Reliability of EEG Measures in Driving Fatigue

Jonathan Harvy, Anastasios Bezerianos, *Senior Member, IEEE*, and Junhua Li, *Senior Member, IEEE*

***Abstract*— Reliability investigation of measures is important in studies of brain science and neuroengineering. Measures' reliability hasn't been investigated across brain states, leaving unknown how reliable the measures are in the context of the change from alert state to fatigue state during driving. To compensate for the lack, we performed a comprehensive investigation. A two-session experiment with an interval of approximately one week was designed to evaluate the reliability of the measures at both sensor and source levels. The results showed that the average intraclass correlation coefficients (ICCs) of the measures at the sensor level were generally higher than those at the source level, except for the directed between-region measures. Single-region measures generally exhibited higher average ICCs relative to between-region measures. The exploration of brain network topology showed that nodal metrics displayed highly varying ICCs across regions and global metrics varied associated with nodal metrics. Single-region measures displayed higher ICCs in the frontal and occipital regions while the between-region measures exhibited higher ICCs in the area involving frontal, central and occipital regions. This study provides an appraisal for the measures' reliability over a long interval, which is informative for measure selection in practical mental monitoring.**

***Index Terms*— Driving fatigue, EEG, brain network, functional connectivity, graph metrics, sensor and source levels.**

I. Introduction

DRIVING fatigue has been considered as one of the fatal causes of traffic accidents, accounting for 20% of all road fatalities worldwide [1]. Prolonged driving on a monotonous road environment reduces driver's vigilance, further resulting in driving fatigue [2]. To date, various psychophysiological signals have been used to assess fatigue and EEG is a relatively reliable and easily-used indicator for fatigue [3], [4]. When selecting a measure, its reliability over time is important as high reliability ensures that driving fatigue can be correctly and accurately assessed. Previous studies only investigated the reliability of single-region measures during different episodes of fatigue [5], while between-region measures have not yet been investigated. This requires a comprehensive investigation of all measures to address how reliable each measure is and compare the reliability between each of them in terms of identifying driving fatigue.

Early EEG studies utilized individual-region measures, such as entropy [6] and spectral power [4], [7], [8], [9], [10], [11] to assess driving fatigue. A study reported decreased sample entropy in the occipital region during driving fatigue [6]. Similar decreases during fatigue were found in central, parietal, occipital regions using entropy. Considering different frequency bands relevant to driving fatigue, previous studies using EEG spectral power reported distinct changes from alert to fatigue. Spectral power in theta and alpha bands increased during fatigue while spectral power in beta band decreased [8], [9], [10]. Increases of spectral power in theta band were found in frontal, central and occipital regions [4], [11]. Spectral power in alpha band increased in central, parietal, occipital, and temporal regions during fatigue [4], [10], [11]. Decreases of beta band during fatigue were observed in frontal, central, temporal, parietal, and occipital regions [4], [10], [11]. Although changes in delta and gamma bands during fatigue have been reported, more prominent changes were frequently reported in theta, alpha, and beta bands [12], [13].

Between-region measures have been increasingly used and widely applied to diverse neuroimaging studies, such as motor imagery performance prediction [14], schizophrenia identification [15], and fatigue identification [16], [17], [18], [19], [20], [21]. Increases of mean phase coherence in frontal and parietal regions were found in the delta and alpha bands under fatigue [17]. In another study, interhemispheric connections in alpha band showed an increase while higher connection strengths were observed for interhemispheric frontal and occipital connections relative to interhemispheric central, parietal, and temporal connections during fatigue [18]. Graph metrics have been utilized to capture the properties of brain functional connectivity during fatigue [19], [20]. In a study using ordinary coherence, total synchronization strengths (in the frequency range of 0.5∼30 Hz) in frontal, central,

Manuscript received 28 January 2022; revised 25 July 2022 and 22 August 2022; accepted 18 September 2022. Date of publication 21 September 2022; date of current version 30 September 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 61806149 and in part by the Guangdong Basic and Applied Basic Research Foundation under Grant 2020A1515010991. *(Corresponding author: Junhua Li.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Institutional Review Board of the National University of Singapore IRB under Approval No. B-15-169, Dated 09 October 2015.

Jonathan Harvy was with the Singapore Institute for Neurotechnology (SINAPSE), National University of Singapore, Singapore 117456. He is now with the Laboratory for Brain-Bionic Intelligence and Computational Neuroscience, Wuyi University, Jiangmen 529020, China.

Anastasios Bezerianos is with the Hellenic Institute of Transport, Center for Research and Technology Hellas (CERTH), 57001 Thermi, Greece.

Junhua Li was with the Singapore Institute for Neurotechnology (SINAPSE), National University of Singapore, Singapore 117456. He is now with the School of Computer Science and Electronic Engineering, University of Essex, CO4 3SQ Colchester, U.K., and also with the Laboratory for Brain-Bionic Intelligence and Computational Neuroscience, Wuyi University, Jiangmen 529020, China (e-mail: juhalee.bcmi@gmail.com).

Digital Object Identifier 10.1109/TNSRE.2022.3208374

and temporal regions, mean degree in delta and theta, and mean clustering coefficient in delta, theta, and alpha displayed significant increases while characteristic path length significantly decreased in delta, theta, alpha, and beta bands during fatigue [19]. A study using phase lag index (PLI) reported increases of connection strengths in delta band and changes of minimum spanning tree metrics in delta and theta bands towards star-like network configuration during fatigue [20]. Directed measures such as partial directed coherence and directed transfer function have also been utilized to detect driving fatigue [21]. Similar to the single-region measures, between-region measures for fatigue assessment were more commonly reported in theta, alpha, and beta bands [12], [13].

Previous studies have utilized various single-region and between-region measures of EEG, capturing different properties of the brain regional and inter-regional activities. To determine their potential usefulness as biomarkers of particular brain functions, we could estimate the reliability of the measures [22]. Most of the previous studies analyzed the reliability of the measures during the resting condition [22], [23], [24], [25]. A study using fMRI reported higher reliability of the second order graph metrics than that of the first order graph metrics [23]. In a MEG study comparing different connectivity measures, higher reliability was observed for amplitude envelope correlation and partial correlation measures while phase-based measures and imaginary partial coherence displayed lower reliability [24]. In a study using EEG at sensor level and source level, the reliability of the graph metrics was higher for within-day sessions relative to between-day recordings and higher at sensor level compared to source level [22]. In the other EEG study utilizing phase-based measures comprising PLI and weighted PLI, the reliabilities of the two measures were found to be similar and the measures in alpha band had the highest reliability [25]. While wPLI exhibited lower global metric reliabilities relative to PLI, wPLI displayed higher reliabilities of regional degree and inter-regional connections [25]. The other studies have also analyzed the reliability of EEG measures using task-based protocol [5], [26]. In a study using working memory tasks, the global metrics were more reliable in the lower frequency bands and during the task compared to during the rest [26]. In a study using single-region measures during different episodes of driving fatigue, mean EEG amplitudes in delta, theta, alpha, and beta bands were generally highly reliable (Pearson's correlation, $r > 0.6$) with the highest observed in delta and theta bands ($r > 0.95$) [5]. Although these studies have discussed the reliability of measures, they did not consider the changes of the measures between different brain states. The studies also reported the reliability of single-region or between-region measures separately, neglecting the comparison between the two categories.

Since subjects may have different baseline values for measures, we proposed to estimate the reliability of the measure changes between alert and fatigue states in this study. Investigating the reliability of measure changes across states, instead of measure values, might be more useful for fatigue detection because it enables the evaluation of the consistency of measures capturing the changes of brain states over time. In this study, we showed comparative results in terms of reliability among measures for driving fatigue and explored those measures at both sensor level and source level.

