A Sticky Weighted Regression Model for Time-Varying Resting State Brain Connectivity Estimation

Aiping Liu, Xun Chen*, Martin J. McKeown and Z. Jane Wang, Senior Member, IEEE.

Abstract—Despite recent progress on brain connectivity modeling using neuroimaging data such as fMRI, most current approaches assume that brain connectivity networks have time-invariant topology/coefficients. This is clearly problematic as the brain is inherently non-stationary. Here we present a time-varying model to investigate the temporal dynamics of brain connectivity networks. The proposed method allows for abrupt changes in network structure via a fused LASSO scheme, as well as recovery of time-varying networks with smoothly changing coefficients via a weighted regression technique. Simulations demonstrate that the proposed method yields improved accuracy on estimating time-dependent connectivity patterns when compared to a static sparse regression model or a weighted time-varying regression model. When applied to real resting state fMRI data sets from Parkinson’s disease (PD) and control subjects, significantly different temporal and spatial patterns were found to be associated with PD. Specifically, PD subjects demonstrated reduced network variability over time, which may be related to impaired cognitive flexibility previously reported in PD. The temporal dynamic properties of brain connectivity in PD subjects may provide insights into brain dynamics associated with PD and may serve as a potential biomarker in future studies.

Index Terms—fMRI, Brain connectivity network, Dynamic, Time varying, Parkinson’s Disease

I. INTRODUCTION

In addition to examining relative changes in blood-oxygen-level-dependent (BOLD) signals as a result of ongoing brain activity, brain connectivity patterns reveal cooperation between different brain regions. Inferring brain connectivity networks from functional magnetic resonance imaging (fMRI) data has been increasingly important for understanding brain functioning both normally, and in disease states.

Many mathematical methods have been developed for brain connectivity modeling. Correlation, mutual information and coherence are among the most straightforward bivariate analysis approaches [1]. Linear decomposition methods such as Principal Component Analysis (PCA) and Independent Component Analysis (ICA) have also been widely employed to explore co-activations among brain regions [2]. Structural Equation Modeling (SEM) [3], Dynamic Causal Models [4] and regression based approaches [5] are popular graphical models which are suitable for modeling brain connectivity networks because their graphical nature assists in the biomedical interpretation of the learned connectivity patterns. To make inference of brain connectivity networks in a sample or even a population, several group level approaches have been proposed, such as Bayesian model selection [6], Group Covariance estimation [7], group $PC_{fd}$ algorithm [8] and the prior information guided group network modeling approach [9].

However, most current approaches assume that the connectivity structure is time-invariant, i.e., without considering temporal variations of the underlying neural activity, and thus the inferred brain connectivity possibly a temporally averaged connectivity pattern [10]. Assessing the temporal dynamics of connectivity patterns may therefore represent an additional dimension to assess brain activity [11].

Several strategies have thus far been proposed to investigate brain connectivity dynamics. Lagged interaction based approaches, such as dynamic Bayesian network modeling [12] and auto-regressive (AR) models [13], examine brain interactions simultaneously and over adjacent time steps. Since these methods assume the lagged interactions themselves are time-invariant, they technically can be considered as stationary mod-
els. State space model based approaches, by combining lagged interaction and filtering theory, estimate non-stationary brain connectivity at each time point [14]. In addition, time-frequency based approaches, such as wavelet transform based coherence analysis, infer resting state dynamic brain connectivity from places in the time-frequency plane [15]. Wavelet based time varying Granger causality analysis has also been used to produce evolving brain connectivity maps that are modulated by task performance [16].

