Covariance Shrinkage for Dynamic Functional Connectivity

Nicolas Honnorat[1(✉)], Ehsan Adeli[2], Qingyu Zhao[2], Adolf Pfefferbaum[1,2], Edith V. Sullivan[2], and Kilian Pohl[1,2]

[1] SRI International, Menlo Park, CA, USA
nicolas.honnorat@sri.com
[2] Department of Psychiatry and Behavioral Sciences, Stanford University, Palo Alto, CA, USA

Abstract. The tracking of dynamic functional connectivity (dFC) states in resting-state fMRI scans aims to reveal how the brain sequentially processes stimuli and thoughts. Despite the recent advances in statistical methods, estimating the high dimensional dFC states from a small number of available time points remains a challenge. This paper shows that the challenge is reduced by *linear covariance shrinkage*, a statistical method used for the estimation of large covariance matrices from small number of samples. We present a computationally efficient formulation of our approach that scales dFC analysis up to full resolution resting-state fMRI scans. Experiments on synthetic data demonstrate that our approach produces dFC estimates that are closer to the ground-truth than state-of-the-art estimation approaches. When comparing methods on the rs-fMRI scans of 162 subjects, we found that our approach is better at extracting functional networks and capturing differences in rs-fMRI acquisition and diagnosis.

1 Introduction

The development of resting-state functional MRI (rs-fMRI) has provided a way to measure spontaneous brain activity across the brain, in vivo and in real-time [1]. Similar patterns of spontaneous brain activity across brain regions are referred to as functional connectivity. Traditionally, these patterns were assumed to be static over the acquisition time of the rs-fMRI, but recent work suggests otherwise [2]. Called Dynamic functional connectivity (dFC) patterns, they are often clustered across time frames of the rs-fMRI series to identify recurring connectivity patterns or 'mind states' [7]. Analysis of these mind states has given new insights into mental disorders and neurodevelopment [9].

According to a recent review [9], the most accurate method for estimating dFC is called Dynamic Conditional Correlation (DCC) [6]. DCC filters the rs-fMRI signal via the generalized autoregressive conditional heteroskedasticity approach (GARCH) and defines a dFC for each time point of the blood-oxygen-level dependent (BOLD) time series as the remaining global connectivity at that

M. D. Schirmer et al. (Eds.): CNI 2019, LNCS 11848, pp. 32–41, 2019.
https://doi.org/10.1007/978-3-030-32391-2_4

time. This global connectivity is captured by a covariance matrix across all voxels or regions of interest estimated using the Exponentially Weighted Moving Average method (EWMA) [6]. EMWA, however, was not designed for estimating large covariance matrices from a small number of samples (or time points) as it is the case in this application. In addition, the estimation is computationally expensive preventing this type of analysis to be directly applied to the entire high resolution fMRI of large data sets.

In this work, we tackle these issues by introducing an accurate and efficient implementation of EWMA. Specifically, we first show that EWMA computes weighted covariance matrices, where the weights are defined according to a continuous sliding window. We then improve the accuracy of that weighted covariance estimation via linear *covariance shrinkage* [5], a statistical method designed for the estimation of large covariance matrices in low sample size settings. Finally, we reformulate the approach to cluster dFCs without having to explicitly compute the covariances themselves. Thus, we can efficiently identify mind states on large data sets. Compared to DCC, the dFC estimated by our approach is closer to the ground-truth on a synthetic data set. When applied to a rs-fMRI data set of 162 subjects, our approach is better at estimating functional networks and at capturing differences in MRI acquisition, and between healthy controls and those with alcohol use disorder.