II. Methodology

A. Experimental Protocol

Thirty healthy students, 18 males and 12 females (age: 23.17 ± 2.72 years, mean ± standard deviation), were recruited from the National University of Singapore. All subjects reported normal or corrected-to-normal vision, with no history of substance addiction or mental disorders. The subjects were required to obtain a full night (>7 h) sleep before the day of the experiment. On the day of the experiment, they were required to avoid consuming caffeine or alcohol. Each subject signed a consent form and was trained to familiarize themselves with the driving equipment before the start of the experiment. The driving simulation was conducted using Logitech G27 Racing Wheel set and Carnetsoft Driving Simulator (http://cs-driving-simulator.com) software. The subjects were instructed to drive a car following a guiding car and to brake as soon as the red taillights of the guiding car lit. Each subject completed two identical driving sessions of 90 minutes, with an interval of approximately one week. The experiment was reviewed and approved by the institutional review board of the National University of Singapore.

B. EEG Data Acquisition and Source Localization

Brain activity was recorded as EEG using wireless EEG recording equipment with 24 dry electrodes (Cognionics, Inc., USA), with a sampling rate of 250 Hz. The impedances of all EEG channels were kept below 20 kΩ. The EEG channels were referenced to the linked mastoids. Preprocessing steps were performed to remove artifacts. Firstly, all EEG channels were rereferenced using common average reference (an alternative reference is infinity [27], [28]). The EEG channels having poor contact with the scalp were removed and then respectively interpolated using the signals from its adjacent channels. The last 5-min portion of EEG was discarded due to the change of the simulation phase into free driving where there was no guiding car. The EEG signals were band-pass filtered at 0.5∼45 Hz. The processed signals were segmented into epochs of a 2-second period. Abnormal epochs containing values with more than 5 times standard deviation from the mean probability distribution were removed using EEGLAB [29]. Based on the self-reported confirmation of fatigue after the experiment and the increased reaction time at the end of the experiment, the epochs between the 0th and 15th minute and between the 70th and 85th minute were considered as alert and fatigue samples respectively. Four subjects having the insufficient number of alert and fatigue epochs in either session after epoch rejection were excluded from further analysis. For the remaining subjects, the remaining epochs were decomposed into components using independent component analysis (ICA). ICA components representing artifacts were removed and the remaining components were used to reconstruct clean EEG epochs. Clean EEG epochs were then obtained in the first session (alert: 391.12 ± 51.81, fatigue:

340.35 ± 91.56) and the second session (alert: 377.15 ± 54.67, fatigue: 363.50 ± 70.93) of the experiment.

The exact low resolution brain electromagnetic tomography (eLORETA) [30] was used in this study to transform the EEG signals at sensor level to the cortical current source densities. The head model of eLORETA was based on the Montreal Neurological Institute average MRI brain map (MNI152) [31]. The solution space was restricted to the cortical gray matter with 6239 voxels at 5 mm spatial resolution. The voxels at the cortical gray matter were then grouped into 80 cortical regions, based on the automated anatomical labeling (AAL) brain atlas. In this study, the 24-channel EEG signals were transformed into 80-region EEG at source level. Single-region and between-region measures were then computed from alert and fatigue epochs at sensor level and source level separately for the first and second sessions.

C. Single-Region Measures

The single-region measures utilized in this study were sample entropy and power spectral density. Sample entropy (SE) is an optimized method of approximate entropy which estimates the complexity of the time series of the data without including the self-matches when computing the probability [32]. The parameters of SE were set to m = 2 and r = 0.2×std [6]. In this study, SE was computed at the whole band (0.5∼45 Hz) for all sensor level and source level regions. Power spectral density (PSD) was utilized to measure the EEG activity at the aforementioned bands relevant to fatigue identification, theta, alpha, and beta bands. PSD was obtained by computing the ratio between the power of each band and the power of the total band [11]. In this study, PSD theta, alpha, and beta were calculated for all sensor level and source level regions. The differences of SE and PSD values between alert and fatigue were obtained before computing the reliability, resulting in 24 and 80 values at sensor level and source level respectively.

D. Between-Region Measures

In this study, phase lag index (PLI) and partial directed coherence (PDC) were utilized as undirected and directed between-region measures respectively. PLI measures between-region synchronizations by computing the instantaneous phase differences between regions, minimizing the effect of volume conduction [33]. PDC is a frequency-based measure based on Granger causality, estimating the directed information flow between regions [34]. For each measure, the individual connections and their corresponding graph metrics were computed. The individual connections of PLI were computed for the alert and fatigue epochs between all sensor level and source level regions, resulting in 276 and 3160 connections respectively. The individual connections of PDC were calculated for the alert and fatigue epochs, resulting in 552 for sensor level and 6320 for source level connections (excluding the self-connections). The graph metrics of PLI and PDC comprised global (clustering coefficient, characteristic path length, global efficiency, and local efficiency) and nodal (nodal efficiency and nodal clustering coefficient) metrics. The PLI and PDC matrices were first thresholded at the sparsity range of 10% to 40% with 1% increment to obtain their distinct characteristics at the different number of edges. Each graph metric was computed at each sparsity threshold and its area under the curve (AUC) along the sparsity range was computed. The difference of the average AUCs in alert and fatigue states were computed before estimating their reliabilities. Each global metric had one value per subject per session while each nodal metric had 24 (80) values per subject per session at sensor (source) level.

E. Reliability and Statistical Analysis

The single-region and between-region measures were obtained for alert and fatigue epochs. The average difference between alert and fatigue across epochs of each subject was then computed separately for the first and second sessions. See the box named 'Compute Reliability' depicted in Fig. 1. Intraclass correlation coefficient (ICC) [35] was then computed to measure the reliability of the differences across subjects over the two sessions. Specifically, 26 difference values for each session were grouped to compute the mean M and the variance V^2. The ICC was computed by

$$ICC = \frac{1}{(N-1)\,V^2}\sum_{n=1}^{N}(x_n^1 - M)(x_n^2 - M)$$

where N is the number of difference values for each session. $x_n^1 \in \{x_1^1, x_2^1, \cdots, x_{26}^1$ and $x_n^2 \in \{x_1^2, x_2^2, \cdots, x_{26}^2\}$ are difference values for session 1 and session 2, respectively. ICC was set to zero when it was a negative value.

To determine whether the ICC distributions of the single-regions and between-regions measures were different, Kruskal-Wallis test was conducted for the sensor level measures and source level measures separately. Wilcoxon signed-rank test was conducted for the comparison between SE and PSD and between PLI and PDC, while Wilcoxon rank-sum test was conducted for the comparison between SE/PSD and PLI/PDC.

III. Results

The ICCs of the measures at sensor level and source level were depicted in Fig. 2. Kruskal-Wallis test showed significant differences ($p<0.05$) among the measures at both sensor level and source level. The results of the post-hoc tests between the measures were shown in Fig. 3 and Fig. 4 for the sensor level and source level measures, respectively.

At sensor level, PSD theta had the highest mean ICC and its ICCs were significantly higher than the ICCs of the other measures. PSD alpha had the second highest mean ICC and its ICCs were significantly higher than the ICCs of the other measures except for sample entropy, PSD beta, and PLI alpha, having similar mean ICCs among them. While PLI alpha was significantly higher than PLI theta, PLI beta, and PDC measures, sample entropy and PSD beta were significantly higher than PLI theta and PLI beta only. The lowest mean ICCs were found for PLI theta and PLI beta.