If brain connectivity networks can be assumed to change slowly and smoothly over time, a sliding window approach is appropriate. By specifying a fixed window length and shifting the window by a given number of data samples, different network learning methods such as correlation [17], [18], [19], covariance [20] and ICA [21], [22] have been applied to estimate the time dependent interactions within each window. However, determining the appropriate window length is critical and difficult: with too small a window length, estimated connectivity patterns suffer from large fluctuations due to noise and thus may not truly reflect the underlying dynamic properties of brain activity; in contrast, too large a window will result in insensitivity to possibly important brain state changes. In order to avoid the assumption that changes occur slowly over time, several studies have reported that functional networks inferred by stationary approaches may be unduly influenced by changes at a few critical time points [10]. These critical time points can be used to segment the entire signal into quasi-stationary sections for the purposes of brain connectivity estimation [10], [23], [24]. Nevertheless, change point detection may be particularly susceptible to artifacts (e.g. due to head motion) in the data. Another important characteristic of both sliding window and change point detection multivariate models is that they assume that different brain regions have temporal variations at the same time scale so that the entire brain dynamics are assumed to switch simultaneously, while in practice, different pairs of brain regions may interact at different temporal scales [11].

Beyond the specific area of fMRI brain connectivity modeling, several time-varying frameworks have been proposed to discover multivariate interactions over time. These include time varying regularized graphical structural estimation [25], linear regression-based Bayesian Network (BN) approaches [26], [27], [28] and change point detection approaches [29]. In a BN framework, both network structure and parameter changes are treated as random processes whose values at each time epoch are modelled via a BN approach. In one change point detection model, a fused penalty used in a preliminary linear regression model is used to detect change points, and then multivariate regression is separately applied to each segment [29]. Another change point detection approach uses penalized regression and Gaussian mixture models [30].

Based on the above discussion, it is clear that modeling brain dynamics often requires fairly strict assumptions be made, such as assuming the networks change smoothly, change suddenly, or change in a piece-wise stationary fashion [11]. Moreover, learning dynamic changes in brain interactions may be complicated by factors such as head movement, measurement noise and other randomized fluctuations. To reduce the influence of random noise, it is reasonable to assume that brain connectivity patterns mostly change smoothly except at critical change points. Based on this assumption, temporally adjacent networks are more likely to share common patterns than temporally distant networks, as assumed in a weighted time-varying regression model [26], yet abrupt changes can still occur at specific change points.

Therefore, in this paper, we propose a sticky weighted time-varying (SWTV) model that estimates the non-stationary process of brain interactions in a temporally penalized, weighted regression fashion. We incorporate a fused penalty [31] into the weighted regression model [26]. The fused penalty is added into a weighted regression model in which we can estimate both smooth changing coefficients and abrupt changing structures so that the change point detection problem will not be separated from the network estimation problem. More importantly, the proposed method allows different pairs of brain regions to exhibit fluctuations at different time scales, as illustrated in simulations in Section III. Finally, we assume connections between spatially disparate brain regions are relatively sparse to facilitate meaningful biological interpretation.

In the remainder of the paper, we will introduce the SWTV model in Section II and perform simulations to validate the proposed method in Section III. Also, in Section IV the proposed SWTV model is applied to resting state fMRI data sets from both Parkinson’s Disease (PD) and control subjects, and significant different temporal and spatial patterns are found to be associated with the disease state.

II. METHODS

In this section, we will briefly introduce regression models used in the brain connectivity network modeling at the single subject level, and then we will present the proposed sticky weighted time varying model to estimate
dynamic interactions between different brain regions in the resting state.

A. Sticky Weighted Time Varying (SWTV) Model

Multivariate linear regression models have been widely used to infer neural interactions. For instance, Structural Equation Modeling (SEM) estimates the brain interactions at zero lag \cite{3}. The multivariate autoregressive (MAR) model focuses on the lagged interactions between different brain ROIs \cite{32}. With a sparsity assumption, LASSO (Least Absolute Shrinkage and Selection Operator), group LASSO and elastic net models try to discover the sparse map of the brain interactions with high computational efficiency \cite{9,33}. In the regression model used in our paper, the fMRI time course of a Region of Interest (ROI) is regarded as the response variable, and is predicted from the time courses of all other ROIs at zero-lag as,

\[ Y = X\beta + e, \]

