Detecting functional connectivity changes in fMRI data*†

Changryong Baek[1], Matt Gampe[2], Benjamin Leinwand[2], Kristen Lindquist[2], Joseph Hopfinger[2], Katheleen M. Gates[2], and Vladas Pipiras[2]

[1]Sungkyunkwan University
[2]University of North Carolina at Chapel Hill

January 31, 2021

Abstract

Most functional Magnetic Resonance Imaging (fMRI) studies implicitly assume that the functional connectivity of brain regions across time is constant. Increasingly there has been interest in assessing if and when changes in connectivity occur during a given fMRI block. The timing and presence of such a change may be meaningful in terms of understanding state changes. Here we describe three methods for detecting changes in connectivity from a data-driven perspective. Connectivity here is quantified using the lag-0 pair-wise covariance estimates as the representation of the dynamic relations among brain regions. We present three methods for change point detection: Dynamic Connectivity Regression, Max-type method, and a PCA based method. The change point detection methods each include different ways to test if two given connectivity patterns from different segments in time are significantly different. These tests can also be used outside of the change point detection approaches to test any two given blocks of data. We compare the three methods for change point detection as well as the complementary significance testing approaches on simulated and empirical data examples.

1 Introduction

Examining changes in brain processes allows researchers to gain insights into key aspects of brain functioning that relate to behaviors, emotions, and cognition. In the early years of fMRI research, researchers quantified changes across time by identifying which specific brain regions seemed to increase (or decrease) in their BOLD activity during a task as compared to baseline, neutral, or some other time point for the individual. The current standard dictates that researchers complement such univariate analyses by also quantifying functional connectivity, defined as how brain region activity covaries across time (Friston; 2011). The connectivity patterns among brain regions likely change across time much like what is seen in univariate analyses. Given the evidence that functional connectivity may meaningfully change across time and the relative lack of focus on this in practice, we provide here an overview of three methods for detecting when change occurs as well as three accompanying approaches for testing if functional connectivity statistically differs from one segment of time to another. We provide these methods publicly and freely in a newly developed R package `detectR`.

*AMS subject classification. Primary: 62M10, 62H15. Secondary: 62G09.

†Keywords and phrases: high-dimensional time series, autocovariances, block multiplier bootstrap, dynamic factor models, principal components, hypothesis tests.

Changes in connectivity patterns occur for a large number of reasons, some of which may be of interest to the researcher. For instance, changes may occur during a task due to habituation (Denny et al.; 2013), learning (Bassett et al.; 2011), or delays in state induction (Waugh and Schirillo; 2012). Changes in connectivity can also occur during resting state due to shifts in self-directed thought (Park et al.; 2018) or due to spontaneous cognitive processes (Kucyi et al.; 2018). Importantly, the temporal dynamics of brain connectivity has been shown to systematically differ across people (Elton and Gao; 2015), suggesting that capturing changes that occur within a block can be of utility.

Researchers often ignore the possibility of changing connectivity patterns in block designs despite the high possibility that changes in connectivity likely occurs here. Functional connectivity relations are typically quantified using instantaneous (i.e., lag 0) covariances (Youssofzadeh et al.; 2017), correlations (Cohen and D'Esposito; 2016), or partial correlations (Huang et al.; 2019). These approaches assume that the patterns of connectivity are constant across time, which we know is not always the case. Some approaches for quantifying continuous change across time, such as using sliding windows (Ombao et al.; 2016), are gaining in popularity. One issue with this is that it is often quantified with windows that are known to be too short to provide appropriate SNRs (Lindquist et al.; 2014). Additionally, these approaches assume gradual and continuous changes in the temporal dynamics between neural networks, which might be true for some processes but not others. In particular, some states, such as emotions or cognitive shifts, may be best considered in non-linear terms such that a state shift occurs when one network assembly changes from one configuration to another (Kucyi et al.; 2018). This paper describes three distinct methods for detecting abrupt changes in functional connectivity.

Another important key to quantifying when a change point occurs in brain connectivity is if the change is significant. Sometimes it is known ahead of time where changes occur or are likely to occur due to task manipulation. Differences in brain connectivity occurring between two different tasks (or two different runs) for the same person can also provide useful information. For instance, one can test if brain processes differed substantially between two tasks, such as baseline versus task similarly to ROI-specific analysis, or across two segments in time for the same task (Gonzalez-Castillo et al.; 2015). In both the case of data-driven change point detection and *a priori* knowledge of where brain processes likely differ, researchers often wish to know if the functional connectivity patterns across two segments of time differ significantly. The same tests that are used to help detect differences in functional connectivity can be done independently of change point detection to assess if connectivity patterns from two given segments significantly differ. We provide an overview and examples of three such tests that can be used even when change point detection is not the focus of the analyses (i.e., predefined segments of time are investigated). We use these tests here as a complement to the change point detection but they are modular and can be used to investigate differences in people, for instance, or any two covariance matrices.

In the end, researchers often have two related but potentially separate goals: (1) to identify when changes in functional connectivity occur during the scan and (2) determine if connectivity patterns differ significantly between two given segments of time. We present and evaluate three options for statistically comparing functional connectivity. We demonstrate their utility in detecting change points and additionally utilize the tests separately outside of the change point detection to compare connectivity patterns in two given segments of time. In all cases, we use functions available in our new R package `detectR`. The paper is structured as follows. First, we provide details regarding the methods including the influence of preprocessing, qualities of the data that contribute to appropriate detection, and the specifics regarding the methods used for detecting change points in connectivity and testing for significance. Next, we provide results from a simulation study to evaluate the relative ability of these methods in detecting change points. Finally, we present a demonstration

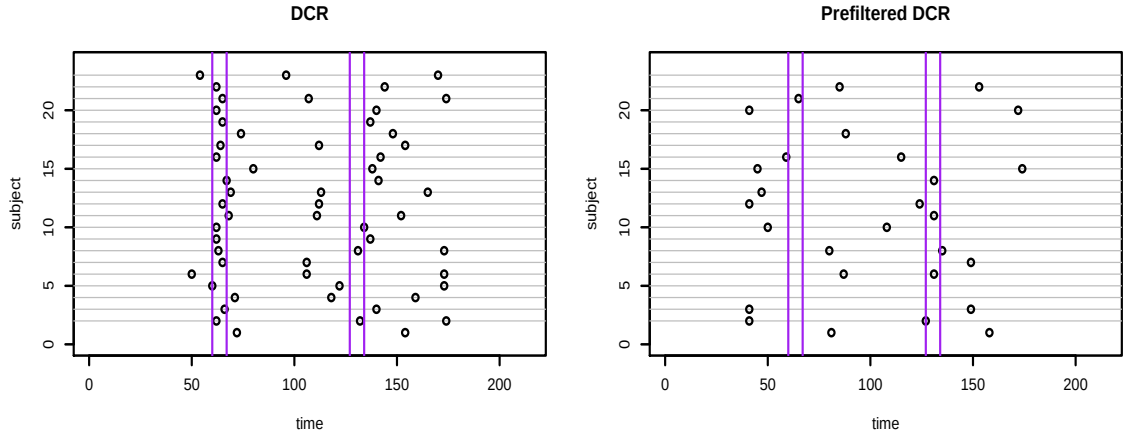

Figure 1: Estimated change points without (left plot) and with (right plot) prefiltering for data in Xu and Lindquist (2015) using DCR method.

of the approaches on empirical data where specific change points are suspected due to the study design. We close with suggestions for researchers.

2 Technical details

In this section, we review and introduce several methods to detect and test changes in connectivity. All methods considered in this paper are available in the R package `detectR`. We also discuss several related issues such as prefiltering and scaling.

2.1 Prefiltering

Before discussing connectivity changes, it is instructive to point out to practitioners that, if trends have not been removed, connectivity changes can be confused with changes in mean (trend) through the methods described below. More precisely, let $X_t = (X_{1,t}, \ldots, X_{d,t})'$, $t = 1, \ldots, T$, be the d-dimensional data of length T. Simply speaking, the methods are applied to detect changes in the covariance matrix $\mathbb{E}(X_t - \mathbb{E}X_t)(X_t - \mathbb{E}X_t)' = \mathbb{E}(X_t X_t') - \mathbb{E}X_t \mathbb{E}X_t'$. In cases where X_t has been rescaled, the methods attempt to detect changes in the correlation matrix. However, if the mean vector $\mathbb{E}X_t$ is not constant, then the methods may also detect possible changes in $\mathbb{E}X_t \mathbb{E}X_t'$. This is widely reported in the change point literature. See, for example, Pitarakis (2004) for detecting univariate variance and mean changes with least squares estimation method. To avoid the confusion between detecting changes in connectivity and trend, data should be properly prefiltered. For instance, a Butterworth filter can be applied to remove the trend. Otherwise, detection methods for connectivity changes may instead detect the changes in mean. Figure 1 compares the change point estimates depending on whether or not the data set used in Xu and Lindquist (2015) has been prefiltered. Two sets of two vertical lines indicate the beginning and end times of two separate interventions in the experiment, and the y-axis represents the subject numbers.