At source level, PSD alpha had the highest mean ICC and its ICCs were significantly higher than the other measures. PLI measures had the lowest mean ICCs and its ICCs were

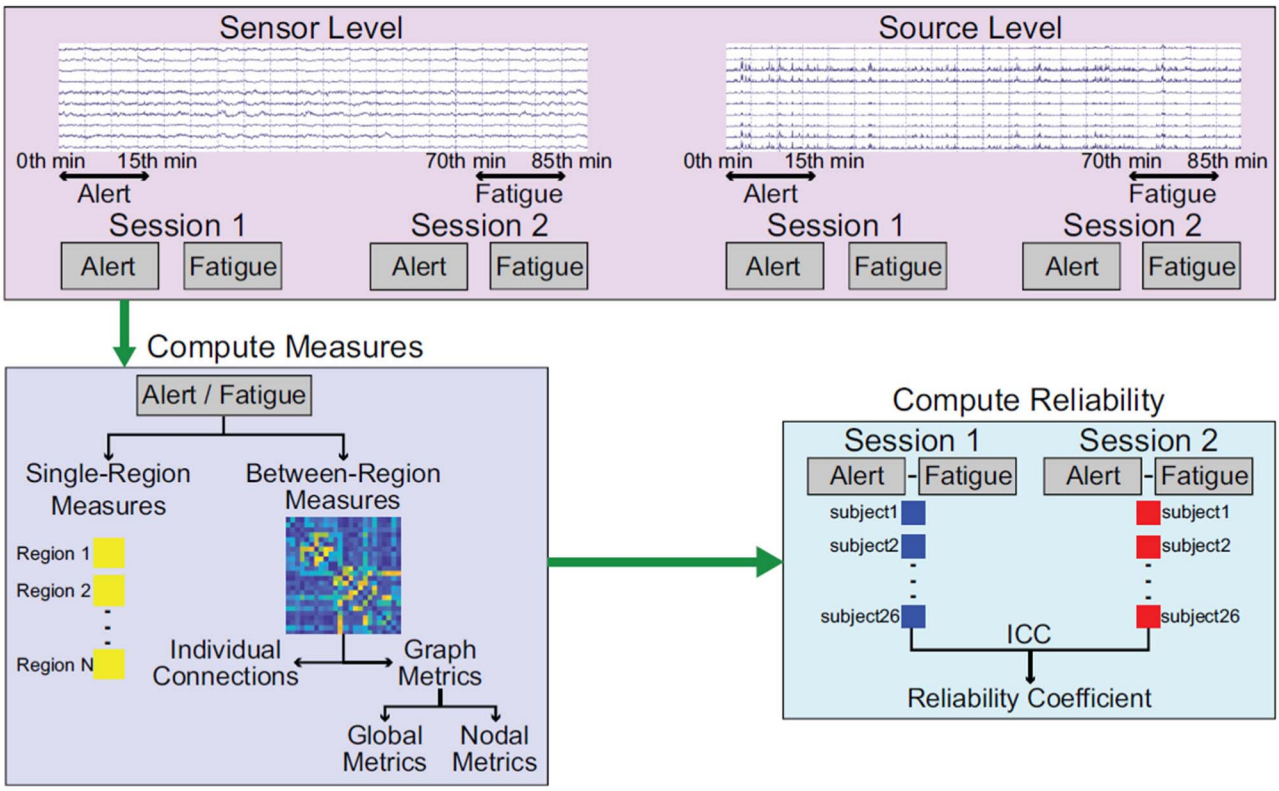

Fig. 1. The illustration of intraclass correlation coefficient (ICC) computation of the measures. The EEG data at sensor level were collected from the two-session experiment. Exact LORETA was used for transforming the EEG at sensor level to EEG at source level. The EEG at both sensor level and source level were segmented into alert and fatigue epochs. For each alert / fatigue epoch, the single-region and between-region measures were calculated. The measure change for each subject was computed by subtracting the average of the measure values in the fatigue epochs from the average of the measure values in the alert epochs. The reliability coefficient was then calculated using ICC across the two sessions.

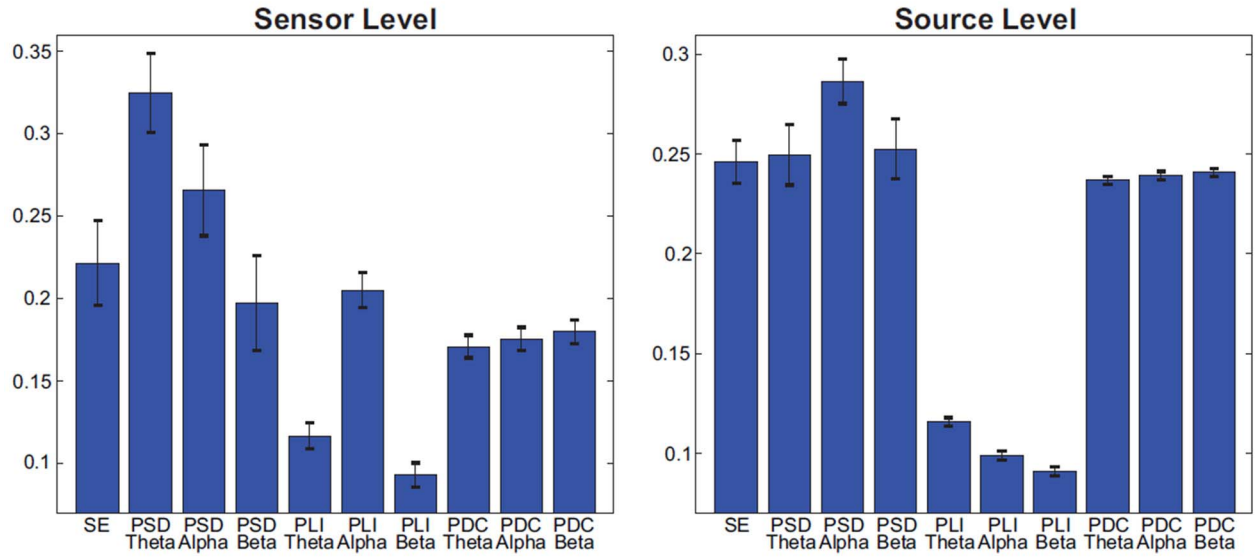

Fig. 2. The means and standard errors of the single-region and between-region measures. The measures at the sensor level were depicted in the left panel while the measures at the source level were shown in the right panel.

significantly lower than the ICCs of the other measures. Sample entropy, PSD theta, and PSD beta had similar mean ICCs among them. PDC measures had similar mean ICCs among the different frequency bands although PDC beta had the highest ICC, followed by PDC alpha and PDC theta, and the differences were significant.

The ICCs of the global and nodal metrics computed from between-region measures were listed in Table I and Table II

TABLE I
RELIABILITY OF THE GLOBAL METRICS OF PLI AND PDC MEASURES

Measure	Graph Metric	ICC (Sensor Level)	ICC (Source Level)
PLI Theta	Clustering Coefficient	0.4179	0.3493
	Characteristic Path Length	0.0300	0.3366
	Global Efficiency	0.0701	0.3505
	Local Efficiency	0.2753	0.3570
PLI Alpha	Clustering Coefficient	0.5763	0.1700
	Characteristic Path Length	0.4595	0
	Global Efficiency	0.6039	0.0274
	Local Efficiency	0.5970	0.0850
PLI Beta	Clustering Coefficient	0	0
	Characteristic Path Length	0	0
	Global Efficiency	0	0
	Local Efficiency	0	0
PDC Theta	Clustering Coefficient	0.3470	0
	Characteristic Path Length	0.4558	0.1133
	Global Efficiency	0.2614	0.0066
	Local Efficiency	0.2261	0
PDC Alpha	Clustering Coefficient	0.3137	0
	Characteristic Path Length	0.4802	0.1127
	Global Efficiency	0.1496	0.0326
	Local Efficiency	0.2254	0
PDC Beta	Clustering Coefficient	0.2467	0
	Characteristic Path Length	0.4985	0.0970
	Global Efficiency	0.0165	0.0713
	Local Efficiency	0.1711	0

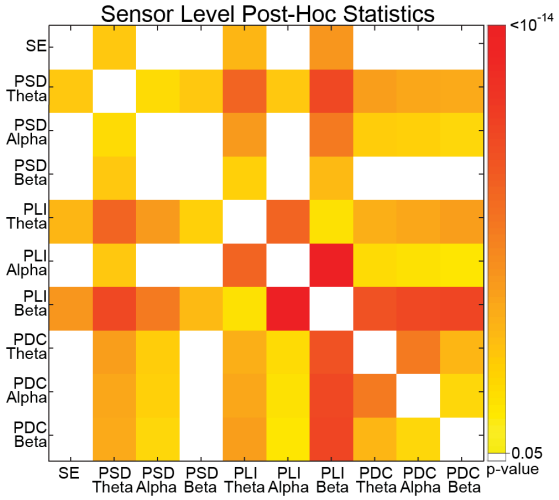

Fig. 3. The post-hoc statistics between the measures at sensor level. White boxes indicate the non-significant differences while the colored boxes refer to the significant differences.