(1)

where vector \( Y \) with length \( T \) means the time course of one brain ROI, \( X \) is \( T \times K \) predictor matrix based on the time courses of all other ROIs with \( K + 1 \) representing the entire number of ROIs, \( \beta \) is the coefficient vector and \( e \) is the Gaussian noise term. Due to the non-stationary nature of the brain activity, the time dependent regression model becomes,

\[ Y_t = X_t\beta^t + e, \]

(2)

where \( t \) represents the time index and we need to estimate the regression coefficient vector at each time point respectively. \( Y_t \) is the response sample at time point \( t \) and \( X_t \) is the \( t \)-th sample row in the predictor matrix. In order to make the connections sparse, one could use an \( l_1 \) penalty on the regression coefficients. However, with only one sample point, the estimator of coefficients would be extremely unstable. Thus in order to estimate time-varying structures/coefficients, yet still allow sparsity, we assume that the underlying networks are changing smoothly over time. Following prior work \cite{26}, we could estimate the coefficients for each time point separately in a weighted time varying model (WTM) as,

\[
\hat{\beta}^t = \arg \min_{\beta^{t^*}} \sum_{t=1}^{T} W^{t^*}(t) (Y_t - X_t\beta^{t^*})^2 + \lambda \| \beta^{t^*} \|_{l_1},
\]

(3)

where \( \beta^{t^*} \) is the coefficient vector we need to estimate at time \( t^* \), \( \hat{\beta}^{t^*} \) is the estimator of \( \beta^{t^*} \), and \( \lambda \) is the parameter for the \( l_1 \) penalty. \( W^{t^*}(t) \) is the weighting of observations from time \( t \) when we estimate the coefficients at time \( t^* \). In general, \( W^{t^*}(t) \) can be defined as any normalized kernel function. In this paper, \( W^{t^*}(t) \) is defined as,

\[
W^{t^*}(t) = \frac{\exp(-((t-t^*)^2/h))}{\sum_{t=1}^{T} \exp(-((t-t^*)^2/h))}.
\]

(4)

This is a normalized Gaussian Radial basis function (RBF) kernel, with \( h \) representing the kernel band-width. Note that this model is essentially a sparse weighted regression model that allows us to estimate the coefficients at each time point separately by reweighting the observations. With the smoothly changing assumption, temporally adjacent coefficients are more likely to be similar than temporally distant coefficients.

A “sticky” weighted time varying model is therefore introduced as,

\[
\minimize_{\beta^{t^*},t^*\in R^T} \sum_{t^*=1}^{T} \sum_{t=1}^{T} W^{t^*}(t) (Y_t - X_t\beta^{t^*})^2 + \lambda \sum_{t^*=1}^{T} \| \beta^{t^*} \|_{l_1} + \gamma \sum_{t^*=2}^{T} \| \beta^{t^*} - \beta^{t^*-1} \|_{l_1},
\]

(5)

where \( \gamma \) is the parameter to control the fused penalty and serves to keep the coefficients temporally consistent except at (possibly several) abrupt change points.

To efficiently solve this optimization problem, we can rewrite the response vector and predictor matrix as \( Y_t^{t^*} = \sqrt{W^{t^*}(t)}Y_t \) and \( X_t^{t^*} = \sqrt{W^{t^*}(t)}X_t \). Let \( Y^{t^*} = (Y_1^{t^*},Y_2^{t^*},\ldots,Y_T^{t^*})' \), \( X^{t^*} = (X_1^{t^*},X_2^{t^*},\ldots,X_T^{t^*})' \), then the weights can be incorporated into the square loss function directly. We can further simplify the objective function by expressing them in a matrix format. Suppose \( \tilde{Y} = (Y^1,Y^2,\ldots,Y^T)' \) is a response vector with length \( TT \), \( \tilde{X} = \text{diag}(X^1,X^2,\ldots,X^T) \) is a block diagonal matrix with dimension \( TT \times TK \), and \( \tilde{\beta} = (\beta^1,\beta^2,\ldots,\beta^T)' \) is the concatenated time varying coefficient vector with length \( TK \). Each \( \beta^t \) corresponds to the coefficient at time point \( t \). The objective function can be formulated as,

\[
\minimize_{\tilde{\beta}} \| \tilde{Y} - \tilde{X}\tilde{\beta} \|_F^2 + \lambda \| \tilde{\beta} \|_{l_1} + \gamma \| C\tilde{\beta} \|_{l_1}.
\]