There are far fewer estimated change points for the prefiltered data. This indicates that some of the estimated change points in the right panel may be due to possible mean changes in the series, rather than changes in connectivity. In the remaining sections, we suppose that the data is prefiltered and assume that the mean has been subtracted, so that $\mathbb{E}X_t = 0$, $t = 1, \ldots, T$.

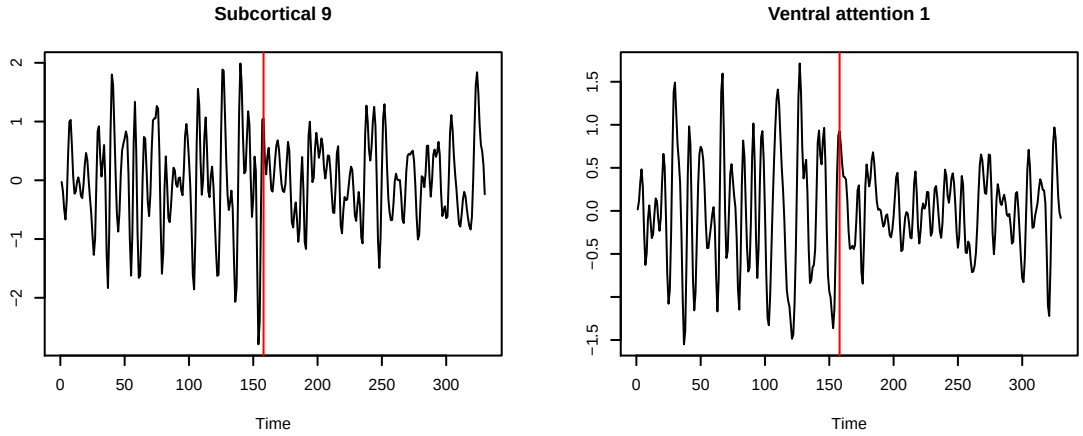

Figure 2: Time plots of fMRI series for two ROIs after prefiltering. Vertical lines indicate the detected changes in the variance.

2.2 Scaling and changes in variance

It is common to define connectivity in terms of correlation matrix $\Sigma_t = \mathbb{E}X_t X_t'$ supposing that X_t is rescaled such that each component series $X_{j,t}$ is replaced by $X_{j,t}/\mathrm{Var}(X_{j,t})^{1/2}$. Another issue (similar to that discussed in the previous section) to be aware of is that changes in correlations might be driven by changes in component variance. For example, Figure 2 depicts two fMRI series from the data presented in Section 4 below, where the variance is clearly visually different over different time blocks, and these differences can be confirmed using basic methods of detecting variance changes (as indicated by vertical lines in the plots). In what follows, we will write $\Sigma_t = \mathbb{E}X_t X_t'$ to mean either the correlation matrix (with rescaling) or the covariance matrix (without rescaling), leaving it to practitioners to choose which matrix to work with. The methods of change-point detection described below can be applied for either matrix.

2.3 Detecting change points in connectivity

This subsection details three change point detection and estimation methods for connectivity changes.

2.3.1 Dynamic Connectivity Regression (DCR)

The DCR algorithm proposed by Cribben et al. (2012) locates a change point if the BIC is significantly reduced by splitting the data into two parts, the first containing data observed until that point, and the second comprised of the data observed after that point. Using BIC or penalization in general is quite popular in the change point literature; see, for example, Bai (2000), Lavielle and Teyssiere (2006), Yao (1988) and references therein. Furthermore, DCR seeks changes in connectivity through the precision matrix $\Omega = \Sigma^{-1}$ and estimates the latter based on the graphical lasso while calculating BIC. For example, if all samples X_t, $t = 1, \ldots, T$, are used, then the estimate of Ω is expressed as

$$\widehat{\Omega} = \underset{\Omega \succ 0}{\mathrm{argmin}} \{\mathrm{tr}(\widehat{\Sigma}\Omega) - \log|\Omega| + \lambda\|\Omega\|_1\}, \tag{2.1}$$

where $\Omega \succ 0$ refers to positive-definite matrices, $\widehat{\Sigma} = T^{-1}\sum_{t=1}^{T} X_t X_t'$ is the sample covariance (or correlation) matrix, $\|\cdot\|_1$ is ℓ_1 norm and λ is the tuning parameter that controls the sparsity of

the precision matrix. Then, the BIC for the whole sample is given by

$$\mathrm{BIC}(1:T) = \mathrm{tr}\left(T \cdot \widehat{\Sigma}\widehat{\Omega}\right) - T \log|\widehat{\Omega}| + k \cdot \log(T),$$

where T is the sample size and k is the number of non-zero off-diagonal elements in the precision matrix estimate $\widehat{\Omega}$. The DCR procedure identifies t_0 as a change point if the BIC reduction by splitting, namely, the quantity

$$\mathrm{BIC}(1:T) - \Big(\mathrm{BIC}(1:t_0) + \mathrm{BIC}(t_0+1:T)\Big)$$

is significantly large. For some $\Delta > 0$, DCR uses a grid search over the points in the interval $(\Delta, T-\Delta)$ to find the largest BIC reduction while avoiding effects related to the boundary. To determine whether such reduction in BIC is indeed significant, a 100(1-α)% quantile of BIC reduction obtained from a stationary bootstrapping procedure is used. Then, the binary segmentation is applied to detect other change points. That is, the same detection procedure is continued on the data before and after the first change point, and this continues if additional change points are detected.

Our implementation of DCR in `detectR` has the following modifications. First, the penalty term in BIC, for example with one change point at t_0, is calculated as $(k_1 + k_2)\log T$ instead of $k_1 \log t_0 + k_2 \log(T - t_0)$, where k_i, $i = 1, 2$ is the number of non-zero off-diagonal parameters before/after break, respectively. Second, we consider the largest reduction of negative log-likelihood without the penalty term in BIC to find a change point (but still use the penalty when testing). This is exactly the quasi-likelihood change point estimator of Bai (2000), where the theoretical justification such as consistency is provided. Our own limited numerical experience shows that our modifications improve DCR. The DCR algorithm has several tuning parameters. They are the minimum distance from the boundaries Δ for grid search, the graphical lasso tuning parameter λ, the mean block size of stationary bootstrap, etc. We refer to Cribben et al. (2012) for more details including a refinement step. Xu and Lindquist (2015) further modified DCR to handle higher dimensional data and reduce computational time, however, their method is still in the spirit of DCR.

2.3.2 Max-type method

Instead of sparse modeling of the precision matrix, Jeong et al. (2016) proposed to use a max-type test statistic to detect connectivity changes. Furthermore, multiple change points are found by a sliding window method rather than binary segmentation. Let $\Sigma_{t,1} = (\sigma^{ij}_{t,1})_{i,j=1,\ldots,d}$, and $\Sigma_{t,2} = (\sigma^{ij}_{t,2})_{i,j=1,\ldots,d}$, be the covariance matrices from subsamples $\{X_{t-w+1}, X_{t-w+1}, \ldots, X_t\}$ and $\{X_{t+1}, X_{t+1}, \ldots, X_{t+w}\}$ of window size w and around time t, respectively. The null hypothesis of no connectivity changes $\mathrm{H}_0 : \Sigma_{t,1} = \Sigma_{t,2}$ is equivalent to $\mathrm{H}_0 : \max_{1\le i\le j\le d} |\sigma^{ij}_{t,1} - \sigma^{ij}_{t,2}| = 0$. To test H_0, Cai et al. (2013) proposed the max-type test statistic

$$M_t = \max_{1\le i\le j\le d} \frac{(\widehat{\sigma}^{ij}_{t,1} - \widehat{\sigma}^{ij}_{t,2})^2}{\mathrm{s.e.}(\widehat{\sigma}^{ij}_{t,1} - \widehat{\sigma}^{ij}_{t,2})^2}$$

and showed its asymptotic convergence to a Gumbel distribution, namely,

$$\mathbb{P}(M_t - 4\log d + \log\log d \le x) \to \exp\left(-(8\pi)^{-1}\exp(-x/2)\right), \quad \text{as} \quad T, d \to \infty.$$

Based on this result, p-values are calculated for M_t for each $t \in (\Delta, T-\Delta)$. The p-values are then transformed to z-values $z_t(w)$. Also, rather than working with one window size, Jeong et al. (2016)

suggested applying several window sizes and calculating the first PC scores z_t using PCA. Then, change points are detected through a thresholding rule for $\{z_t\}$ based on a local false discovery rate. The rule assumes that z_t follow a semiparametric mixture model

$$\pi_0 f_0 + (1-\pi_0) f_1, \tag{2.2}$$

where π_0 is the prior probability of observing z_t from null (no change), f_0 and f_1 are the null and non-null log-concave densities, respectively. The local FDR is given by $\pi_0 f_0(z)/(\pi_0 f_0(z)+(1-\pi_0)f_1(z))$, and once it is less then or equal to a prespecified constant c, a change point is detected. The expectation-maximization (EM) algorithm is used to estimate the semiparametric mixture model (2.2). There are no modifications to this method in detectR.