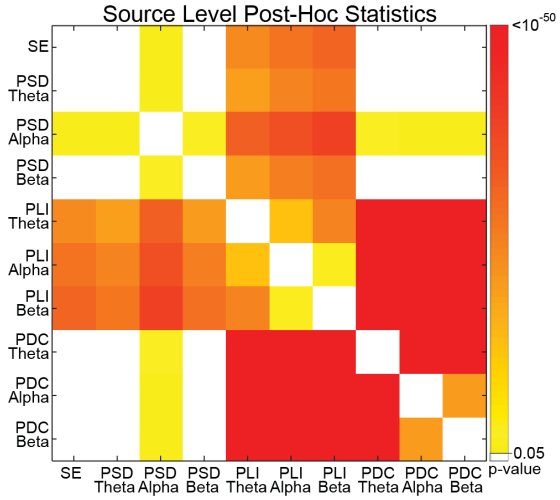

Fig. 4. The post-hoc statistics between the measures at source level. White boxes indicate the non-significant differences while the colored boxes refer to the significant differences.

respectively. Among the PLI measures at sensor level, PLI alpha had the highest ICCs for both global and nodal metrics, followed by PLI theta. PLI beta had poor ICCs (ICC < 0.2) for the global and nodal metrics. The graph metrics of PLI alpha at source level had lower ICCs than the respect metrics at sensor level. The graph metrics of PLI beta showed poor ICCs while the graph metrics of PLI theta at source level generally had higher ICCs than the corresponding metrics at

TABLE II
RELIABILITY OF THE NODAL METRICS (MEAN ± STD) OF PLI AND PDC MEASURES

Measure	Graph Metric	ICC (Sensor Level)	ICC (Source Level)
PLI Theta	Nodal Clustering Coefficient	0.1663 ± 0.1450	0.1921 ± 0.1455
	Nodal Efficiency	0.1369 ± 0.1428	0.1657 ± 0.1332
PLI Alpha	Nodal Clustering Coefficient	0.3149 ± 0.1969	0.1536 ± 0.1386
	Nodal Efficiency	0.3275 ± 0.2188	0.1272 ± 0.1312
PLI Beta	Nodal Clustering Coefficient	0.0459 ± 0.0863	0.1220 ± 0.1453
	Nodal Efficiency	0.0485 ± 0.0847	0.1301 ± 0.1350
PDC Theta	Nodal Clustering Coefficient	0.2123 ± 0.1411	0.1115 ± 0.1293
	Nodal Efficiency	0.1009 ± 0.1319	0.1728 ± 0.1622
PDC Alpha	Nodal Clustering Coefficient	0.1937 ± 0.1225	0.1168 ± 0.1376
	Nodal Efficiency	0.1059 ± 0.1228	0.1812 ± 0.1683
PDC Beta	Nodal Clustering Coefficient	0.1623 ± 0.1039	0.1207 ± 0.1437
	Nodal Efficiency	0.0987 ± 0.1100	0.1844 ± 0.1776

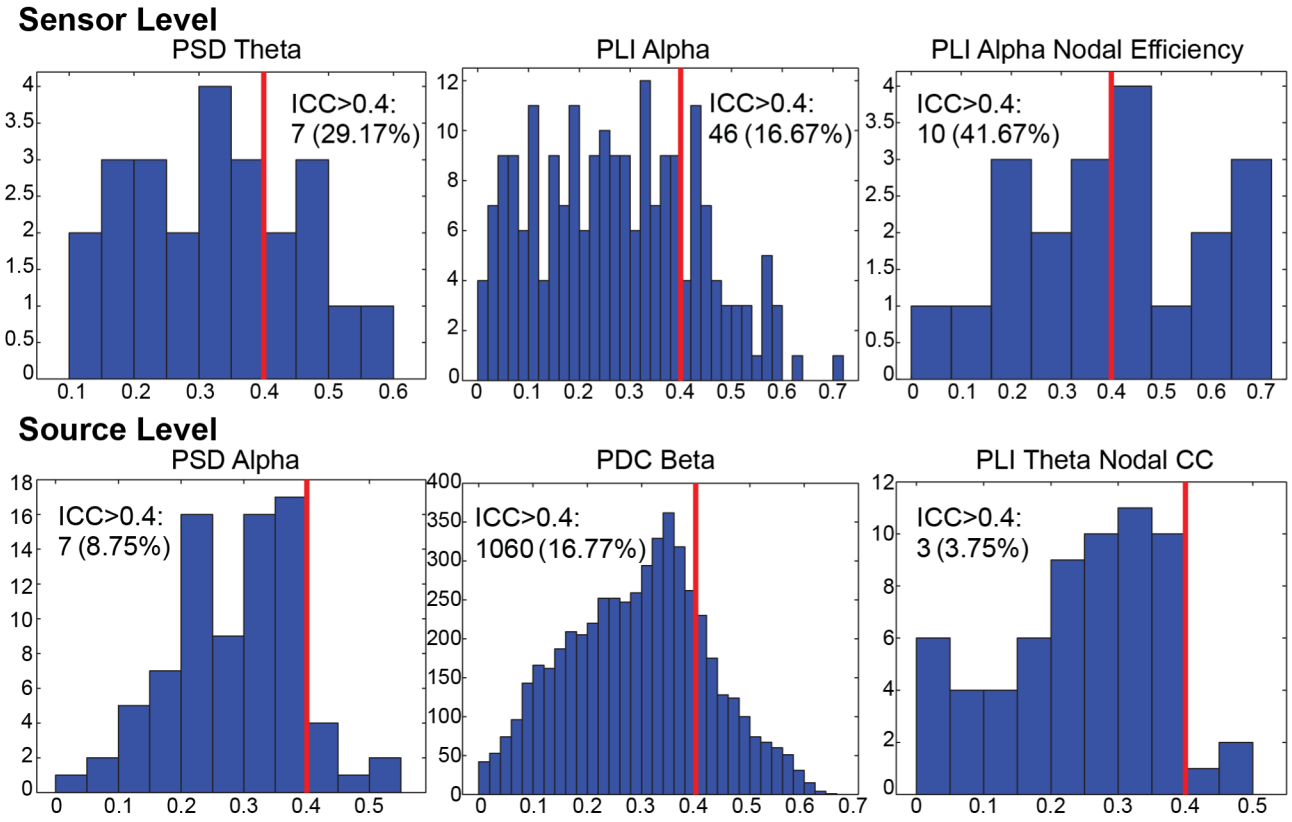

Fig. 5. The ICC distributions of the selected measures for single-region, between-region, and nodal metrics. The top panel and bottom panel showed the distributions of the measures at sensor and source level respectively. The horizontal axis represents ICC values and the vertical axis represents the number of occurrences. The red lines in each plot refer to ICC = 0.4 and the percentages indicate the regions / connections having ICCs higher than 0.4.

sensor level, except for the clustering coefficient. For the PDC measures, global metrics had higher ICCs at the sensor level than the metrics at source level except for global efficiency of PDC beta. The nodal clustering coefficients of PDC measures were higher at sensor level than the metrics at source level, while the opposite case occurred for the nodal efficiency of PDC measures.