(6)

where \( C \) is a sparse \((T-1)K \times TK \) matrix with two nonzero elements in each row. More specifically, \( C((T-1)*(i-1) + j,K*(j-1)+i) = -1 \) and \( C((T-1)*(i-1) + j,K*j+i) = 1, i = 1,2,\ldots,K, j = 1,2,\ldots,T-1 \). By this formulation, it becomes a generalized fused LASSO problem which can be
solved using the smoothing proximal gradient (SPG) method [34]. SPG is one efficient algorithm with the convergence rate of \(O(\frac{1}{\epsilon})\) where \(\epsilon\) is the precision of the algorithm; Per-iteration complexity of SPG is linear with the number of nonzero elements of the constructed sparse network \(C\) [34]. In this paper, we use the SPG optimization toolbox [34], and the implementation of the SWTV model is described in Table. I.

B. Model Selection

The parameters of stationary regression model are usually determined by cross validation (CV) which separates the data into training and testing sets. However, the standard CV approach cannot be employed directly in our time-varying case, since each sample corresponds to a specific time point and the structures and coefficients may be different across time.

Therefore, in order to apply cross validation, we first up-sample the data by a factor of two: the odd samples represent the original data points and even samples are the interpolated data points. For the purpose of model selection, we assume that the corresponding even samples have the same temporal properties as those of odd samples. In the following simulation studies, by treating the odd samples as the training set and even samples as the testing set, we can select optimal parameters of the model.

For each fixed set of parameters \(\lambda, \gamma\) and the band width \(h\), we can estimate time-varying coefficients as described in Table. I and use the cross validation to select the optimal values. However, for large scale data sets, cross validation can be time consuming, and it may not be feasible for the large scale problems. Instead, the gradient descent approach can be applied to iteratively update each parameter as described in [35], [36]. It sequentially applies three line searches along each descent direction to minimize the corresponding mean square error of current CV until the error convergences. For the static sparse regression model used in the simulation part, we utilize the stability selection approach which is proved to enhance selection accuracy [37].

C. Brain Connectivity Network Modeling and Statistical Analysis

To perform inference of brain connectivity networks, we utilize a linear regression approach. We treat each ROI in turn as the response vector and all other ROIs as constituting the predictor matrix. This corresponding coefficient vector would give the strength of connectivity from all other ROIs to the response ROI which are the estimated directed connections. In this way, the time varying coefficient vector for each ROI is estimated one by one until we obtain the whole brain network.

To quantify and compare the temporal variability of the inferred networks, we define the network variation as,

\[
V = \frac{1}{T-1} \sum_{t=2}^{T} ||G(t) - G(t-1)||^2_F,
\]

where \(t\) represents the time index and \(G(t)\) represents the brain connectivity network which is a matrix estimated using SWTV model as described before at time point \(t\). Network variation calculates the average of distance between two brain connectivity networks at adjacent time points. It quantifies the changes of network structures as well as connectivity strengths. This metric measures the ability of switching or oscillation of the networks across the time. While a relatively simple measure, we found \(V\) an intuitive way to quantify and compare the temporal changes of brain connectivity networks: with a fixed \(\gamma\), a higher \(V\) implies higher moment-to-moment variability in the networks (Fig. 4).

III. Simulation

To validate the proposed method, we performed simulations to compare the performance of SWTV model with both that of the weighted time varying model and a static sparse regression model. We considered different simulation settings where the different variables changed in the same time scale with and without autocorrelation, and also when they changed in different time scales.

In brief, the simulated data were generated from a Gaussian model with changing structures and coefficients as \(Y_t = X_t \beta^t + e_t\). \(X^t\) was a randomly generated sample row at time point \(t\) with \(K\) variables (i.e., a \(1 \times K\) row vector), \(\beta^t\) was a time dependent coefficient vector with same length \(K\) (\(K = 20\)) and \(e_t\) was white Gaussian noise.