2.3.3 PCA based method

Introduced and studied by Han and Inoue (2015), this method detects change points in connectivity of fMRI data after dimension reduction. More specifically, let $\widehat{f}_t$ be a r-vector series of the first r PC factors, where r is considerably smaller than d. The idea is then effectively to test for change in the covariance structure by adapting approaches available in the low-dimensional setting. Setting

$$A_k = \sqrt{T}\Big(k^{-1}\sum_{t=1}^{k} \widehat{f}_t\widehat{f}_t' - (T-k)^{-1}\sum_{t=k+1}^{T} \widehat{f}_t\widehat{f}_t'\Big),$$

the test statistic is given by

$$W = \max_{\Delta \le k \le T-\Delta} \mathrm{vech}(A_k)'\widehat{S}_k^{-1}\mathrm{vech}(A_k),$$

where vech is a vectorization operation for the lower-triangular part of a symmetric matrix and $\widehat{S}_k$ is a suitable estimator of the so-called long-run covariance matrix representing the series of covariance matrices of $\widehat{f}_t\widehat{f}_t'$ across all lags, for example, given by

$$\widehat{S}_k = \frac{k}{T}\sum_{h=-q}^{q}\Big(1-\frac{h}{q+1}\Big)\widehat{\Gamma}_1(h) + \frac{T-k}{T}\sum_{h=-q}^{q}\Big(1-\frac{h}{q+1}\Big)\widehat{\Gamma}_2(h),$$

where $\widehat{\Gamma}_1(j)$ and $\widehat{\Gamma}_2(j)$ are the sample autocovariance matrices of $\widehat{f}_t\widehat{f}_t'$ at lag j for $t = 1,\dots,k$ and $t = k+1,\dots,T$, respectively. Then, under suitable assumptions (in particular, assuming X_t follows a dynamic factor model) and the null hypothesis of no connectivity changes, Han and Inoue (2015) showed the convergence in distribution

$$W \xrightarrow{d} \sup_{\epsilon \le \pi \le 1-\epsilon} \frac{(B_p(\pi)-\pi B_p(1))'(B_p(\pi)-\pi B_p(1))}{\pi(1-\pi)}, \tag{2.3}$$

where $B_p(\cdot)$ is a $p = r(r+1)/2$ dimensional vector of independent Brownian motions on $[0,1]$ and $\Delta/T \to \epsilon$. The critical values for the test are obtained by numerically approximating (2.3).

For multiple connectivity changes, one can use binary segmentation or a sliding window. For the latter method, one calculates the test statistic W using only $\{X_t, X_{t+1},\dots,X_{t+w}\}$ for each $t = 1,\dots,T-w$. The threshold value can be similarly calculated based on (2.3), however, the block wild bootstrap method described in Baek et al. (2021) is used for better approximation of the critical value. That is, the empirical quantile of W is computed from bootstrap samples $\{w_i\widehat{f}_t\widehat{f}_t'\}$ where the weights $\{w_i\}$ are i.i.d. $\mathcal{N}(0,1)$, i refers to the block of time t. Then, the change points are identified by finding suitable local maxima over the threshold. The package `detectR` finds local maxima if it is located Δ away from boundaries and above the threshold for at least 7 consecutive times (Eichinger and Kirch; 2018).

Remark 2.1 Note that DCR and Max-type tests have advantage over PCA based method in the sense that they also provide the nodes where change occurs.

2.4 Testing for changes in connectivity

In many fMRI studies, experiments are conducted to find changes in connectivity across different segments of time based on tasks or to compare different people and naturally call for suitable statistical tests of such changes or differences in connectivity patterns. Thus, suppose that $\mathbb{X} = \{X_t\}_{t=1,\dots,T}$ and $\mathbb{Y} = \{Y_t\}_{t=1,\dots,T}$ are samples representing two fMRI signals and that we are interested in testing the equal connectivity structures, $\mathrm{H}_0 : \Sigma_{\mathbb{X}} = \Sigma_{\mathbb{Y}}$. This problem is similar to change point detection in the sense that a change point would occur at the midpoint of the sample combining $\mathbb{X}$ and $\mathbb{Y}$. For this reason, the methods for testing of changes in connectivity can be defined analogously to those for detecting change points.

2.4.1 DCR test

Let $\mathbb{Z} = \{\mathbb{X}, \mathbb{Y}\}$ be the combined sample. Then, one can test for connectivity change by checking the difference of BIC given by

$$\mathrm{BIC}(\mathbb{Z}) - \Big(\mathrm{BIC}(\mathbb{X}) + \mathrm{BIC}(\mathbb{Y})\Big),$$

where $\mathrm{BIC}(\mathbb{W})$ denotes the BIC calculated from the sparse estimation of the precision matrix as in (2.1) but using the sample $\mathbb{W}$. The p-value can again be obtained from a stationary bootstrap. That is, a stationary bootstrap sample is applied to each sample $\mathbb{X}$ and $\mathbb{Y}$, and the combined bootstrap sample is used calculate the bootstrap BICs.

2.4.2 Max-type test

Let $Z_t = (Z_{j,t})_{j=1,\dots,d^*} = \mathrm{vech}(X_t X_t' - Y_t Y_t')$ be the $d^* = d(d-1)/2$ dimensional vector of length T. Note that the null hypothesis of no connectivity change is equivalent to testing $\mathrm{H}_0 : \mathbb{E}Z_t = 0$. The max-type test is based on the test statistic

$$\sqrt{T} \max_{j=1,\dots,d^*} \Big| T^{-1} \sum_{t=1}^{T} Z_{j,t} \Big|. \tag{2.4}$$

The p-value is calculated based on the block wild bootstrap method. See Baek et al. (2021) for more details on the block wild bootstrap including a long-run variance adjustment proposed by Zhang and Wu (2017) and a sum-type test where the sum of squares of sample means are considered.

2.4.3 PCA-based test

This test is based on suitable low-dimensional PC factors $\mathbb{X}^{pc}$ and $\mathbb{Y}^{pc}$. Calculating the sample covariance matrices from $\mathbb{X}^{pc}$ and $\mathbb{Y}^{pc}$ gives the vectorized difference as

$$\widehat{\Delta}_\gamma = \mathrm{vec}\left(\widehat{\Sigma}_{\mathbb{X}^{pc}} - \widehat{\Sigma}_{\mathbb{X}^{pc}}\right).$$

Under suitable assumptions, Baek et al. (2021) showed the asymptotic normality of $\sqrt{T}\widehat{\Delta}_\gamma$, leading to the test statistic

$$\frac{T}{2} \widehat{\Delta}_\gamma \widehat{W}^{-1} \widehat{\Delta}_\gamma, \tag{2.5}$$

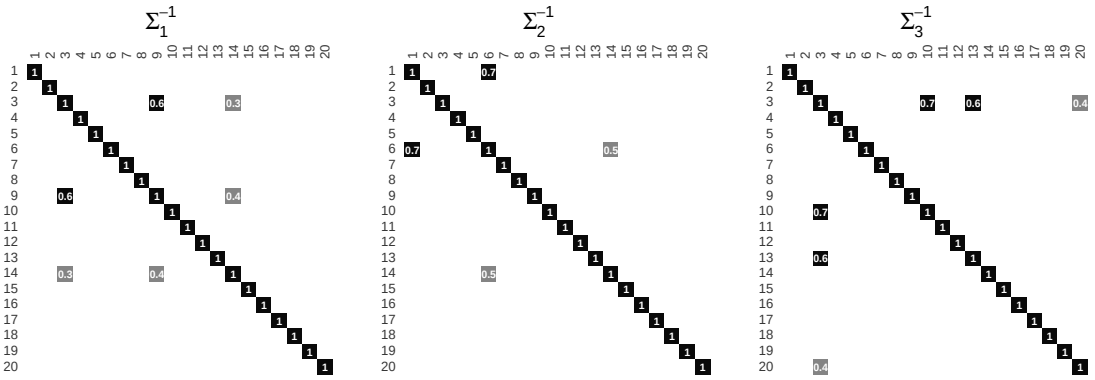

Figure 3: Precision matrices for DGP3. Only non-zero entries are indicated in the plots.

where $\widehat{W}$ is a consistent estimator of the limiting covariance matrix of $\sqrt{T}\widehat{\Delta}_\gamma$. The test statistic 2.5 asymptotically follows a χ^2 distribution with $r(r+1)/2$ degree of freedom. See Baek et al. (2021) for the exact formula of $\widehat{W}$ and more details.