The measures with the highest ICCs in different categories, such as single-region, individual connections, and nodal metrics, were selected for visualization. The ICC distributions of the selected measures were depicted in Fig. 5. For the single-region category, PSD theta and PSD alpha had the highest mean ICCs at sensor level and source level respectively. PLI alpha and PDC beta had the highest mean ICCs for

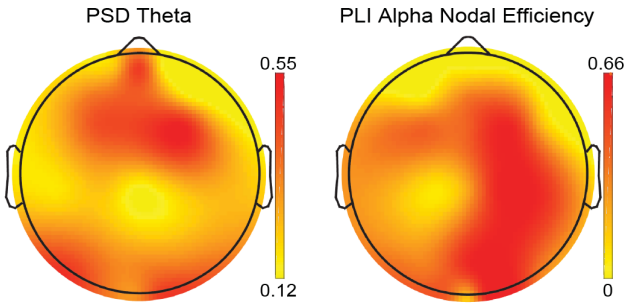

Fig. 6. The topographies of ICC values of PSD theta and PLI alpha (nodal efficiency) at the sensor level.

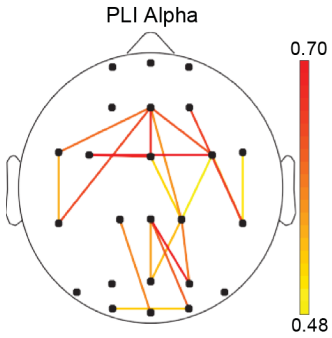

Fig. 7. The top 20 connections of PLI alpha having the highest ICC values at the sensor level.

between-region measures. For the nodal metrics, PLI alpha (nodal efficiency) and PLI theta (nodal clustering coefficient) had the highest mean ICCs. In Fig. 5, single-region measures and nodal metrics had a higher percentage of regions with ICC $>$ 0.4 at sensor level relative to the measures at source level while the between-region measures with ICC $>$ 0.4 had similar percentage of individual connections at sensor level and source level.

The ICCs of the selected single-region measure, nodal metric, and between-region measure at the sensor level were depicted in Fig 6 and Fig 7. In Fig. 6, higher ICCs for PSD theta were observed in frontal and occipital regions while higher ICCs for PLI alpha (nodal efficiency) were mainly found from frontal to occipital regions in the right hemisphere. In Fig. 7, twenty connections with the highest ICCs were shown for PLI alpha. The connections were mainly found between frontal and central regions and between central and parietal/occipital regions.

The selected measures at source level were depicted in Fig. 8, Fig. 9, and Fig. 10 for single-region measure, nodal metric, and between-region measure, respectively. In Fig. 8, the right frontal, right temporal, right parietal and occipital regions had higher ICCs for PSD alpha. For PDC beta, twenty connections with the highest ICCs were shown in Fig. 9. The connections were observed mainly from right superior frontal gyrus (medial part) and right anterior cingulate gyrus to the other parietal, temporal, and occipital regions. For PLI theta (nodal clustering coefficient), higher ICCs were found in frontal, temporal, and occipital regions, shown in Fig. 10.

IV. Discussion

In this study, we comprehensively investigated the reliability of the EEG measures for driving fatigue identification. Our study explored the reliability of measure changes, instead of measure values, to evaluate the consistency of the changes from alert to fatigue. We estimated the reliability across two sessions with a long interval in between, instead of two episodes within a session, since such estimation is closer to the practical use of fatigue detection which requires reliable performance across days of operation. We compared the reliability of the single-region measures with the between-region measures and discussed the results in detail below.

From the single-region measures at sensor level, we observed differences in the ICCs of PSD measures relative to SE. Among the single-region measures, PSD theta (significant, $p<0.05$) and PSD alpha (not significant, $p>0.05$) had higher mean ICCs relative to SE while PSD beta had lower mean ICCs (not significant, $p>0.05$). Previous spectral EEG study also found that EEG activity in theta band had the highest correlation coefficients between two episodes of driving fatigue, followed by that in alpha and beta bands [5]. At source level, the ICCs of PSD alpha were significantly higher than PSD theta, PSD beta, and SE. In this study, higher ICCs were found at lower frequency bands. This might reflect the distinct consistencies of the single-region measures in particular frequency bands during driving fatigue.

At both sensor level and source level, single-region measures generally had higher mean ICCs than individual connections from between-region measures. This observation might indicate the difference between the regional activities and inter-regional interactions in identifying brain state changes. While the consistency of regional activities from alert to fatigue depends only on the individual regions, the consistency of inter-regional interactions relies on the changes involving any two regions. This more complex mechanism in between-region interactions might be reflected by their overall lower reliability relative to the reliability of the regional activities.

The ICCs of the measures at sensor level were generally higher than those at source level, except for PDC measures. For single-region measures, the higher percentage of regions with ICC $>$ 0.4 were found at sensor level relative to source level, suggesting that single-region measures are more reliable at sensor level compared to those at source level. This finding is in agreement with the previous statement in a study investigating the reliability of EEG measures at both sensor and source levels [22], probably due to the volume conduction effect at the sensor level which was highly repeatable across sessions and subjects [24], [37].

At sensor level, PLI measures generally exhibited lower mean ICCs except for PLI alpha. In the previous study comparing MEG-based between-region measures [24], phase-based measures also showed relatively lower reliabilities. The low ICCs of PLI might be caused by its method of minimizing the volume conduction effect [24] and relying on subtle properties of the signals which were harder to estimate and more variable across subjects [37]. Compared to the other between-region measures, PLI alpha at sensor level displayed higher

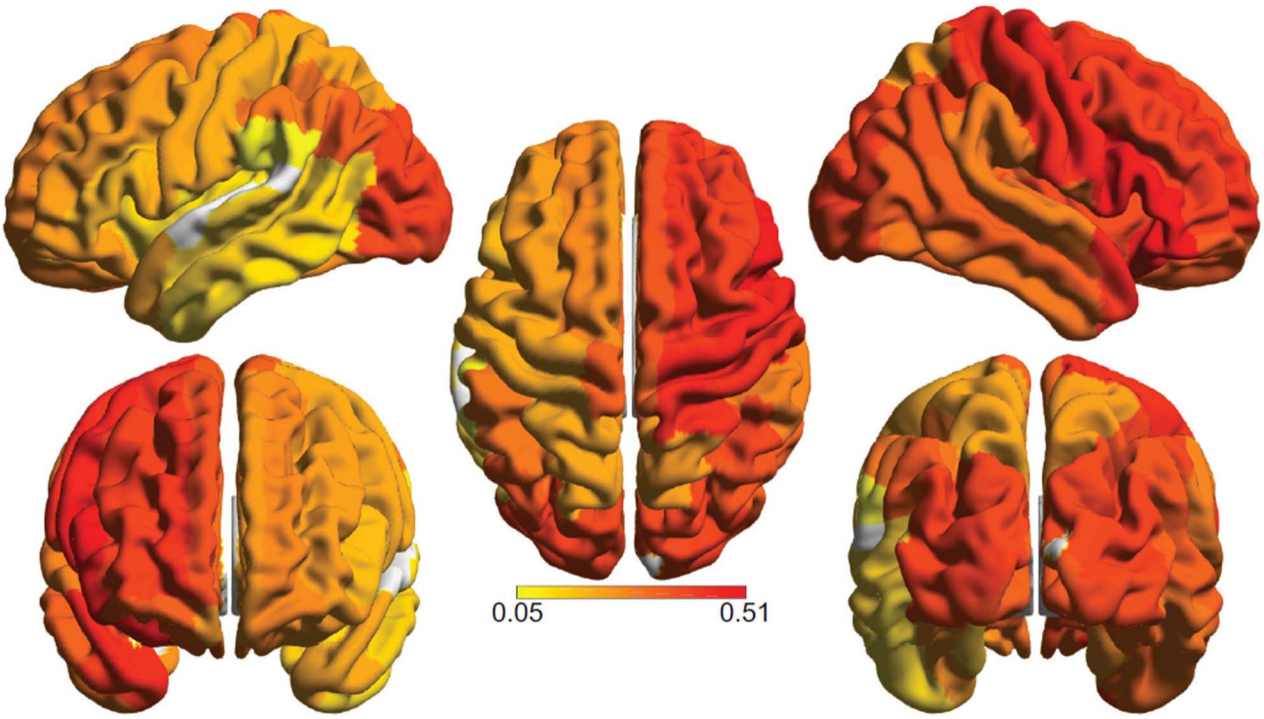

Fig. 8. The topography of the ICCs of PSD alpha at the source level.