More specifically:

1) We first generated the changing coefficients \(\beta^t\). In the first and second simulations, we assumed that all the variables changed in the same time scale as \(N\) and the total length of sample size was \(T = 3 \times N\) (an example of this is shown in Fig. 1(a)). In the third simulation, different coefficients could have different time scales as shown in Fig. 2(a). The averaged time scale and sample size were set to \(N\) and \(T = 3 \times N\) respectively.

2) The design matrix \(X\) was randomly generated containing \(T\) observations and \(K\) predictors. The error vector \(e\) was Gaussian noise ~ \(N(0, 1)\). The response vector \(Y\) was generated by \(Y^t = X^t \beta^t + e_t\) with \(t = 1, \ldots, T\). In the second simulation, to generate the
TABLE I: Implementation of Sticky Weighted Time Varying Model

| Step 1: Weighting the response vector \( Y \) and the predictor matrix \( X \). |
| 1. Based on \( h \), calculate \( W_{i'}(t) \) according to Eq. (4) for \( t = 1, 2, \ldots, T \). |
| 2. \( Y_{i'}^t \leftarrow \sqrt{W_{i'}(t)}Y_i \), \( t = 1, 2, \ldots, T \), and \( Y^t = (Y_{i'}^1, Y_{i'}^2, \ldots, Y_{i'}^T) \). |
| 3. \( X_i^t \leftarrow \sqrt{W_{i'}(t)}X_i \), \( t = 1, 2, \ldots, T \), and \( X^t = (X_{i'}^1, X_{i'}^2, \ldots, X_{i'}^T) \). |

| Step 2: Constructing the objective function. |
| 1. \( \hat{Y} \leftarrow (Y^1, Y^2, \ldots, Y^T) \). |
| 2. \( \hat{X} \leftarrow \text{diag}(X^1, X^2, \ldots, X^T) \). |
| 3. Construct a sparse matrix \( C \in R^{(T-1)K \times TK} \) as, |
| \( C((j-1) + j, K * (j - 1) + i) \leftarrow -1 \), \( i = 1, 2, \ldots, K \), \( j = 1, 2, \ldots, T - 1 \). |
| 4. Formulate the objective function as in Eq. (6) by inputting \( \hat{Y}, \hat{X}, \hat{\beta}, C, \lambda, \gamma \). |

| Step 3: Estimating time-varying coefficients \( \hat{\beta} = (\beta^1, \beta^2, \ldots, \beta^T)' \). |
| Apply SPG toolbox \([34]\) to solve the optimization problem in Eq. (6) and obtain the estimated time varying coefficients \( \hat{\beta} \). |

Output: Time varying coefficients \( \hat{\beta} = (\beta^1, \beta^2, \ldots, \beta^T)' \).
Fig. 1: Results for the first and second simulations: (a) The true Model. (b)-(d), models learned by static LASSO, Weighted Time Varying, and Sticky Weighted Time Varying models respectively in the first simulation. (e)-(g), models learned by static LASSO, Weighted Time Varying, and Sticky Weighted Time Varying models respectively in the second simulation. The time index is along the x-axis, the variable index is along the y-axis and the color bar represents the coefficients’ strength.

then compare the temporal and spatial patterns of the inferred connectivity networks.

A. Subjects and fMRI resting state data set

Twelve PD subjects and ten healthy control subjects were recruited from Pacific Parkinson’s Research Center (PPRC) at the University of British Columbia (UBC). All the experiments were approved by the Ethics Board at UBC, and all the subjects provided informed consent prior to experiment participation.