3 Simulation study

In this section, we evaluate the performance of the methods in Section 2. Data is simulated according to the following general model,

$$X_t = \Sigma_t^{1/2} e_t, \quad t = 1, \ldots, T,$$

where Σ_t are $d \times d$ positive definite matices, e_t is a d-dimensional noise and the sample size is $T = 450$. We have used an autoregressive model of order 1 with parameter .5 to generate noise in each dimension, mimicking the temporal dependence of fMRI data. The sparsity of the matrices used here was informed by prior simulation work on change point detection for fMRI (Xu and Lindquist; 2015). The choices of Σ_t are made according to one of the following data generating processes (DGPs):

(DGP1) The data has no connectivity change and $\Sigma_t = I_d$ for $1 \leq t \leq 450$, where I_d is a $d \times d$ identity matrix.

(DGP2) The data has no connectivity change and $\Sigma_t = B_d$ for $1 \leq t \leq 450$, where $B_d = (b_{ij})$ is a band matrix with $b_{ij} = 0$ if $|i - j| > 5$. If $|i - j| \leq 5$, $b_{ij} = .5$ if $i \neq j$ and $b_{ii} = 1$.

(DGP3) There are two changes at times 150 and 300 in the precision matrix with dimension $d = 20$. The precision matrices for this model are given in Figure 3.

(DGP4) This model has two connectivity changes at times 150 and 300: $\Sigma_t = I_d$ for $1 \leq t \leq 150$, $\Sigma_t = B_d$ for $151 \leq t \leq 300$, and again $\Sigma_t = I_d$ for $301 \leq t \leq 450$.

The following tuning parameters are used in the considered methods. For DCR, we set $\Delta = 40$, the number of bootstrap replications is 100, and the stationary bootstrap has mean block size of 10. The penalty parameter λ is searched in the interval $[.001, 10]$. For the max-type test, we use the same $\Delta = 40$, window sizes of $30, 40, 50, 60, 70$ are applied to obtain the first PC scores, and the cut-off constant for the local FDR is set to $c = .1$. For the PCA-based method, we use a window size of 50 together with the number of factors $r = 3$. The data adaptive bandwidth rule is used for long-run variance bandwidth q. The number of block wild bootstrap replicates is 199. The block

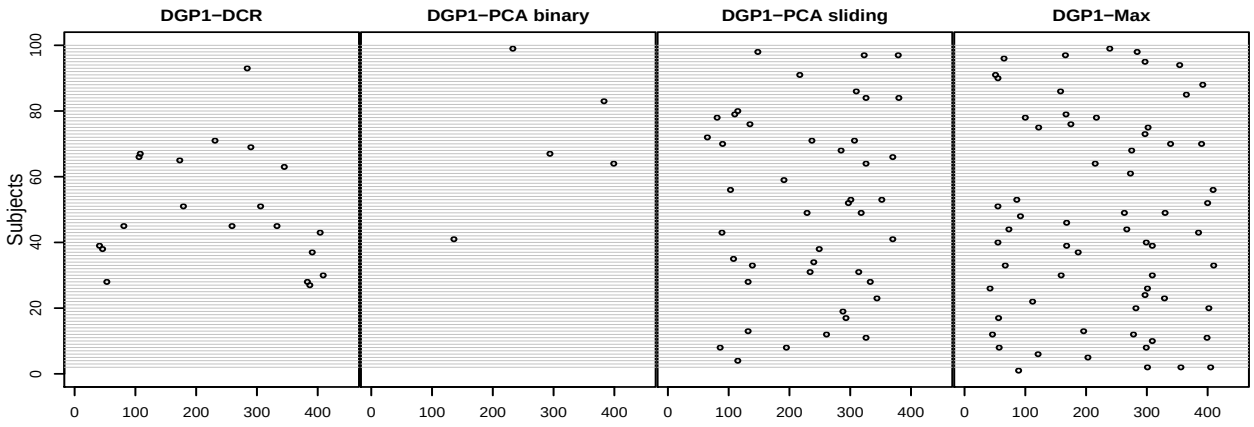

Figure 4: The estimated change points in DGP1 with 20 nodes. No change point in DGP1 and an identity matrix is used for covariance structure.

length of wild bootstrap is set to data adaptively although a square root rule $\sqrt{T}$ is also available in `detectR`; see Baek et al. (2021) for details. Also, the significance level $\alpha = .05$ is used in both DCR and PCA-based methods. The number of nodes (dimension d) is taken as either 20 or 250 to be consistent with the numbers often used in the fMRI literature.

The performance of the proposed methods is assessed by calculating the detection rate at each time point among 100 replications (subjects) as in Jeong et al. (2016), and also displaying the estimated change points. Figure 4 shows the estimated change points for DGP1 with 20 nodes. The PCA binary method seems to achieve correct 5% significant level while the other methods are over-sized. The DGP2 also has no connectivity change with a band matrix and the results are summarized in Figure 5. The DCR and PCA binary methods show no sign of significant changes in connectivity while PCA sliding and Max-type tests are oversized, suggesting that spurious change points are detected.

On the other hand, in the models DGP3 and DGP4 in Figures 6 and 7, two connectivity changes occur at time points 150 and 300, respectively. All methods successfully detect both change points. The DCR method shows the highest detection rate for DGP3, while the PCA-based method with binary segmentation shows compatible performance. When the band matrix is considered as in DGP4, the DCR method performs the best followed by the PCA binary method.

We also considered more realistic situations by increasing the number of nodes from $d = 20$ to $d = 250$ in DGP2 and DGP4. Figures 8 and 9 suggest that the DCR and PCA binary methods seem to achieve correct size while showing the most satisfactory detection rates. However, the PCA sliding method estimates the true change points with the smallest variability. For example, if we calculate the rate in the neighborhood of true change point within 5 points, then PCA sliding gives 69% and 66% while DCR gives 38% and 43%, respectively. In general, the DCR and PCA binary methods perform best when the number of nodes is very large. However, the computational time of DCR is enormous taking up to a few hours for a single run while PCA binary takes less than a few minutes.

Finally, we examine the finite sample performance of the testing methods for changes in connectivity. Empirical sizes and powers are calculated over 200 replications with dimension $d = 20$ or $d = 250$ for sample size $T = 150$. We used two covariance matrices, an identity matrix I_d and a band matrix B_d. Table 1 summarizes the resulting empirical sizes and powers, where DCR represents the DCR test in Section 2.4.1, BMB is the block multiplier bootstrap method for (2.4), LVBWR is the long-run variance adjustment of the BMB method with Rademacher multiplier,

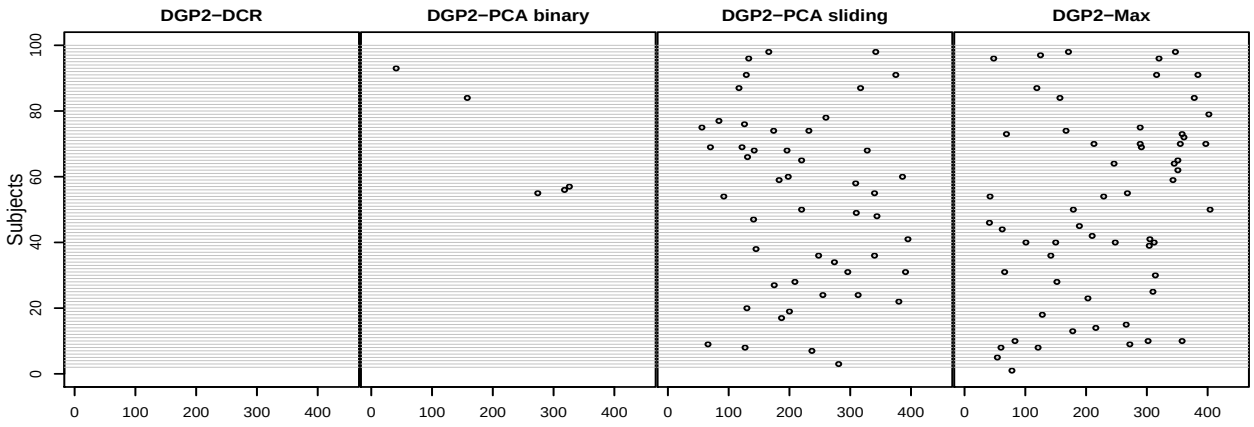

Figure 5: The estimated change points in DGP2 with 20 nodes. No change point in DGP2 and a band matrix is used for covariance structure.