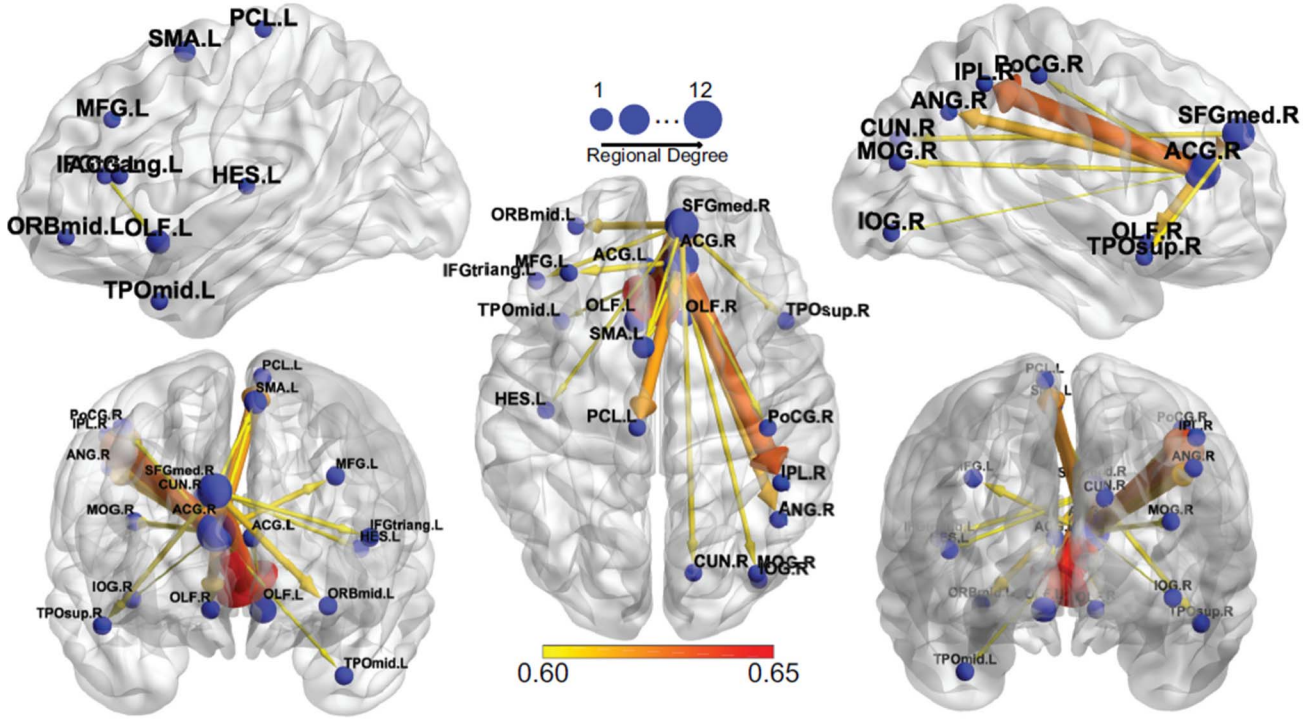

Fig. 9. The top 20 connections of PDC beta having the highest ICC values at the source level.

mean ICCs in individual connections and graph metrics. The observation agreed with a study investigating the reliabilities of PLI and weighted PLI, reporting higher global and the median of inter-regional PLI in alpha band relative to that in theta and beta bands [25]. Connectivity in alpha band was also reported as relatively more dominant and reliable compared to that in the other bands for driving fatigue assessment [12]. The dominance of PLI alpha at sensor level might suggest the high consistency of the measure in identifying fatigue.

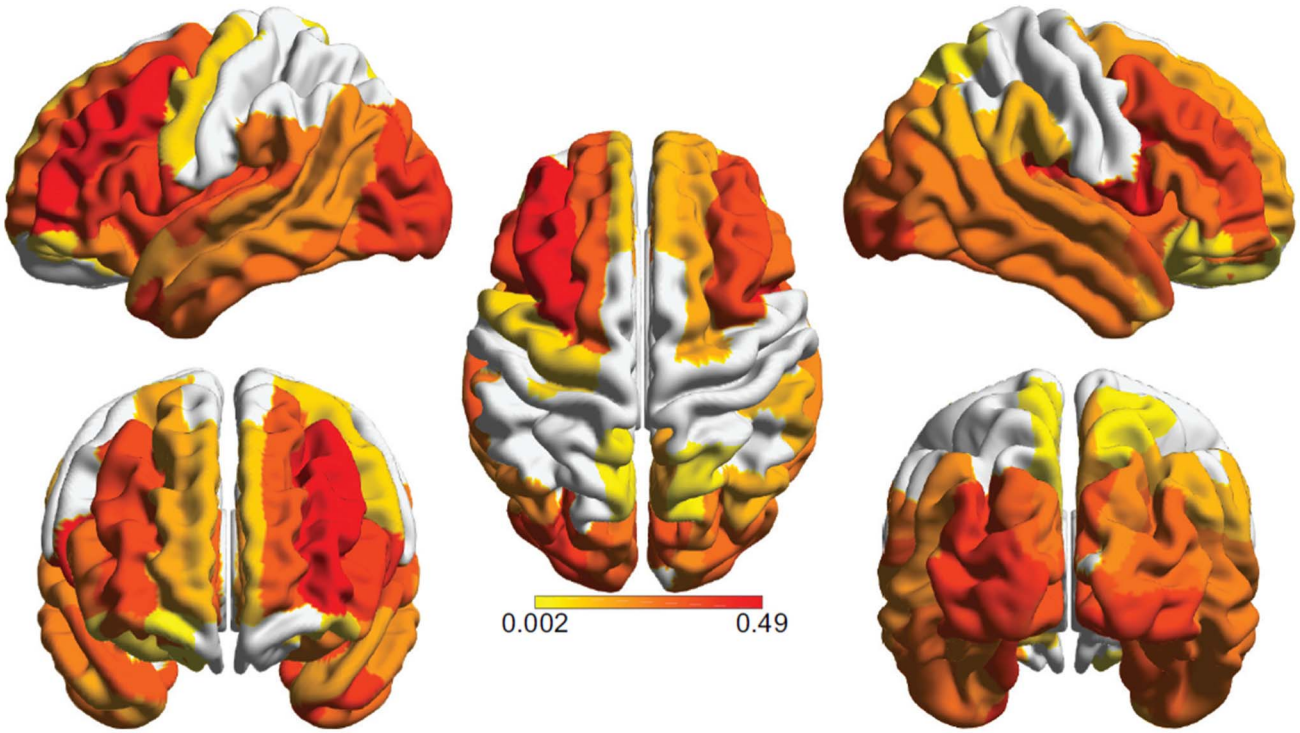

Fig. 10. The ICCs of PLI theta (nodal clustering coefficient) at the source level.

On the other hand, PDC measures had significantly higher ICCs ($p<0.05$) at source level than those at sensor level. This might signify that the PDC measures at source level are relevant to identifying fatigue. A previous study had assessed mental fatigue using PDC measures at the source level, revealing the less efficient and asymmetrical organization of the cortical connectivity [36].