A 3 Tesla scanner (Philips Gyroscan Intera 3.0T; Philips Medical Systems, Netherlands) equipped with a head-coil was used to collect data in the resting state. Before scanning, all the subjects were instructed to lie on their back in the scanner and have several minutes to acclimatize themselves to the scanner environment with eyes closed. Blood oxygenation level-dependent (BOLD) contrast echo-planar (EPI) T2*-weighted images were taken with the following specifications with a repetition time of 1985 ms, echo time of 37 ms, flip angle 90°, field of view (FOV) 240.00 mm, matrix size 128×128, with pixel size 1.9 mm×1.9 mm. The duration of each functional run was 4 mins during which we obtained 36 axial slices with 3 mm thickness and 1 mm gap thickness. The FOV was set to include the cerebellum
Fig. 2: Results for the third simulation: (a) The true Model. (b)-(d), models learned by static LASSO, Weighted Time Varying, and Sticky Weighted Time Varying models respectively. The time index is along the x-axis, the variable index is along the y-axis and the color bar represents the coefficients’ strength.

Fig. 3: Simulation Results. (a) F1 scores of the first simulation. (b) F1 scores of the second simulation. (c) F1 scores of the third simulation. Red lines represent the F1 scores of the proposed method, blue lines represent the F1 scores of weighted time varying model and the green lines represent the F1 scores of the static model.

B. Results

To apply the proposed method on the subjects with PD and control subjects, we need to choose the band width $h$, the sparse penalty parameter $\lambda$ and the fused penalty parameter $\gamma$ in the SWTV model. We conducted parameter selection using gradient descent approach for each subject. The optimal parameters for each subject varied across a broad range of values ($h = 38.9091 \pm 26.1769$, $\lambda = 0.5117 \pm 0.4382$, $\gamma = 1.5341 \pm 0.8308$). Using the optimal values, the density of learned connectivity networks varied from 0.0294 to 0.3326, making it difficult to compare two groups. To fairly compare the connectivity patterns of patient and control groups, we chose fixed parameters/densities for all the subjects in ventrally and also include the dorsal surface of the brain. 48 Freesurfer-derived ROIs in total were chosen in this study as shown in Table. II
TABLE II: The index and name of 48 selected brain ROIs. ‘L’ represents the brain left side and ‘R’ represents the brain right side.

<table>
<thead>
<tr>
<th>Index</th>
<th>Name</th>
<th>Index</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1</td>
<td>L-Cerebellum</td>
<td>R1</td>
<td>R-Cerebellum</td>
</tr>
<tr>
<td>L2</td>
<td>L-PMd</td>
<td>R2</td>
<td>R-PMd</td>
</tr>
<tr>
<td>L3</td>
<td>L-PMv</td>
<td>R3</td>
<td>R-PMv</td>
</tr>
<tr>
<td>L4</td>
<td>L-Pre-SMA</td>
<td>R4</td>
<td>R-Pre-SMA</td>
</tr>
<tr>
<td>L5</td>
<td>L-SMA-proper</td>
<td>R5</td>
<td>R-SMA-proper</td>
</tr>
<tr>
<td>L6</td>
<td>L-ACC</td>
<td>R6</td>
<td>R-ACC</td>
</tr>
<tr>
<td>L7</td>
<td>L-Caudate</td>
<td>R7</td>
<td>R-Caudate</td>
</tr>
<tr>
<td>L8</td>
<td>L-Cerebellum-Cortex</td>
<td>R8</td>
<td>R-Cerebellum-Cortex</td>
</tr>
<tr>
<td>L9</td>
<td>L-PFC</td>
<td>R9</td>
<td>R-PFC</td>
</tr>
<tr>
<td>L10</td>
<td>L-Pallidium</td>
<td>R10</td>
<td>R-Pallidium</td>
</tr>
<tr>
<td>L11</td>
<td>L-Putamen</td>
<td>R11</td>
<td>R-Putamen</td>
</tr>
<tr>
<td>L12</td>
<td>L-Somatosensory</td>
<td>R12</td>
<td>R-Somatosensory</td>
</tr>
<tr>
<td>L13</td>
<td>L-Thalamus-Proper</td>
<td>R13</td>
<td>R-Thalamus-Proper</td>
</tr>
<tr>
<td>L14</td>
<td>ctx-L-caudaldmed-middle-frontal</td>
<td>R14</td>
<td>ctx-R-caudaldmed-middle-frontal</td>
</tr>
<tr>
<td>L15</td>
<td>ctx-L-cuneus</td>
<td>R15</td>
<td>ctx-R-cuneus</td>
</tr>
<tr>
<td>L16</td>
<td>ctx-L-inferiorparietal</td>
<td>R16</td>
<td>ctx-R-inferiorparietal</td>
</tr>
<tr>
<td>L17</td>
<td>ctx-L-inferiortemporal</td>
<td>R17</td>
<td>ctx-R-inferiortemporal</td>
</tr>
<tr>
<td>L18</td>
<td>ctx-L-lateraloccipital</td>
<td>R18</td>
<td>ctx-R-lateraloccipital</td>
</tr>
<tr>
<td>L19</td>
<td>ctx-L-middletemporal</td>
<td>R19</td>
<td>ctx-R-middletemporal</td>
</tr>
<tr>
<td>L20</td>
<td>ctx-L-precentral</td>
<td>R20</td>
<td>ctx-R-precentral</td>
</tr>
<tr>
<td>L21</td>
<td>ctx-L-precuneus</td>
<td>R21</td>
<td>ctx-R-precuneus</td>
</tr>
<tr>
<td>L22</td>
<td>ctx-L-superiorparietal</td>
<td>R22</td>
<td>ctx-R-superiorparietal</td>
</tr>
<tr>
<td>L23</td>
<td>ctx-L-superiortemporal</td>
<td>R23</td>
<td>ctx-R-superiortemporal</td>
</tr>
<tr>
<td>L24</td>
<td>ctx-L-supramarginal</td>
<td>R24</td>
<td>ctx-R-supramarginal</td>
</tr>
</tbody>
</table>