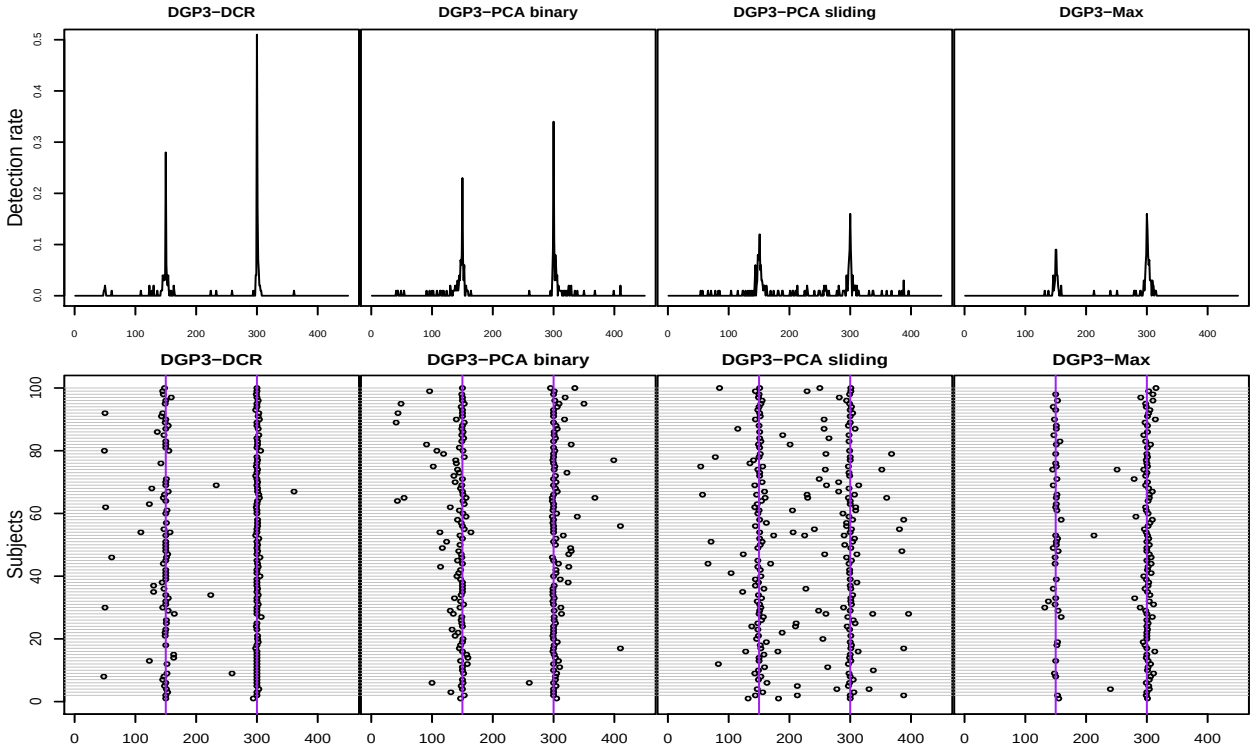

Figure 6: The detection rates (top) and the estimated change points in DGP3 with 20 nodes. Connectivity changes at time points 150 and 300.

BSUM is the sum-based test, PCA is the PCA-based test in Section 2.4.3. For more details on LVBWR and BSUM, see Baek et al. (2021). All considered models achieve reasonable sizes for both 20 and 250 nodes. However, LVBWR and PCA-based tests show the highest power for both moderate and very large number of nodes.

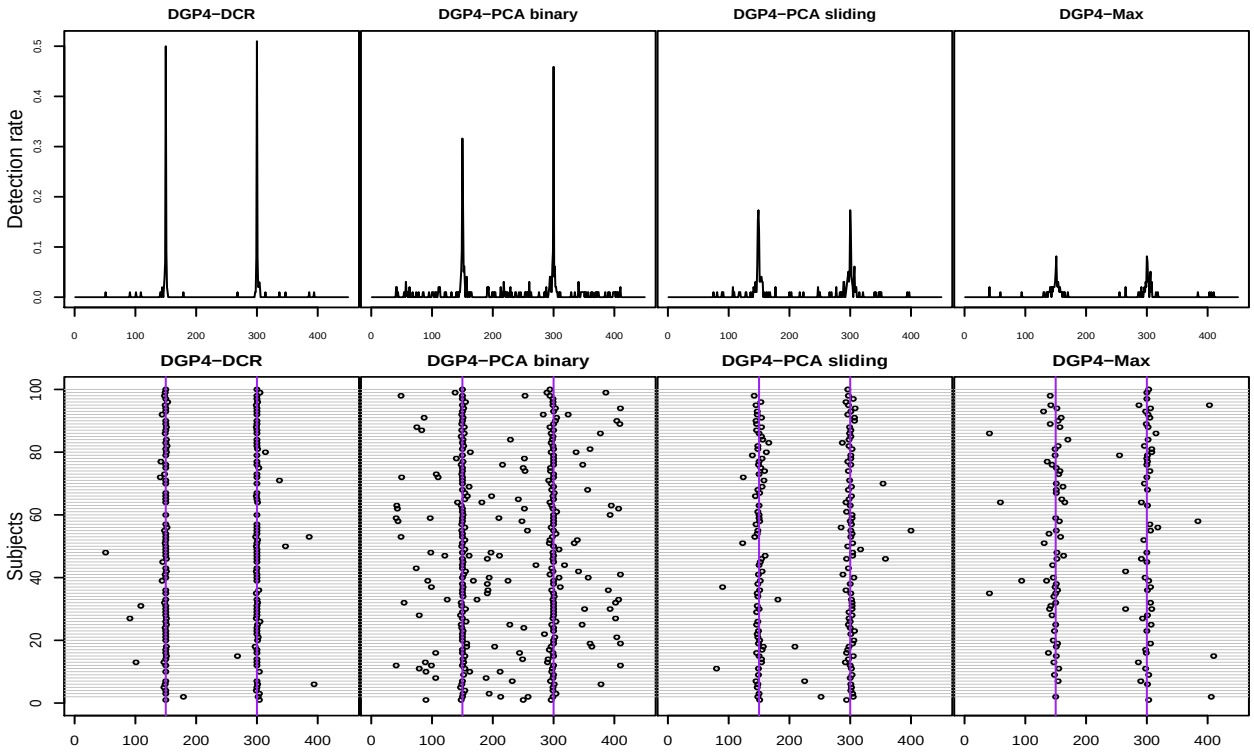

Figure 7: The detection rates (top) and the estimated change points in DGP4 with 20 nodes. Connectivity changes at time points 150 and 300.

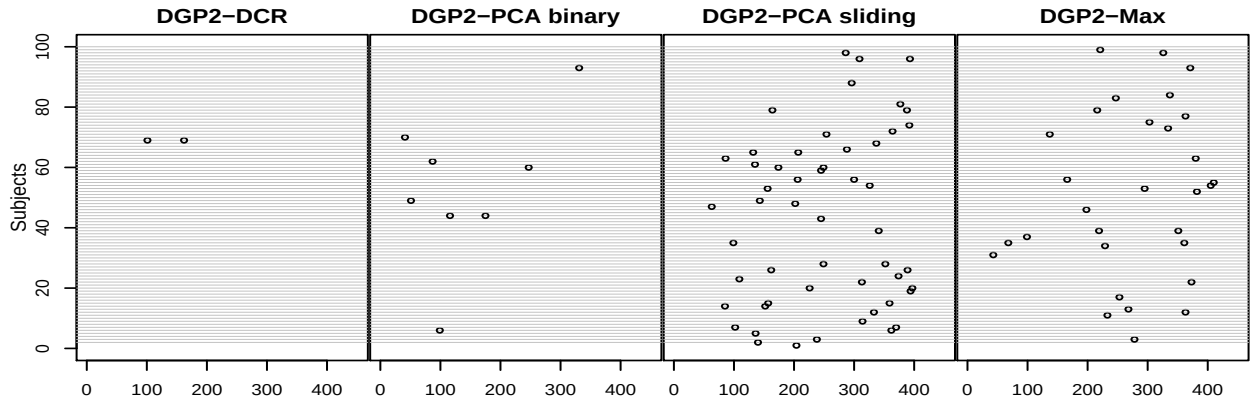

Figure 8: The estimated change points in DGP2 with 250 nodes. DGP2 has no connectivity changes.

4 Empirical study

We now turn to data collected for the purposes of probing the methods' ability to detect change points in an experimental design where we know where change points are expected to occur. We also test if the connectivity patterns are significantly different for the three tasks used here to assess the degree to which our manipulation worked.