2 Covariance Shrinkage for dFC

2.1 EWMA-Based dFC Estimation Using Continuous Sliding Windows

Let n be the number of time points and p the number of voxels or regions of interests of a rs-fMRI, then we denote with $X = [x_1, ..., x_n]$ the $p \times n$ matrix storing all the BOLD measurements. We assume that these measurements have been processed by a neuroimaging pipeline including standard motion correction, temporal band-pass filtering, and BOLD signal normalization. The entries of each row of X are therefore assumed to have a zero mean and unit variance. To relate EWMA to a continuous sliding window approach, we introduce a continuous sliding window $w_t := [w_t(1), ..., w_t(n)]$ for each time point t. The entries of w_t are positive and add up to 1.

$$\begin{aligned} w_1 \quad &:= (1, 0, 0, .., 0)\,, \\ w_t(i) &:= \begin{cases} 1-\theta & \texttt{if} \quad i = t \\ \theta\ w_{t-1}(i) & \texttt{otherwise} \end{cases}, \end{aligned} \tag{1}$$

where $\theta \in [0, 1]$ specifies the weight of previous time points. We then compute the weighted covariance matrix with respect to time point t as

$$\begin{aligned} C_t &:= \textstyle\sum_{i=1}^{n} w_t(i) \left(x_i - \sum_{j=1}^{n} w_t(j) x_j\right) \left(x_i - \sum_{j=1}^{n} w_t(j) x_j\right)^T \\ &= \textstyle\sum_{i=1}^{n} w_t(i) x_i x_i^T - \sum_{ij}^{n} w_t(i) w_t(j) x_i x_j^T . \end{aligned} \tag{2}$$

This covariance computation is mathematically equivalent to the recursive computation carried out by the EWMA in [6].

2.2 Linear Covariance Shrinkage

This new mathematical formulation offers a way to improve the EWMA estimations used for fMRI connectivity analysis. As mentioned, computing $C_t \in \mathbb{R}^{p\times p}$ is not reliable as the number of measurements p is much larger than the number of samples or time points n. Linear covariance shrinkage [5] mitigates this issue by replacing the empirical covariance with a linear combination between itself and its trace $Tr(\cdot)$:

$$C_t^* = (1-\lambda_t)C_t + \lambda_t \frac{Tr\left(C_t\right)}{p} I. \tag{3}$$

Under the assumption that the measurements are Gaussian distributed and $w_t(i) = 1/n$, the optimal setting of the parameter λ can be directly computed from C_t and n according to the Oracle Approximating Shrinkage [3]:

$$\lambda_t = min\left(1, \frac{\left(1-\frac{2}{p}\right) Tr\left(C_t^2\right) + Tr^2\left(C_t\right)}{\left(n+1-\frac{2}{p}\right)\left[Tr\left(C_t^2\right) - \frac{Tr^2(C_t)}{p}\right]}\right). \tag{4}$$

We found that the proof in [3] can be generalized to embed covariance shrinkage into EMWA (i.e., Eq. (2)) and thus drop the assumption that $w_t(i) = 1/n$. We replaced the original Wishart distribution moments in [3] with moments obtained for Wishart matrices weighted by w_t and we followed the derivations step by step. We finally found that Eq. (4) still holds, if one replaces the number of time points n with the *effective number of samples*:

$$n_w = \frac{\left(\sum_i w_t(i)\right)^2}{\sum_i \left(w_t(i)\right)^2}. \tag{5}$$

2.3 Efficient Implementation

Typical dFC analysis involves clustering the covariance matrices, which requires computing λ_t for each time point and ℓ_2 distance between each pair of covariances C_s^* and C_t^*. We found that the following computational trick can be used to efficiently compute λ and the ℓ_2 distance without computing the matrices themselves. Specifically, we introduce a matrix $K := X^T X \in \mathbb{R}^{n\times n}$, which is a very small matrix that can be computed at once, at the very beginning of the fMRI analysis. Denoting the entry-wise product between matrices with $\odot$, we found that Eq. (4) can be efficiently estimated with

$$\begin{aligned}
Tr\left(C_t\right) &= w_t^T Diag(K) - w_t^T K w_t, \\
Tr\left(C_t^2\right) &= w_t^T \left(K \odot K\right) w_t - 2 w_t^T K Diag(w_t) K w_t + \left(w_t^T K w_t\right)^2, \\
Tr\left(C_s C_t\right) &= w_s^T \left(K \odot K\right) w_t - L_{st} + \left(w_s^T K w_t\right)^2, \\
\text{where } L_{st} &= w_s^T K Diag(w_t) K w_s + w_t^T K Diag(w_s) K w_t.
\end{aligned} \tag{6}$$