In this study, the pattern of the reliability of the measures at source level might not resemble that at sensor level, particularly for distinct frequency bands. This observation might be caused by the variability induced by the source localization method [37]. In this case, the use of multimodality such as EEG-fMRI might be more beneficial for reliability estimation of driving fatigue indicators at sensor level and source level. A previous study has assessed driving fatigue using EEG-fNIRS, showing increasing alpha suppression in the occipital region and increasing HbO in the frontal cortex, supplementary motor area, primary motor cortex, and parieto-occipital cortex [38].

Based on the results of the graph metrics, we found that nodal metrics had relatively high standard deviations. On the right panel of Fig. 6, regions were observed with high contrast of high and low ICC values. Similar to individual connections, the highly varying ICCs might be caused by the complexity to compute the metrics. In this case, the reliability of nodal efficiency and clustering coefficient of a node depended on its changes relative to the changes of all of the other nodes. Synchronous changes between states with all other nodes could result in high reliability while asynchronous changes with one or more nodes could lower the reliability. The global effect of this dependency with all of the other nodes might also suggest the ICCs of the global metrics which were varying within the ICC value ranges of the corresponding nodal metrics. At sensor level, several global metrics could achieve higher ICCs relative to the mean ICCs of the single-region measures, possibly indicating the synchronous changes among all nodes. The other cause might be that particular nodes were more relevant for fatigue estimation, which will be discussed in conjunction with the regions of the other measures.

Particular regions were observed with higher ICCs compared to the other regions. In PSD theta at sensor level, higher ICCs were found in frontal and occipital regions as shown on the left panel of Fig. 6. At source level, right frontal, right temporal, right parietal and occipital regions of PSD alpha displayed higher ICCs relative to the other regions, depicted in Fig. 8. Based on the nodal metrics, regions from right frontal to right occipital showed high ICC values at sensor level for PLI alpha (nodal efficiency), shown on the right panel of Fig. 6. Frontal, temporal, and occipital regions had higher ICC values relative to the other regions for PLI theta (nodal clustering coefficient) at source level, shown in Fig. 10. Based on the results, higher ICCs were mainly found in frontal and occipital regions. In the previous studies, power increases during fatigue have been reported in occipital region [11], [39], [40], [41] and in frontal region [10], [11], [42], [39]. These regions might be more sensitive to induced fatigue, involved in the pathophysiology of chronic fatigue [43] and cognitive control [44], [45] (frontal) as well as visual processes (occipital).

For the PLI alpha at sensor level (see Fig. 7), the connections having high ICCs were observed between frontal and

central and between central and parietal/occipital regions. The previous study using transfer entropy also revealed connectivity changes around central and parietal regions during a transition state from high to low vigilance level [44]. These high ICCs of connections involving the central region might explain the importance of sensorimotor regions (central) in relation to the visual (occipital) and cognitive control (frontal) processing during fatigue [44], [45].

For PDC beta at source level, connections from right frontal to right occipital regions displayed high ICC values. In the previous study using PDC at the source level, an asymmetrical pattern of connectivity was observed where the right hemispheric connectivity was preserved during fatigue [36]. The dominance of right hemispheric activity might be associated with sustained visual attention [46], [47].

To conclude, this study presented the reliability of the proposed measure changes between alert and fatigue states. The reliability of single-region and between-region measures were computed at sensor level and source level. At both sensor level and source level, single-region measures had higher mean ICCs than the individual connections of between-region measures. Nodal metrics displayed highly varying ICCs, suggesting the dependence of a region on the changes of the other regions. The global effect of this interdependency was reflected in the ICCs of the global metrics, varying within the ICC range of the respective nodal metrics. In this study, between-region measures showed high reliability for PLI alpha (sensor level) and PDC measures (source level) while single-region measures showed high reliability for PSD at lower frequency bands (theta and alpha). All in all, the reliability of measures during driving reveals their capability of consistently identifying driving fatigue. Such reliability across sessions with a long interval is important for the selection of measures in real-time fatigue monitoring.

REFERENCES

[1] G. Zhang, K. K. Yau, X. Zhang, and Y. Li, "Traffic accidents involving fatigue driving and their extent of casualties," *Accident Anal. Prevent.*, vol. 87, pp. 34–42, Feb. 2016, doi: 10.1016/J.AAP.2015.10.033.

[2] P. Thiffault and J. Bergeron, "Monotony of road environment and driver fatigue: A simulator study," *Accident Anal. Prevention*, vol. 35, no. 3, pp. 381–391, May 2003, doi: 10.1016/S0001-4575(02)00014-3.

[3] S. K. L. Lal and A. Craig, "A critical review of the psychophysiology of driver fatigue," *Biol. Psychol.*, vol. 55, no. 3, pp. 173–194, Feb. 2001, doi: 10.1016/S0301-0511(00)00085-5.

[4] C. Papadelis *et al.*, "Monitoring sleepiness with on-board electrophysiological recordings for preventing sleep-deprived traffic accidents," *Clin. Neurophysiol.* vol. 118, no. 9, pp. 1906–1922, 2007.

[5] S. K. L. Lal and A. Craig, "Reproducibility of the spectral components of the electroencephalogram during driver fatigue," *Int. J. Psychophysiol.*, vol. 55, no. 2, pp. 137–143, Feb. 2005, doi: 10.1016/J.IJPSYCHO.2004.07.001.

[6] H. Wang, A. Dragomir, N. I. Abbasi, J. Li, N. V. Thakor, and A. Bezerianos, "A novel real-time driving fatigue detection system based on wireless dry EEG," *Cogn. Neurodyn.*, vol. 12, no. 4, pp. 365–376, Aug. 2018, doi: 10.1007/S11571-018-9481-5.

[7] J. He *et al.*, "Boosting transfer learning improves performance of driving drowsiness classification using EEG," in *Proc. Int. Workshop Pattern Recognit. Neuroimag. (PRNI)*, Jun. 2018, pp. 1–4.

[8] M. Simon *et al.*, "EEG alpha spindle measures as indicators of driver fatigue under real traffic conditions," *Clin. Neurophysiol.*, vol. 122, no. 6, pp. 1168–1178, 2011.

[9] H. J. Eoh, M. K. Chung, and S.-H. Kim, "Electroencephalographic study of drowsiness in simulated driving with sleep deprivation," *Int. J. Ind. Ergonom.*, vol. 35, no. 4, pp. 307–320, Apr. 2005.

[10] B. T. Jap, S. Lal, P. Fischer, and E. Bekiaris, "Using EEG spectral components to assess algorithms for detecting fatigue," *Expert Syst. Appl.*, vol. 36, no. 2, pp. 2352–2359, 2009.

[11] C. Zhao, M. Zhao, J. Liu, and C. Zheng, "Electroencephalogram and electrocardiograph assessment of mental fatigue in a driving simulator," *Accident Anal. Prevention*, vol. 45, pp. 83–90, Mar. 2012.

[12] J. Harvy, N. Thakor, A. Bezerianos, and J. Li, "Between-frequency topographical and dynamic high-order functional connectivity for driving drowsiness assessment," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 3, pp. 358–367, Jan. 2019.

[13] J. Harvy, E. Sigalas, N. Thakor, A. Bezerianos, and J. Li, "Performance improvement of driving fatigue identification based on power spectra and connectivity using feature level and decision level fusions," in *Proc. 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2018, pp. 102–105.

[14] R. Zhang *et al.*, "Efficient resting-state EEG network facilitates motor imagery performance," *J. Neural Eng.*, vol. 12, no. 6, Dec. 2015, Art. no. 066024, doi: 10.1088/1741-2560/12/6/066024.

[15] T. Wang, A. Bezerianos, A. Cichocki, and J. Li, "Multikernel capsule network for schizophrenia identification," *IEEE Trans. Cybern.*, vol. 52, no. 6, pp. 4741–4750, Jun. 2022, doi: 10.1109/TCYB.2020.3035282.

[16] J. Li *et al.*, "Mid-task break improves global integration of functional connectivity in lower alpha band," *Frontiers Hum. Neurosci.*, vol. 10, p. 304, Jun. 2016.