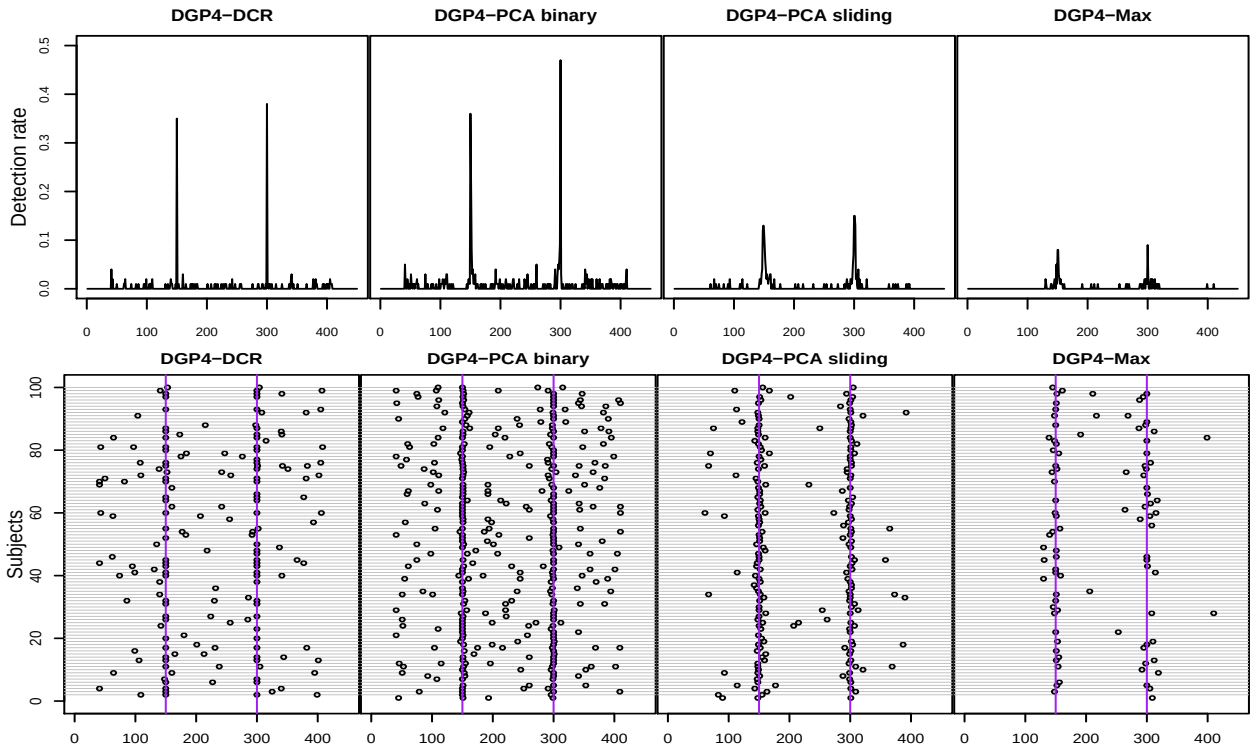

Figure 9: The detection rates (top) and the estimated change points in DGP4 with 250 nodes. Connectivity changes at time points 150 and 300.

	Model	DCR	BMB	LVBWR	BSUM	PCA
$d = 20$	I_d vs I_d	0.000	0.005	0.070	0.000	0.035
$d = 250$	I_d vs I_d	0.080	0.000	0.060	0.000	0.050
$d = 20$	B_d vs B_d	0.045	0.000	0.050	0.000	0.015
$d = 250$	B_d vs B_d	0.020	0.000	0.050	0.000	0.030
$d = 20$	I_d vs B_d	1.000	0.965	1.000	1.000	1.000
$d = 250$	I_d vs B_d	0.780	0.555	1.000	0.125	1.000

Table 1: Empirical sizes and powers for the proposed tests for changes in connectivity.

4.1 Data acquisition

4.1.1 Study design

Data were collected as part of a larger study that aimed to subset brain data based on task. In order to investigate the ability to detect brain connectivity changes, we collected data on a subset of the total participant set ($N = 6$). The goal here was to gather data with which we can probe the ability for methods to detect change points. The design used here could best be described as "continuous", as scanning continued during the shifts from resting to task, and from the first task to the second task. The conditions used in this continuously gathered design are: (1) resting state (t = 1:150 scans); (2) Motor+Auditory task (t = 151:240); and (3) Face+Motor+Auditory task (t = 241:330). Two individuals with very low movement provided appropriate data. The change point detection methods are very sensitive to abrupt changes in the data, which are often an artifact of motion. For succinctness and space concerns we focus on one individual here. Participants

were screened to ensure that they were free of neurological or psychiatric disorders, had no metal in their bodies, were not currently pregnant, had normal or corrected-to-normal vision and were right-handed. All procedures were approved by the Institutional Review Board at the University of North Carolina at Chapel Hill, and all participants provided written informed consent.

Participants viewed images on a translucent screen through an angled mirror attached to the head coil. Open source software OpenSesame (http://osdoc.cogsci.nl; Mathôt, Schreij, and Theeuwes; 2012) was used to present stimuli and record button-press responses. Prior to beginning the MRI session, participants completed a brief (2 minute) training run on a laptop outside of the scanner. This practice run was conducted to provide exposure to the stimuli and to ensure that the subjects understood the response requirement.

The Motor+Auditory condition was designed to engage the motor system, without engaging the visual processing systems. In this condition, participants were instructed to keep their eyes closed for the entire duration of the run, and to press the button on the MRI compatible response box (Current Designs Pyka 5-Finger Response Pad) every time a specified tone was played through the headphones. The tone was presented at the same rate (once every 800 msec) as the face image was presented in the Face+Motor+Auditory condition. This was done to ensure that the motor system was engaged at the same temporal frequency in this condition as the other task condition. The auditory stimulus was a 220 Hz tone, played for 100 msec. There was 700 msec between successive auditory tones, and participants were asked to press the response button after each tone. This task continued for 3 minutes, beginning at 302 seconds after the resting state.

The Face+Motor+Auditory condition was designed to engage all of the systems utilized in the Motor+Auditory condition, plus the visual areas including the fusiform face area of the ventral visual system. The highly similar response requirements between the Motor+Auditory and Face+Motor+Auditory conditions, with the addition of face processing in the latter, provides a test of whether the change point detection method is sensitive enough to detect these relatively small differences between conditions within the same subject. The visual images were face stimuli taken form the NimStim database (Tottenham et al.; 2009); there was a total of 20 faces used (each appeared once in every 16 second stimulation period, in randomized order), and the auditory stimulus (220 Hz tone) was the one used in the Motor+Auditory condition. On each trial, a face image was presented simultaneously with the tone. The tone was presented for a duration of 100 msec, and the face image was presented for 500 msec duration. Subjects were instructed to press the response button for each presentation of the face/tone. Following the offset of the face, there was 300 msec of a fixation-cross-only screen. This task continued for 3 minutes, beginning at 482 seconds into the run.

Participants were highly familiar with the stimuli and tasks before the current "continuous" run was conducted, because they practiced these tasks before entering the scanner and because each participant performed 6 runs during fMRI data collection before the current "continuous" run reported here. Those first 6 runs included at least one full run of each of the conditions alone (only one condition per run), as well as a visual-only run in which the faces were presented but no motor response was required. Immediately before the current "continuous" run, the participants were instructed that they were going to do a single long run with the three conditions of Rest, Motor+Auditory, and Face+Motor+Auditory, and they were informed of the order and the task requirements.

4.1.2 MRI scanning protocol

Images were collected using a 3T Siemens PRISMA MRI system at the University of North Carolina Biomedical Research Imaging Center. Functional images included 37 transverse slices (3x3x3 mm^3

	264 ROIs		9ROIs	
Block Comparison	DCR	PCA	DCR	PCA
resting vs motor+auditory	0.000	0.003	0.000	0.019
resting vs face+motor+auditory	0.000	0.024	0.020	0.021
motor+auditory vs face+motor+auditory	0.00	0.144	0.060	0.667

Table 2: Test results on the prespecified blocks. p-values provided for each comparison.

resolution), collected interleaved inferior to superior. Images were acquired using a T2-weighted echo-planar imaging (EPI) sequence (TR = 2000 ms, TE = 26 ms, flip angle = 80 degrees), and the first three scans of each run were discarded to allow for magnetic field stabilization. After discarding the first three scans, the fMRI runs reported here each lasted 448 seconds. A structural scan was also acquired for each participant (T1-weighted; TR = 2400 ms; TE = 2.22 ms; flip angle = 8, FOV = 256 mm, 208 slices, 0.8x0.8x0.8 mm^3 resolution; 398 seconds duration).

4.1.3 fMRI processing

MRI data were preprocessed using SPM12 (Wellcome Department of Imaging Neuroscience, University College London, UK). Preprocessing included spatial realignment and slice-time correction. The mean image constructed from realignment was used to determine parameters for coregistration and spatial normalization into the standard MNI-space using the EPI-template included in SPM12. The fMRI data were normalized at a 2x2x2 mm^3 resolution, and were then smoothed with an 8 mm (FWHM) isotropic kernel. The first components of white matter and cerebral spinal fluid data were regressed out of the data, and a Butterworth filter with lower bound of .02 and upper bound .1 Hz was applied prior to analysis. Region of interests were defined using the Power atlas (Power et al.; 2011). Two node sizes are considered: (1) the 264 brain regions in the Power atlas and (2) a subset of 9 of those with one brain region taken from the sensory-somatomotor hand, auditory, cingulo-opercolar, visual, fronto-parietal, default mode, salience, ventral, and dorsal networks. The specific ROI chosen was the one within each network that evidenced the greatest variability across the continuous run as these likely changed the most across the tasks.