Now let $\gamma := \lambda_s Tr(C_s)/p$ and $\delta := \lambda_t Tr(C_t)/p$, the ℓ_2 distance between two covariances C_s^* and C_t^* is efficiently obtained by

$$\begin{aligned} ||C_s^* - C_t^*||_2^2 &= ||(1-\lambda)C_s + \gamma I - (1-\mu)C_t - \delta I||_2^2 \\ &= 2(\gamma-\delta)(1-\lambda)Tr(C_s) + 2(\delta-\gamma)(1-\mu)Tr(C_t) \\ &\quad +(1-\mu)^2 Tr(C_t^2) - 2(1-\lambda)(1-\mu)Tr(C_s C_t) \\ &\quad +(1-\lambda)^2 Tr(C_s^2) + (\gamma-\delta)^2 p. \end{aligned} \tag{7}$$

This trick, inspired from the Support Vector Machine literature [4], reduces the computational burden by several order magnitudes during our experiments. It provides us with a mean to process full-resolution scans in reasonable time, and spares a significant amount of computer memory by preventing the computation of the covariances matrices, which would have contained billions of entries.

We further improve the efficiency of the computations by exploiting the relation between consecutive EWMA time windows. More specifically, we recursively compute the traces of the covariance matrices (Eq. (6)) by introducing intermediate variables:

$$\begin{aligned} \alpha_t &:= {w_t}^T Diag(K), \\ \epsilon_t &:= K w_t, \\ \beta_t &:= {w_t}^T \epsilon_t, \end{aligned} \tag{8}$$

so that $Tr\left(C_t\right) = \alpha_t - \beta_t$. Now let $K_{.t}$ denote the column of K corresponding to time point t, then the algorithm for recursively computing the intermediate variables is as follows

$$\begin{array}{ll} \texttt{Initialization} & \texttt{Recursion} \\ \alpha_1 = K_{11} & \alpha_t = \theta\alpha_{t-1} + (1-\theta)\, K_{tt} \\ \epsilon_1 = K_{.1} & \epsilon_t = \theta\epsilon_{t-1} + (1-\theta) K_{.t} \\ \beta_1 = K_{11} & \beta_t = \theta^2\beta_{t-1} + 2\theta\,(1-\theta)\,\epsilon_{t-1}(t) + (1-\theta)^2\, K_{tt} \end{array} \tag{9}$$

Similarly, $Tr\left(C_t^2\right)$ and $Tr\left(C_s C_t\right)$ in Eq. (6) can also be efficiently computed based on the above intermediate variables. This strategy allows estimating dFCs from full resolution fMRI in a few minutes using a standard office computer instead of repeatedly computing C_t matrices which would require terabytes of memory. Furthermore, it allows clustering the dFC in reproducible transient mind states efficiently, by computing the distance matrix D across all dFCs

$$D_{st} = ||C_s^* - C_t^*||_2^2, \tag{10}$$

and applying a clustering method to D. We establish the validity of this approach by computing such matrices D during our experiments, and extracting from these matrices a measurement indicative of the "clusterability" of the time points, which are potential biomarkers.

3 Experiments

3.1 Data Sets

The first data set consists of rs-fMRI of 162 subjects [8]: 18 were diagnosed with alcohol dependence (ALC) while the remaining samples were labeled as controls.

