. J. (2019). Sliding window correlation analysis: Modulating window shape for dynamic brain connectivity in resting state. *NeuroImage*, *189*, 655–666. https://doi.org/10.1016/j.neuroimage.2019.02.001

Nguyen, V. T., Sonkusare, S., Stadler, J., Hu, X., Breakspear, M., & Guo, C. C. (2017). Distinct cerebellar contributions to cognitive-perceptual dynamics during natural viewing. *Cerebral Cortex*, *27*(12), 5652–5662. https://doi.org/10.1093/cercor/bhw334

Noble, S., Scheinost, D., & Constable, R. T. (2019). A decade of test-retest reliability of functional connectivity: A systematic review and meta-analysis. *NeuroImage*, *203*, 116157. https://doi.org/10.1016/j.neuroimage.2019.116157

Preti, M. G., Bolton, T. A., & Van De Ville, D. (2017). The dynamic functional connectome: State-of-the-art and perspectives. *NeuroImage*, *160*, 41–54. https://doi.org/10.1016/j.neuroimage.2016.12.061

Ren, Y., Lv, J., Guo, L., Fang, J., & Guo, C. C. (2017). Sparse coding reveals greater functional connectivity in female brains during naturalistic emotional experience. *PLoS One*, *12*(12), e0190097. https://doi.org/10.1371/journal.pone.0190097

Sonkusare, S., Breakspear, M., & Guo, C. (2019). Naturalistic stimuli in neuroscience: Critically acclaimed. *Trends in Cognitive Sciences*, *23*(8), 699–714. https://doi.org/10.1016/j.tics.2019.05.004

Tailby, C., Masterton, R. A., Huang, J. Y., Jackson, G. D., & Abbott, D. F. (2015). Resting state functional connectivity changes induced by prior brain state are not network specific. *NeuroImage*, *106*, 428–440. https://doi.org/10.1016/j.neuroimage.2014.11.037

Tambini, A., Ketz, N., & Davachi, L. (2010). Enhanced brain correlations during rest are related to memory for recent experiences. *Neuron*, *65*(2), 280–290. https://doi.org/10.1016/j.neuron.2010.01.001

Termenon, M., Jaillard, A., Delon-Martin, C., & Achard, S. (2016). Reliability of graph analysis of resting state fMRI using test-retest dataset from the Human Connectome Project. *NeuroImage*, *142*, 172–187. https://doi.org/10.1016/j.neuroimage.2016.05.062

Tong, Y., Hocke, L. M., & Frederick, B. B. (2019). Low frequency systemic hemodynamic "Noise" in resting state BOLD fMRI: Characteristics, causes, implications, mitigation strategies, and applications. *Frontiers in Neuroscience*, *13*, 787. https://doi.org/10.3389/fnins.2019.00787

Viviano, R. P., Raz, N., Yuan, P., & Damoiseaux, J. S. (2017). Associations between dynamic functional connectivity and age, metabolic risk, and cognitive performance. *Neurobiology of Aging*, *59*, 135–143. https://doi.org/10.1016/j.neurobiolaging.2017.08.003

Waites, A. B., Stanislavsky, A., Abbott, D. F., & Jackson, G. D. (2005). Effect of prior cognitive state on resting state networks measured with functional connectivity. *Human Brain Mapping*, *24*(1), 59–68. https://doi.org/10.1002/hbm.20069

Wang, J., Han, J., Nguyen, V. T., Guo, L., & Guo, C. C. (2017). Improving the test-retest reliability of resting state fMRI by removing the impact of sleep. *Frontiers in Neuroscience*, *11*, 249. https://doi.org/10.3389/fnins.2017.00249

Wang, J., Ren, Y., Hu, X., Nguyen, V. T., Guo, L., Han, J., & Guo, C. C. (2017). Test-retest reliability of functional connectivity networks during naturalistic fMRI paradigms. *Human Brain Mapping*, *38*(4), 2226–2241. https://doi.org/10.1002/hbm.23517

Wang, X., Zhou, T., Wang, P., Zhang, L., Feng, S., Meng, X., … Zhang, Y. (2019). Dysregulation of resting-state functional connectivity in patients with Cushing's disease. *Neuroradiology*, *61*(8), 911–920. https://doi.org/10.1007/s00234-019-02223-y

Wehrle, F. M., Michels, L., Guggenberger, R., Huber, R., Latal, B., O'Gorman, R. L., & Hagmann, C. F. (2018). Altered resting-state functional connectivity in children and adolescents born very preterm short title. *NeuroImage: Clinical*, *20*, 1148–1156. https://doi.org/10.1016/j.nicl.2018.10.002

Yang, F. N., Xu, S., Spaeth, A., Galli, O., Zhao, K., Fang, Z., … Rao, H. (2019). Test-retest reliability of cerebral blood flow for assessing brain function at rest and during a vigilance task. *NeuroImage*, *193*, 157–166. https://doi.org/10.1016/j.neuroimage.2019.03.016

Yeo, B. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., … Buckner, R. L. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology*, *106*(3), 1125–1165. https://doi.org/10.1152/jn.00338.2011

Zalesky, A., & Breakspear, M. (2015). Towards a statistical test for functional connectivity dynamics. *NeuroImage*, *114*, 466–470. https://doi.org/10.1016/j.neuroimage.2015.03.047

Zalesky, A., Fornito, A., Cocchi, L., Gollo, L. L., & Breakspear, M. (2014). Time-resolved resting-state brain networks. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(28), 10341–10346. https://doi.org/10.1073/pnas.1400181111

Zang, Y. F., He, Y., Zhu, C. Z., Cao, Q. J., Sui, M. Q., Liang, M., … Wang, Y. F. (2007). Altered baseline brain activity in children with ADHD revealed by resting-state functional MRI. *Brain and Development*, *29*(2), 83–91. https://doi.org/10.1016/j.braindev.2006.07.002

Zhang, C., Baum, S. A., Adduru, V. R., Biswal, B. B., & Michael, A. M. (2018). Test-retest reliability of dynamic functional connectivity in resting state fMRI. *NeuroImage*, *183*, 907–918. https://doi.org/10.1016/j.neuroimage.2018.08.021

Zhang, X., Guo, L., Li, X., Zhang, T., Zhu, D., Li, K., … Liu, T. (2013). Characterization of task-free and task-performance brain states via functional connectome patterns. *Medical Image Analysis*, *17*(8), 1106–1122. https://doi.org/10.1016/j.media.2013.07.003

Zhang, X., Yu, Y., Shi, Z. S., Xu, K., Feng, J. H., Li, Z. Y., … Wang, W. (2020). Increased resting state functional irregularity of T2DM brains with high HbA1c: sign for impaired verbal memory function? *Brain Imaging and Behavior*, *15*, 772–781. https://doi.org/10.1007/s11682-020-00285-8

Zuo, X. N., & Xing, X. X. (2014). Test-retest reliabilities of resting-state FMRI measurements in human brain functional connectomics: a systems neuroscience perspective. *Neuroscience and Biobehavioral Reviews*, *45*, 100–118. https://doi.org/10.1016/j.neubiorev.2014.05.009

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Zhang, X., Liu, J., Yang, Y., Zhao, S., Guo, L., Han, J., & Hu, X. (2022). Test–retest reliability of dynamic functional connectivity in naturalistic paradigm functional magnetic resonance imaging. *Human Brain Mapping*, *43*(4), 1463–1476. https://doi.org/10.1002/hbm.25736