4.2 Testing for differences in connectivity: *a priori* cut points

We first wished to identify which, if any, of the tasks produced significantly different connectivity matrices. Since max-type test requires the balanced sample size between two groups, here we only report results on DCR and PCA-based tests. The p-values are summarized in Table 4.2. It is observed that all tests clearly indicate that resting state is significantly different from tasks regardless of how many ROIs we have considered. However, between tasks, the results are mixed. PCA-based test indicates that they are not significantly different for all ROIs we considered, however, DCR tests shows much stronger evidence of rejecting the null hypothesis of same brain connectivity especially when high dimensional ROIs are considered.

4.3 Change point detection

With the knowledge that resting state connectivity in our design differed significantly from the two tasks, which did not significantly differ from each other, we would anticipate the detection of a change point near $t = 151$, as that is when the first task began. We used $\Delta = 50$ for the minimum

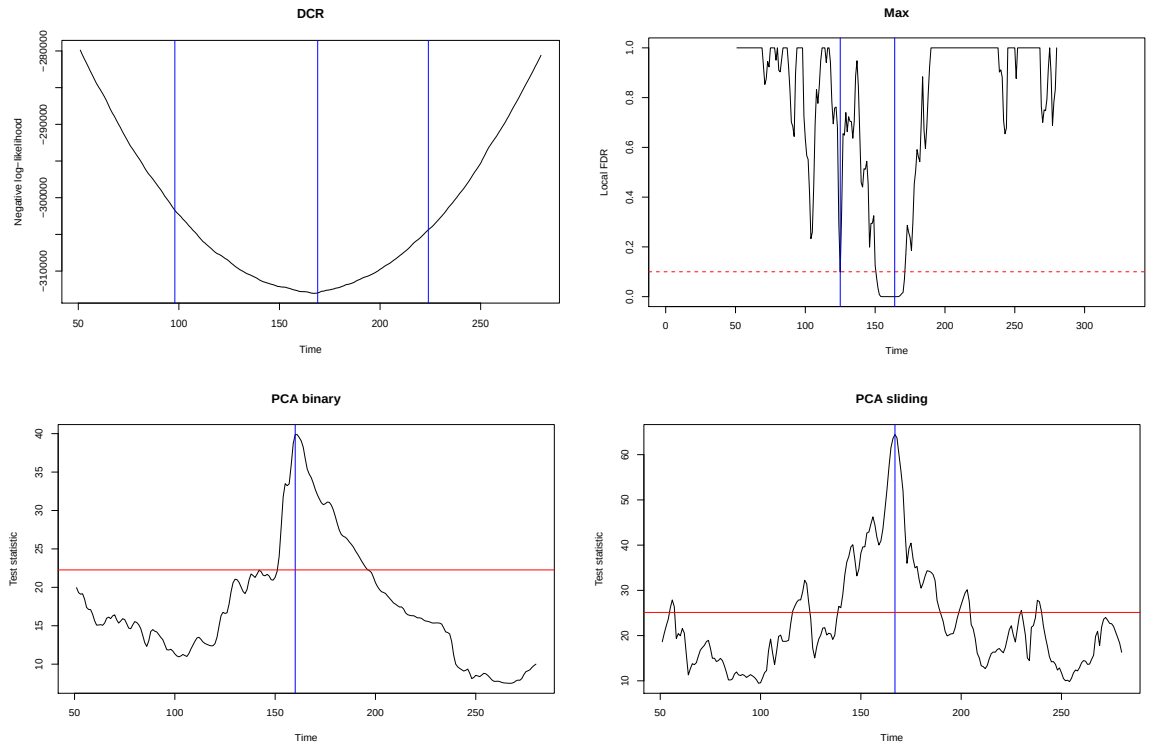

Figure 10: The test statistics and final estimated change points in blue lines when 264 brain regions are used.

distance between change points and away from boundaries, and used the same tuning parameters as in Section 3.

The results from using the full 264 brain regions are depicted in Figure 10. The top left panel shows the negative log-likelihood when the time point t is assumed to be a change point. The first candidate for a change point is the one with the smallest negative log-likelihood, which here is found to be $t = 160$. Now, we apply bootstrap to determine if the BIC reduction is significant. Indeed the first candidate $t = 160$ gives empirical p-value of 0, so it is identified as the change point. Then, the same procedure is repeated over before/after the first change point until no further splitting is possible. Finally, the DCR finds two more breaks at $t = 98$ and $t = 224$.

The top right plot of Figure 10 shows the max-type method in Jeong et al. (2016). The estimated local FDR is depicted, and the local minimums below the prespecified constant $c = .1$ are identified as change points. This gives the estimated breaks as $t = 125$ and $t = 164$, but as the plot suggests, the evidence is stronger around 164. For the PCA-based method, the test statistic W which is the maximum over a range of time points is located at $t = 160$, and the test statistic is above the critical value at the horizontal line, so it is significant change point. The binary segmentation PCA algorithm found maximum tests statistics at $t = 110$ and 233. However, both were not significant under 5% significance level. The PCA-based method with sliding window simultaneously identify the change points once the rolling test statistic is the local maximum above the critical value. We used the 5 consecutive time points to define local maximum, that is, the test statistics that should be above the threshold for at least 5 consecutive time points. The final estimated change point is $t = 167$. All methods seem to indicate change point around $t = 160$, which is anticipated, but a subsequent change point around $t = 240$ was not as reliably found as the first one.

Finally, for the selected 9 ROIs, the DCR method gives 109, 161, and 268 as change points, the max-type test gives 55, 157 and 256, the PCA-based method with binary segmentation gives 129

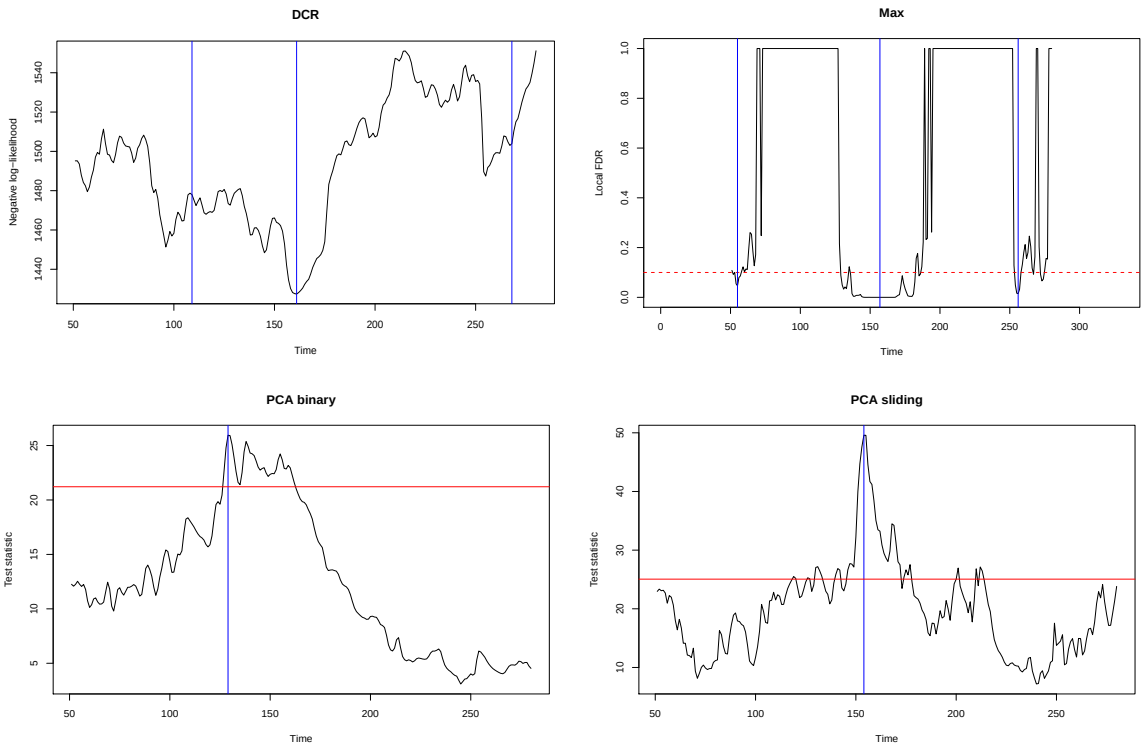

Figure 11: The test statistics and final estimated change points in blue lines when 9 ROIS are selected.

while the sliding method identified 154 as change point (see Figure 11). These results are similar to those found using the set of 264 brain regions by finding change points close to the expected shift from resting state to task that occurred at $t = 151$. However, we see additional change points occurring closer to the begin of the second task beginning ($t = 240$) as well as change points that occurred during the resting state task and are unexplained by the task design.