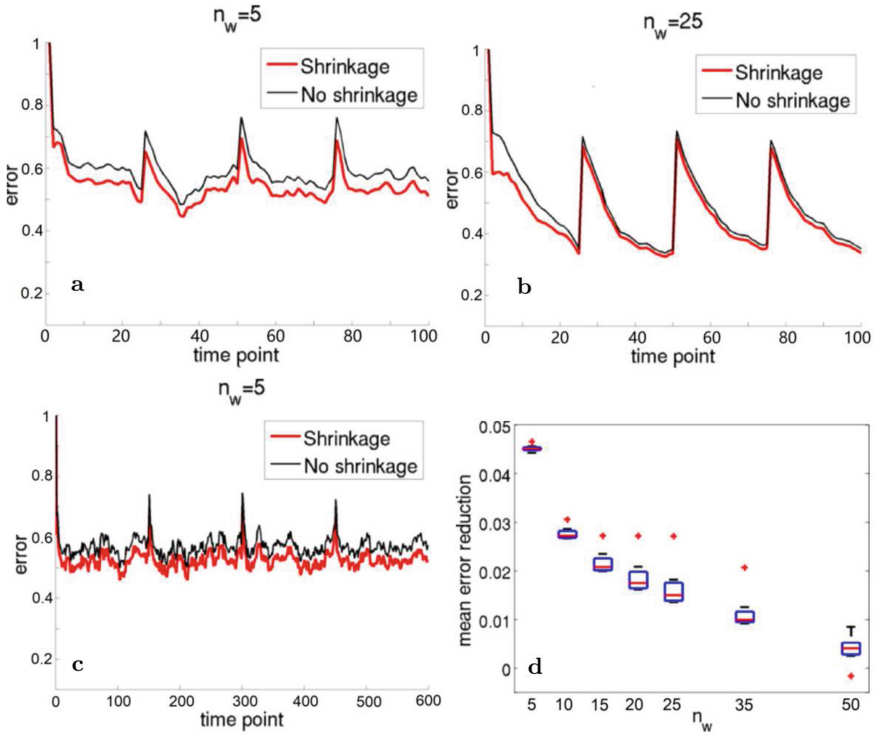

Fig. 1. Average of the Pearson distances measured for the data sets with 25 time points per brain state between the ground-truth and the dFCs estimated by EWMA, which we refer to as the "estimation error" for (a) $n_w = 5$, (b) $n_w = 25$. (c) for 150 time points per brain state and $n_w = 5$. (d) estimation error measured without covariance shrinkage minus the estimation error with covariance shrinkage. Covariance shrinkage consistently reduces the estimation error. This improvement is larger for small n_w, i.e., short time windows.

The total group comprising 73 female and 89 male participants, ranged in age from 23–80 years. Furthermore three different acquisition sequences were used to acquire the data. The fMRI repetition time (TR) of these protocols was 2 s for 61 subjects, 2.2 s for 67, and 2.648 s for the remaining 34. The processing of the rs-fMRI scan consisted of spatial smoothing with 5 mm FWHM, temporal de-trending, and band-pass filtering between 0.01 and 0.1 Hz. The processed scans were non-rigidly registered to an atlas of 111 regions of interest [10]. We computed the average BOLD signal inside each region to produce low-resolution time series. Furthermore, we kept the high-resolution time series, which consisted of the 175473 BOLD series within the gray matter. As a result, the spatial resolution of our high resolution series was thousand times finer than that of typical dFC studies [9,11].

We created a synthetic data set by first defining four distinct brain states. To do so, we computed for each subject the correlation matrix of the low-resolution time series. We selected for the ground-truth brain states the four correlation matrices with the largest ℓ_2 distances to each others. To create a synthetic rs-fMRI scan, we generated for each state a random band-passed signal by (i) creating a random timeseries, with a sampling rate of 0.5 Hz, by randomly sampling from the normal Gaussian distribution $N(0, 1)$, (ii) band-pass filtering these time series using Butterworth filter of order 2 and a band-pass filter of $[0.01, 0.1]$ Hz, and (iii) multiplying the square root of the state correlation matrix with the filtered time series. Lastly, we concatenated the four random band-passed signals to create a synthetic rs-fMRI sequence. This process was repeated 10 times for each data set. We created 8 of these data sets with different number of time points per brain state: 25, 50, 75, 100, 125, 150, 200 and 250.

3.2 Implementations

The baseline for our experiments was defined by an implementation of DCC, which was applied to all our data sets. Specifically, we first filtered rs-fMRI time series via GARCH. Next, we estimated the dFC by applying our efficient implementation of EWMA without covariance shrinkage. Note, we ensured on the synthetic data set that the outcome of that implementation was exactly the same as the EWMA implementation originally proposed in [6]. This implementation was applied to each data set seven times with the effective number of samples n_w being $\{5, 10, 15, 20, 25, 35, 50\}$. We then repeated these computations a second time using our proposed approach, the efficient EWMA implementation with covariance shrinkage.