Fig. 4: (a) The averaged number of detected connections within the networks as sparsity parameter $\lambda$ increases, with a fixed fused penalty parameter $\gamma = 0.5$. The blue line represents the averaged number of detected connections as a function of $\lambda$. The red dashed line represents the averaged number of common connections with network detected by a smaller $\lambda$. E.g., the first point of the red line is the number of averaged common connections between networks with $\lambda = 0.1$ and $\lambda = 0$. (b) The network variations of networks as fused penalty parameter $\gamma$ increases, with a fixed $\lambda = 0.5$.

Although a few studies have been conducted on time variation in connectivity networks, the exact time scale of brain activities is unclear, and varies between subjects.

This is especially true in resting state studies where subjects are asked to lie quietly and not think of anything in particular, so the exact temporal patterns of brain activity may vary across the population. Similar to the choice of sliding window length, a small bandwidth $h$ will suffer from large fluctuations while a large band width $h$ may reduce sensitivity to fluctuations in the signal. Following prior work [18] as well as our preliminary studies, we set the bandwidth to 32s (16 points). A comprehensive comparisons of brain connectivity variation scales will be conducted in future work.

Fig. 4 (a) demonstrates the relationship between the averaged number of connections and the sparsity parameter $\lambda$ as applied to one control subject. Suppose $\lambda_0 = 0$, $\lambda_1 = 0.1$, $\lambda_2 = 0.2$, $\cdots$, $\lambda_{20} = 2$, it is apparent that the averaged number of connections decreases with sparsity parameter $\lambda$ increasing. We compare the averaged common connections between the inferred networks with $\lambda_i$ and $\lambda_{i-1}$ ($i = 2, \cdots, 20$), as shown in Fig. 4 (a). Reassuringly, we observed that the connectivity inferred
with a larger $\lambda$ is mostly contained in the estimated networks with a smaller $\lambda$. In other words, important connections will always be selected. In our study, we learned the networks with a fixed sparsity parameter $\lambda$ of 0.5 for a fair comparison at the population level. We also compared the temporal patterns with fixed sparsity ($0.1$) across all the subjects. Fig. 4 (b) demonstrates the relationship between the value of network variation and the fused penalty parameter $\gamma$ when applied to one control subject. The network variation generally decreases with increasing values of $\gamma$. However, since we are interested in the relative differences between control and patient groups, we set $\gamma = 1.5$ for all subjects. We also compared the connectivity networks between two groups when $\gamma = 0.5$.