5 Conclusions

The paper provided an evaluation of three methods for detecting change points in functional connectivity across a block. None of the methods - DCR, Max-type, or PCA based - performed unequivocally better than the others. We also compared the results of these methods on an empirical fMRI example, finding much consistency in inferences along with some minor differences across methods. All analyses were conducted with the newly developed `detectR` R package.

All methods were generally able to detect change points that existed in the data generating models. When the data were generated to not have change points, both Max-type and PCA with sliding windows recovered many false positive break points. DCR and PCA with binary segmentation did not, suggesting that these methods may be first steps to deciding if there are any true change points or not. An added benefit of PCA is that it is computationally less intensive than DCR, and thus provides a good starting point for analyses. Should change points be suggested, then the qualities of the data matrix may help guide next steps. DCR and PCA binary performed well for any node size when the matrices were very sparse. PCA sliding and Max-type generally performed better when the data matrices contained more non-zero values.

In our data example we utilized two lag-0 covariance matrices (one with 9 ROIs, the other

with 264) from one data participant across three blocks with no breaks. First we examined if significant differences in functional connectivity existed across resting state and two task blocks. For both the large and small sized connectivity maps the PCA method found a significant difference in connectivity between resting state and the two tasks but not between the two tasks. DCR detected differences between the two tasks in addition to between the resting state and task blocks. Consistent with these significant differences, we found that the methods detected a change point in functional connectivity shortly after the participant shifted to task for both the 9 and 264 sized matrices, with the esception that the PCA with binary segmentation method missed this shift in the 9 ROI case. Additional change points were also detected by most methods in resting state (which may be true or not, and were outside the control of the experiment design). DCR additionally found a change point during the first task. Taken together, we can assert that a shift in brain processing as assessed with functional connectivity of the full functional brain likely occurred when going from resting state to active tasks, and perhaps one other during resting state.

As indicated in the test, prior to the change point detection analyses described here the researcher must make a few key decisions that are similar across all methods used here. First, one must decide whether or not to detrend the data to remove any shifts in the mean across time. This prefiltering is common in fMRI. In the present context, neglecting to do so would capture changes in the means of the brain regions along with shifts in connectivity. Second, the researcher must consider if changes in variance of each ROI are of interest, or solely shifts in the covariance of ROIs. If changes in the variance are of no interest, the research might rescale the data by using the correlation matrix rather than covariance matrix as input to these tests. Finally, the researcher needs to select the level of geographical resolution to use. Should the entire functional brain be use, changes may be missed if only connectivity in a subset of ROIs (nodes) shifts across time. A drawback from focusing in on a few regions may be that changes that unexpectedly occur elsewhere may be missed.

This project evaluated and compared three state-of-the-art methods for detecting nonlinear change points in functional connectivity. The results, while promising and indicating utility of these methods, point to areas of growth. For one, it seems that PCA with sliding windows and Max-type might be overly sensitive in detecting change points where none truly exist. This drawback is balanced by the true detection of change points when they do indeed exist. Hence, some work to decrease the rate of false positives is warranted for these methods.

Acknowledgments

The authors thank Seok-Oh Jeong for proving code for max-type method. The work of the first author was supported in part by the National Research Foundation of Korea (NRF-2019R1F1A1057104). The work of Hopfinger and Gates and data acquisition was supported by the National Institutes of Health - National Institute of Biomedical Imaging and Bioengineering (R01 EB021299).

References

Baek, C., Gates, K. M., Leinwand, B. and Pipiras, V. (2021), 'Two sample tests for high-dimensional autocovariances', *Computational Statistics & Data Analysis* **153**, 107067.

Bai, J. (2000), 'Vector autoregressive models with structural changes in regression coefficients and in variance-covariance matrices', *Annals of Economics and Finance* **1**(2), 303–339.

Bassett, D. S., Wymbs, N. F., Porter, M. A., Mucha, P. J., Carlson, J. M. and Grafton, S. T. (2011), 'Dynamic reconfiguration of human brain networks during learning', *Proceedings of the National Academy of Sciences* **108**(18), 7641.

Cai, T., Liu, W. and Xia, Y. (2013), 'Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings', *Journal of the American Statistical Association* **108**(501), 265–277.

Cohen, J. R. and D'Esposito, M. (2016), 'The segregation and integration of distinct brain networks and their relationship to cognition', *Journal of Neuroscience* **36**(48), 12083–12094.

Cribben, I., Haraldsdottir, R., Atlas, L. Y., Wager, T. D. and Lindquist, M. A. (2012), 'Dynamic connectivity regression: determining state-related changes in brain connectivity', *Neuroimage* **61**(4), 907–920.

Denny, B. T., Fan, J., Liu, X., Guerreri, S., Mayson, S. J., Rimsky, L., New, A. S., Siever, L. J. and Koenigsberg, H. W. (2013), 'Insula–amygdala functional connectivity is correlated with habituation to repeated negative images', *Social Cognitive and Affective Neuroscience* **9**(11), 1660–1667.

Eichinger, B. and Kirch, C. (2018), 'A MOSUM procedure for the estimation of multiple random change points', *Bernoulli* **24**(1), 526–564.

Elton, A. and Gao, W. (2015), 'Task-related modulation of functional connectivity variability and its behavioral correlations', *Human brain mapping* **36**(8), 3260–3272.

Friston, K. J. (2011), 'Functional and effective connectivity: a review', *Brain connectivity* **1**(1), 13–36.

Gonzalez-Castillo, J., Hoy, C. W., Handwerker, D. A., Robinson, M. E., Buchanan, L. C., Saad, Z. S. and Bandettini, P. A. (2015), 'Tracking ongoing cognition in individuals using brief, whole-brain functional connectivity patterns', *Proceedings of the National Academy of Sciences* **112**(28), 8762–8767.

Han, X. and Inoue, A. (2015), 'Tests for parameter instability in dynamic factor models', *Econometric Theory* **31**(5), 1117–1152.

Huang, B., Zhang, K., Sanchez-Romero, R., Ramsey, J., Glymour, M. and Glymour, C. (2019), 'Diagnosis of autism spectrum disorder by causal influence strength learned from resting-state fmri data', *arXiv preprint arXiv:1902.10073* .

Jeong, S.-O., Pae, C. and Park, H.-J. (2016), 'Connectivity-based change point detection for large-size functional networks', *NeuroImage* **143**, 353–363.

Kucyi, A., Tambini, A., Sadaghiani, S., Keilholz, S. and Cohen, J. R. (2018), 'Spontaneous cognitive processes and the behavioral validation of time-varying brain connectivity', *Network Neuroscience* **2**(4), 397–417.

Lavielle, M. and Teyssiere, G. (2006), 'Detection of multiple change-points in multivariate time series', *Lithuanian Mathematical Journal* **46**(3), 287–306.

Lindquist, M. A., Xu, Y., Nebel, M. B. and Caffo, B. S. (2014), 'Evaluating dynamic bivariate correlations in resting-state fmri: a comparison study and a new approach', *NeuroImage* **101**, 531–546.

Ombao, H., Lindquist, M., Aston, J. and Thompson, W. (2016), *Handbook of Neuroimaging Data Analysis*, Chapman and Hall/CPC.

Park, H.-J., Friston, K. J., Pae, C., Park, B. and Razi, A. (2018), 'Dynamic effective connectivity in resting state fmri', *Neuroimage* **180**, 594–608.

Pitarakis, J.-Y. (2004), 'Least squares estimation and tests of breaks in mean and variance under misspecification', *The Econometrics Journal* **7**, 32–54.

Power, J. D., Cohen, A. L., Nelson, S. M., Wig, G. S., Barnes, K. A., Church, J. A., Vogel, A. C., Laumann, T. O., Miezin, F. M., Schlaggar, B. L. et al. (2011), 'Functional network organization of the human brain', *Neuron* **72**(4), 665–678.

Waugh, C. E. and Schirillo, J. A. (2012), 'Timing: A missing key ingredient in typical fMRI studies of emotion', *Behavioral and Brain Sciences* **35**(3), 170–171.

Xu, Y. and Lindquist, M. A. (2015), 'Dynamic connectivity detection: an algorithm for determining functional connectivity change points in fmri data', *Frontiers in neuroscience* **9**, 285.

Yao, Y.-C. (1988), 'Estimating the number of change-points via schwarz'criterion', *Statistics & Probability Letters* **6**(3), 181–189.

Youssofzadeh, V., Akhtar, Z., Amiri, A. M. and Falk, T. H. (2017), An automated framework for emotional fmri data analysis using covariance matrix, *in* '2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP)', IEEE, pp. 760–763.

Zhang, D. and Wu, W. B. (2017), 'Gaussian approximation for high dimensional time series', *The Annals of Statistics* **45**(5), 1895–1919.