3.3 Findings on the Synthetic Data

Figure 1 plots the error of each approach with respect to the synthetic data set consisting of brain states of 25 time points and the effective sample size being (a) $n_w = 5$ and (b) $n_w = 25$. The error is defined by the Pearson distance between the estimated and ground-truth covariance matrices averaged across the corresponding 10 scans. Our proposed approach exhibits a lower error for almost all time points and the difference to the baseline was especially large for the smaller effective sample size. The results obtained with different number of time points per brain states were very similar, as shown for instance by Fig. 1(c). The box plot of Fig. 1(d) confirms this finding as we observe better estimations using covariance shrinkage for all the parameter values. The improvement is larger for the smallest n_w, which corresponds to the shortest dFC patterns. These results strongly advocate for the use of covariance shrinkage when using EWMA to estimate dFC.

3.4 Results on the Real rs-fMRI Data

Figure 2 plots the distance matrix D for both approaches with respect to the high resolution data and three effective number of samples n_w. For all three settings,

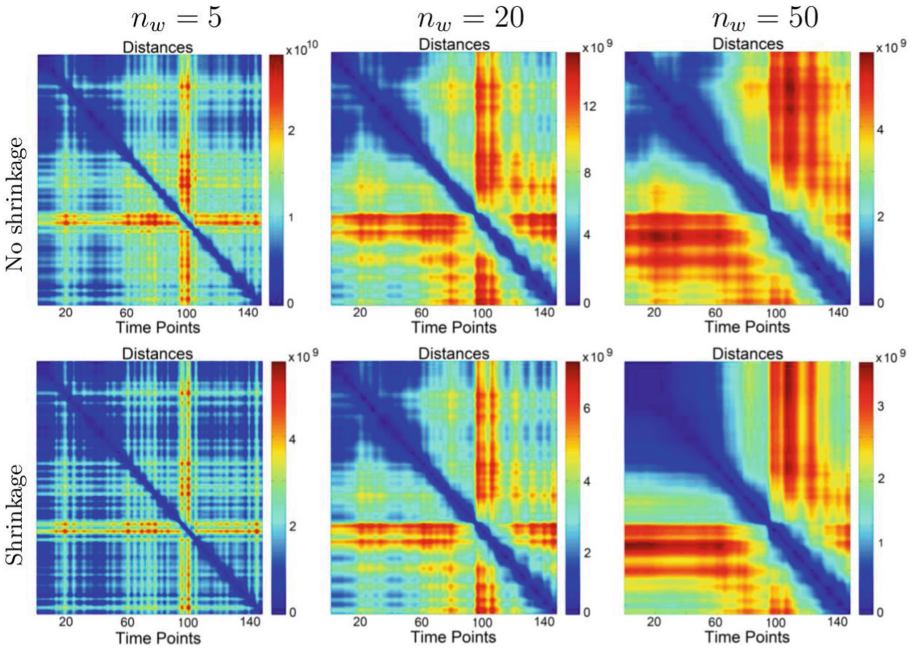

Fig. 2. The distance matrix D based on the dFC estimation of both approaches applied to the high resolution scans with different effective sample sizes n_w. Covariance shrinkage generates sharper distance matrices which would yield more reliable clusterings.

the pattern for the distance matrices computed by the proposed approach is sharper, especially for the smallest n_w. This observation suggests that clustering dFC into distinct mind states is easier when using covariance shrinkage.

We confirmed these qualitative findings by computing the quartile coefficient of dispersion (QCD) of all the distance matrices D obtained during our experiments. To compute the QCD, we first computed the first quantile $Q1$ and the third quartile $Q3$ of the distribution defined by the entries of D. We then defined

$$QCD := \frac{Q3 - Q1}{Q3 + Q1}. \tag{11}$$

We chose this metric as it is robust against outliers and is confined to $[-1, 1]$, which allows us to compare distance matrices of different magnitudes. More importantly, the larger the QCD value is, the easier it is to partition the matrix into distinct clusters, and the easier it is to separate dFCs into mind states. Figure 3(a) plots the QCDs of the two approaches when applied to the high resolution scans with the effective sample size of $n_w = 5$. It is striking that the QCD values obtained using our method are always higher than the ones derived using DCC, which indicates that dFC clustering would result in more reliable mind states.