[17] W. Kong, Z. Zhou, B. Jiang, F. Babiloni, and G. Borghini, "Assessment of driving fatigue based on intra/inter-region phase synchronization," *Neurocomputing*, vol. 219, pp. 474–482, Jan. 2017.

[18] B. T. Jap, S. Lal, and P. Fischer, "Inter-hemispheric electroencephalography coherence analysis: Assessing brain activity during monotonous driving," *Int. J. Psychophysiol.*, vol. 76, no. 3, pp. 169–173, Jun. 2010.

[19] C. Zhao, M. Zhao, Y. Yang, J. Gao, N. Rao, and P. Lin, "The reorganization of human brain networks modulated by driving mental fatigue," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 3, pp. 743–755, Mar. 2016.

[20] J. Chen, H. Wang, C. Hua, Q. Wang, and C. Liu, "Graph analysis of functional brain network topology using minimum spanning tree in driver drowsiness," *Cogn. Neurodyn.*, vol. 12, no. 6, pp. 569–581, Dec. 2018.

[21] G. N. Dimitrakopoulos *et al.*, "Driving mental fatigue classification based on brain functional connectivity," in *Proc. Int. Conf. Eng. Appl. Neural Netw.*, 2017, pp. 465–474.

[22] B. Moezzi, B. Hordacre, C. Berryman, M. C. Ridding, and M. R. Goldsworthy, "Test-retest reliability of functional brain network characteristics using resting-state EEG and graph theory," *BioRxiv*, Aug. 2018, Art. no. 385302.

[23] U. Braun *et al.*, "Test–retest reliability of resting-state connectivity network characteristics using fMRI and graph theoretical measures," *NeuroImage*, vol. 59, no. 2, pp. 1404–1412, 2012.

[24] G. L. Colclough, M. W. Woolrich, P. K. Tewarie, M. J. Brookes, A. J. Quinn, and S. M. Smith, "How reliable are MEG resting-state connectivity metrics?" *NeuroImage*, vol. 138, pp. 284–293, Sep. 2016.

[25] M. Hardmeier, F. Hatz, H. Bousleiman, C. Schindler, C. J. Stam, and P. Fuhr, "Reproducibility of functional connectivity and graph measures based on the phase lag index (PLI) and weighted phase lag index (wPLI) derived from high resolution EEG," *PLoS ONE*, vol. 9, no. 10, Oct. 2014, Art. no. e108648.

[26] L. Deuker *et al.*, "Reproducibility of graph metrics of human brain functional networks," *NeuroImage*, vol. 47, no. 4, pp. 1460–1468, 2009.

[27] L. Dong *et al.*, "MATLAB toolboxes for reference electrode standardization technique (REST) of scalp EEG," *Frontiers Neurosci.*, vol. 11, pp. 1–8, Oct. 2017, doi: 10.3389/FNINS.2017.00601.

[28] D. Yao, "A method to standardize a reference of scalp EEG recordings to a point at infinity," *Physiol. Meas.*, vol. 22, no. 4, pp. 693–711, 2001, doi: 10.1088/0967-3334/22/4/305.

[29] A. Delorme and S. Makeig, "EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *J. Neurosci. Methods*, vol. 134, no. 1, pp. 9–21, Mar. 2004.

[30] R. D. Pascual-Marqui, "Discrete, 3D distributed, linear imaging methods of electric neuronal activity. Part 1: Exact, zero error localization," 2007, *arXiv:0710.3341*.

[31] M. Hata *et al.*, "Functional connectivity assessed by resting state EEG correlates with cognitive decline of Alzheimer's disease—An eLORETA study," *Clin. Neurophysiol.*, vol. 127, no. 2, pp. 1269–1278, Feb. 2016.

[32] J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," *Amer. J. Physiol.-Heart Circulatory Physiol.*, vol. 278, no. 6, pp. H2039–H2049, Jun. 2000.

[33] C. J. Stam, G. Nolte, and A. Daffertshofer, "Phase lag index: Assessment of functional connectivity from multi channel EEG and MEG with diminished bias from common sources," *Hum. Brain Mapping*, vol. 28, no. 11, pp. 1178–1193, Nov. 2007.

[34] L. A. Baccalá and K. Sameshima, "Partial directed coherence: A new concept in neural structure determination," *Biol. Cybern.*, vol. 84, no. 6, pp. 463–474, May 2001.

[35] J. J. Bartko, "The intraclass correlation coefficient as a measure of reliability," *Psychol. Rep.*, vol. 19, no. 1, pp. 3–11, 1966.

[36] Y. Sun, J. Lim, K. Kwok, and A. Bezerianos, "Functional cortical connectivity analysis of mental fatigue unmasks hemispheric asymmetry and changes in small-world networks," *Brain Cogn.*, vol. 85, pp. 220–230, Mar. 2014.

[37] K. Mahjoory, V. V. Nikulin, L. Botrel, K. Linkenkaer-Hansen, M. M. Fato, and S. Haufe, "Consistency of EEG source localization and connectivity estimates," *NeuroImage*, vol. 152, pp. 590–601, May 2017.

[38] C.-H. Chuang, Z. Cao, J.-T. King, B.-S. Wu, Y.-K. Wang, and C.-T. Lin, "Brain electrodynamic and hemodynamic signatures against fatigue during driving," *Frontiers Neurosci.*, vol. 12, p. 181, Mar. 2018.

[39] A. M. Strijkstra, D. G. M. Beersma, B. Drayer, N. Halbesma, and S. Daan, "Subjective sleepiness correlates negatively with global alpha (8–12 Hz) and positively with central frontal theta (4–8 Hz) frequencies in the human resting awake electroencephalogram," *Neurosci. Lett.*, vol. 340, no. 1, pp. 17–20, 2003.

[40] R.-S. Huang, T.-P. Jung, and S. Makeig, "Tonic changes in EEG power spectra during simulated driving," in *Proc. Int. Conf. Found. Augmented Cognition*, 2009, pp. 394–403.

[41] C.-T. Lin *et al.*, "Tonic and phasic EEG and behavioral changes induced by arousing feedback," *NeuroImage*, vol. 52, no. 2, pp. 633–642, Aug. 2010.

[42] A. Craig, Y. Tran, N. Wijesuriya, and H. T. Nguyen, "Regional brain wave activity changes associated with fatigue," *Psychophysiology*, vol. 49, no. 4, pp. 574–582, 2012, doi: 10.1111/J.1469-8986.2011.01329.X.

[43] D. B. Cook, P. J. O'Connor, G. Lange, and J. Steffener, "Functional neuroimaging correlates of mental fatigue induced by cognition among chronic fatigue syndrome patients and controls," *NeuroImage*, vol. 36, no. 1, pp. 108–122, May 2007, doi: 10.1016/J.NEUROIMAGE.2007.02.033.

[44] C.-S. Huang, N. R. Pal, C.-H. Chuang, and C.-T. Lin, "Identifying changes in EEG information transfer during drowsy driving by transfer entropy," *Frontiers Hum. Neurosci.*, vol. 9, p. 570, Oct. 2015.

[45] N. Lawrence, T. Ross, R. Hoffmann, H. Garavan, and E. Stein, "Multiple neuronal networks mediate sustained attention," *J. Cognit. Neurosci.*, vol. 15, no. 7, pp. 1028–1038, Oct. 2003, doi: 10.1162/089892903770007416.

[46] R. Whitehead, "Right hemisphere processing superiority during sustained visual attention," *J. Cognit. Neurosci.*, vol. 3, no. 4, pp. 329–334, Oct. 1991, doi: 10.1162/JOCN.1991.3.4.329.

[47] M. Corbetta and G. L. Shulman, "Control of goal-directed and stimulus-driven attention in the brain," *Nature Rev. Neurosci.*, vol. 3, no. 3, pp. 201–215, 2002, doi: 10.1038/NRN755.