Fig. 5 and Fig. 6 demonstrate the examples of time varying brain connectivity networks of typical normal and PD subjects at different time points where the networks are learned with fixed parameters $h = 32s, \lambda = 0.5, \gamma = 1.5$. The proposed method could estimate the brain connectivity networks with both changing structures and coefficients. When compared with the normal subject, we note that the PD subject shows a sparser network. In addition, the PD subject has more distributed connections while the normal control subject tends to incorporate more hub regions in brain connectivity networks.

To measure temporal properties of the learned time varying brain connectivity networks, we compared the
network variations between control and PD groups in Fig. 7. The averaged network variations were significantly lower in the PD group, whether or not the sparsity parameter \( \lambda = 0.5 \) (Fig. 2(a)) or fixed sparsity (0.1) (Fig. 7(b)). Fig. 8 compares the averaged time period, defined as the duration of non-zero values, between control and PD groups. We note that the PD group has a larger time period with different parameters compared with that of control group. If we consider “switching ratio”, defined as number of time points with switching from zero to non-zero states to the total length of time points, we note that the PD group had a significantly smaller switching ratio as shown in Fig. 9.

We have also investigated the spectrum properties of the inferred brain connectivity networks. We note that the most dominant low frequency connectivity fluctuations are below 0.02 Hz, and specifically at around 0.005 and 0.015 Hz, which are consistent with previous studies [20]. While we found no significant differences between groups in the mean frequency, we suspect this could be due to the relatively small subject size in our study.

In addition to the temporal dynamics, we also studied the spatial patterns learned by the fixed sparsity (0.1) by examining “consistent” connections over time. We define consistent connections as those connections that appear at least once at one time point in all subjects within a given group. As shown in Fig. 10, the PD group has fewer cortico-basal ganglia connections and more cortico-cortical connections compared to the control group. The alterations in cortico-cortical and cortico-basal connectivity may reflect compensatory connections to ameliorate the effects of the diseased basal ganglia [38], [39], [40].

V. DISCUSSION AND CONCLUSION

It is clear that the brain is inherently non-stationary. Therefore, studying dynamic properties of brain connectivity networks could extend our understanding of brain functioning. In this paper, a penalized weighted regression model is presented to estimate both smooth and abrupt changes in the time dependent brain connectivity patterns. Compared with previous multivariate time varying approaches introduced for fMRI brain connectivity modeling, the proposed SWTV model is more flexible and allows different pairs of brain regions to have different dynamic time scales. While the proposed method is designed for the time evolving networks estimation, when the underlying models are static, the proposed method could still accurately estimate networks with appropriate parameters.

When applied to real fMRI resting state data consisting of 12 subjects with PD and 10 control subjects, PD subjects had significantly reduced network variation, likely related to impaired cognitive flexibility in PD. This highlights the importance of establishing dynamic properties in PD subjects.

While the proposed method appears promising, there are a number of limitations. We had to estimate certain parameters, such as sparsity and temporal bandwidth. Further work will be required to more comprehensively investigate time varying brain connectivity patterns over a broad range of time scales. Resting state data is particularly challenging in this regard, as different subjects will undoubtedly have different temporal patterns. We used a very simple metric to estimate temporal variability of the network, this could be expanded in future work. A previous study has suggested that larger brain regions tend to show greater connectivity variability, while the smaller regions are more stable [11]. Nevertheless the disease related changes of the time-varying patterns in brain connectivity such as we observed might be the potential biomarker for future studies [22], [41].

ACKNOWLEDGMENT

This study was partially supported by the PPRI/UBC Chair in Parkinson’s Disease (MJM) and Canadian Natural Sciences and Engineering Research Council (NSERC) Grants.

REFERENCES


0018-9294 (c) 2013 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.


Fig. 10: Connections that consistently appear in at least one time point in all subjects in (a) the control group and (b) the PD group.