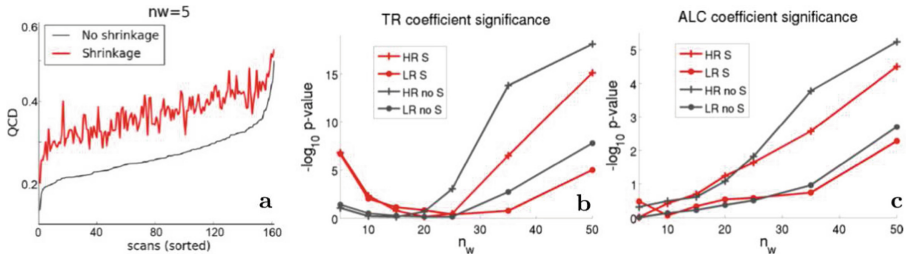

Fig. 3. (a) QCD measures obtained by applying the two approaches to all high-resolution timeseries with the effective sample size being $n_w = 5$. To ease visualization, the scans are sorted according to their QCD values under the baseline approach. (b–c) negative log p-value of the correlation between QCD and fMRI repetition time (TR) and alcohol diagnosis (ALC) for high-resolution data (HR) and low-resolution data (LR) as well as with shrinkage (S) or without covariance shrinkage (no S).

For both high resolution and low resolution clinical data sets, we also modelled the relation between QCD and explanatory variables by computing the following generative additive model

$$QCD \sim \alpha + \beta_0 sex + \beta_1 age + \beta_2 ALC + \beta_3 TR. \tag{12}$$

We then plotted in Fig. 3 the negative log of the p-value for those explanatory variables, whose correlation with QCD is significant ($p < 0.05$). Specifically, the QCDs generated by both implementations recorded significant correlations with alcohol diagnosis ALC and acquisition protocol TR. Figure 3(b) and (c) show agreement between the results obtained for low and high resolution data sets. Specifically, for small effective sample sizes the impact of the explanatory variables on the QCDs is similar for high and low resolution data. For large effective sample sizes, the effect of the explanatory variables on the QCD is always higher for the baseline approach. Assuming that the results obtained for our synthetic data set translate to real fMRI data, these results would suggest that the findings generated by the approach without covariance shrinkage, i.e., DCC, overestimate the impact of explanatory variables for large effective sample sizes.

3.5 Networks Extraction from High Resolution Data

For the final experiment, we extracted the Default Mode Network (DMN) from the dFC by slightly modifying both approaches. Specifically, we confined the computations of Eq. 2+3 to the correlations between the average BOLD signal within the Posterior Cingulate Cortex (PCC) and all the gray matter voxels in the atlas. Of the two approaches, Fig. 4 reveals that the Default Mode Networks extracted from a small number of dFC matrices by our method are closer to the typical DMN (scan average). Indeed the Spearman correlation with respect to the scan average was 0.72 on the high resolution data and 0.85 on the low resolution

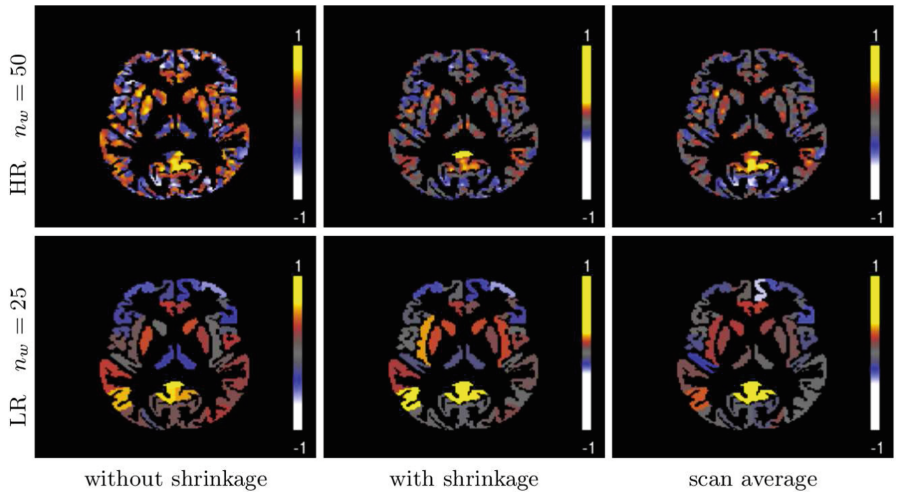

Fig. 4. Default mode network (DMN) extraction by averaging over 5 randomly selected dFCs estimated without (Left) and with shrinkage (Middle); Standard DMN extraction by averaging over the entire time series (scan average). On both low and high resolution data, the networks generated from the approach with shrinkage are much closer to the scan average than without shrinkage.

data for the approach with covariance shrinkage. The correlations drop down to 0.61 on the high-resolution data and 0.75 on the low resolution data for the approach without covariance shrinkage. In summary, this experiment suggests that covariance shrinkage improves functional networks extraction when only a few BOLD measurements are available for the estimation.

4 Conclusion

In this paper, we demonstrated that linear covariance shrinkage improved the estimation of dynamic resting-state fMRI connectivity, in particular when short brain connectivity states are extracted. We showed how the computation of large covariance matrices can be averted to scale the processing to full resolution fMRI scans, which suppressed the need of defining summary statistics for functional brain atlases regions. Our experiments on synthetic data demonstrate that our approach produced dFC estimates that are closer to the ground-truth. On real data, we found that our approach is better at extracting functional networks and capturing differences in rs-fMRI acquisition and diagnosis.

Acknowledgements. This research work was funded by the National Institute on Alcohol Abuse and Alcoholism (NIAAA) under the grants AA005965, AA013521, AA010723, and AA026762.

References

1. Biswal, B., Zerrin Yetkin, F., Haughton, V., Hyde, J.: Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. Magn. Reson. Med. **34**(4), 537–541 (1995)
2. Chang, C., Glover, G.: Time-frequency dynamics of resting-state brain connectivity measured with fMRI. NeuroImage **50**(1), 81–98 (2010)
3. Chen, Y., Wiesel, A., Eldar, Y.C., Hero, A.O.: Shrinkage algorithms for MMSE covariance estimation. IEEE Trans. Signal Process. **58**(10), 5016–5029 (2010)
4. Hofmann, T., Schölkopf, B., Smola, A.: Kernel methods in machine learning. Ann. Stat. **36**(3), 1171–1220 (2008)
5. Ledoit, W.: Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. J. Empirical Finan. **10**(5), 603–621 (2003)
6. Lindquist, M., Xu, Y., Nebel, M., Caffo, B.: Evaluating dynamic bivariate correlations in resting-state fMRI: a comparison study and a new approach. NeuroImage **101**, 531–546 (2014)
7. Liu, X., Zhang, N., Chang, C., Duyn, J.: Co-activation patterns in resting-state fMRI signals. NeuroImage **180**(Part B), 485–494 (2018)
8. Pfefferbaum, A., et al.: Accelerated aging of selective brain structures in human immunodeficiency virus infection. Neurobiol. Aging **35**(7), 1755–1768 (2014)
9. Preti, M., Bolton, T., Van De Ville, D.: The dynamic functional connectome: state-of-the-art and perspectives. NeuroImage **160**, 41–54 (2017)
10. Rohlfing, T., Zahr, N., Sullivan, E., Pfefferbaum, A.: The SRI24 multichannel atlas of normal adult human brain structure. Hum. Brain Mapp. **31**(5), 798–819 (2014)
11. Xie, H., et al.: Efficacy of different dynamic functional connectivity methods to capture cognitively relevant information. NeuroImage **188**, 502–514 (2019)