

Modeling Dynamic Functional Connectivity using a Wishart Mixture Model

Søren F.V. Nielsen^{*}, Kristoffer H. Madsen^{*,†},
Mikkel N. Schmidt^{*} and Morten Mørup^{*}

^{*} DTU Compute, Technical University of Denmark, Denmark

[†]Danish Research Centre for Magnetic Resonance,
Copenhagen University Hospital Hvidovre, Denmark

June 21, 2017

Abstract

Dynamic functional connectivity (dFC) has recently become a popular way of tracking the temporal evolution of the brains functional integration. However, there does not seem to be a consensus on how to choose the complexity, i.e. number of brain states, and the time-scale of the dynamics, i.e. the window length. In this work we use the Wishart Mixture Model (WMM) as a probabilistic model for dFC based on variational inference. The framework admits arbitrary window lengths and number of dynamic components and includes the static one-component model as a special case. We exploit that the WMM framework provides model selection by quantifying models generalization to new data. We use this to quantify the number of states within a prespecified window length. We further propose a heuristic procedure for choosing the window length based on contrasting for each window length the predictive performance of dFC models to their static counterparts and choosing the window length having largest difference as most favorable for characterizing dFC. On synthetic data we find that generalizability is influenced by window length and signal-to-noise ratio. Too long windows cause dynamic states to be mixed together whereas short windows are more unstable and influenced by noise and we find that our heuristic correctly identifies an adequate level of complexity. On single subject resting state fMRI data we find that dynamic models generally outperform static models and using the proposed heuristic points to a window-length of around 30 seconds provides largest difference between the predictive likelihood of static and dynamic FC.

^{*}Corresponding email: sfvn@dtu.dk. Søren F.V. Nielsen, Mikkel N. Schmidt and Morten Mørup were supported by Lundbeckfonden (fellowship grant R105-9813 to Morten Mørup). Kristoffer H. Madsen was supported by a Novo Nordisk Foundation Interdisciplinary Synergy Grant (NNF14OC0011413)

1 INTRODUCTION

It is a well known fact that the brain's way of integrating and segregating information changes over time and is perturbed by various cognitive tasks. In functional magnetic resonance imaging (fMRI) the functional brain "network" is often described using functional connectivity (FC) models, i.e. the correlation between segregated regions of interest, and these networks are known to change during task. In recent years studies of resting-state FC have shown to also exhibit dynamic properties indicating that FC during rest is non-stationary. Thus tracking temporal changes in FC during resting-state has become a popular research topic in recent years [1, 2, 3]. We see two main advantages of modeling the temporal changes in FC; first there has been some evidence that the dFC can be used to characterize different psychiatric disorders, such as PTSD [4] and schizophrenia [5]. Secondly, we hope by modeling dynamic functional connectivity (dFC) to gain a better understanding of the resting-state and the spontaneous changes in coupling between regions not associated by task activation [6].

Most models of dFC use the *sliding-window* approach [6, 7], where the assumption is that the FC is stable in subsegments of the data. Allen et. al. [6] applied this to a large cohort of healthy subjects where the extracted region-by-region covariance matrices from each window were clustered using the k-means algorithm. The 7 states extracted showed a highly non-stationary behavior where coupling in and to the default mode network notably varied over the states. We are though still faced with a number of problems in estimating reliable dFC patterns [8, 9, 10]. On one hand we face the problem of timescales, i.e. window length of dFC patterns is in most cases assumed to be known or set to some value based on the acquisition parameters in the experiment. Window-free methods exist, such as hidden Markov models in the context of microstates for EEG/MEG [11] and for dFC in fMRI [12, 13]. However, these models are more expressive and thus it becomes of importance to control the "over-characterization" in the training. Finally, in both microstate analysis and dFC models the complexity, i.e. the number of states to extract, is always a problem [13]. In k-means we have no natural way to choose the number of states, and thus heuristics such as the Gap-criterion is often used.

In this work we will use a Bayesian formulation of the Wishart mixture model (WMM) [14, 15, 16]. The Wishart distribution is defined as the distribution of the scatter matrix of zero-mean multivariate Gaussian data, and is thus a natural likelihood function for windowed functional connectivity. Whereas [14] used expectation-maximization (EM) and [16] Gibbs sampling we presently consider variational inference and use the WMM as a probabilistic analogy to the sliding-window k-means clustering approach. The probabilistic treatment will allow us to tap into features of the Bayesian modeling framework such as prediction. We will investigate a predictive likelihood framework to estimate the number of states in dFC problems, both in a synthetic setting and in resting-state fMRI data and propose a heuristic for choosing the window length.

2 METHODS

We first briefly present some notation. Let $\mathbf{x}_t \in \mathbb{R}^p$ be a zero-mean distributed signal at time point t . We consider data of L symmetric-positive semi-definite matrices C_ℓ of size $p \times p$ in which there are K clusters. In this paper we will use Gram matrices, i.e. $C_\ell = \sum_{t \in W_\ell} \mathbf{x}_t \mathbf{x}_t^T$, in which W_ℓ is the ℓ 'th window set.

2.1 Bayesian Wishart Mixture Model

The Bayesian Wishart Mixture Model (WMM) for K states can be written in terms of the generative model,

$$\begin{aligned} \boldsymbol{\pi} &\sim \text{Dir}(\boldsymbol{\alpha}) \quad z_\ell \sim \text{Cat}(\boldsymbol{\pi}) \quad \eta \sim \mathcal{G}^{-1}(a_0, b_0) \\ \boldsymbol{\Sigma}^{(k)-1} &\sim \mathcal{W}(\eta \mathbf{I}_p, \nu_0) \quad \mathbf{C}_\ell \sim \mathcal{W}(\boldsymbol{\Sigma}^{(z_\ell)}, \nu_\ell), \end{aligned} \quad (1)$$

in which $\boldsymbol{\pi}$ is a vector of length K containing the proportions of the states, $\boldsymbol{\alpha}$ is the prior vector of length K for the Dirichlet distribution, z_ℓ is the categorical (hard) assignment of window ℓ , η is the prior on the "scale" of the cluster centres, $\boldsymbol{\Sigma}^{(k)-1}$ is the k 'th cluster centres inverted Gram matrix, ν_0 is the degrees of freedom for each cluster centre and ν_ℓ is the degrees of freedom for the ℓ 'th window. The prior on the cluster centres, $\boldsymbol{\Sigma}^{(k)-1}$, and the parameter η is mostly chosen for convenience in the inference procedure and makes all updates closed form. The η parameter works as a regularizer on the cluster centres. This becomes very important in high dimensions and a low number of data points, as is the case in most fMRI scenarios. As for the degrees of freedom for each window, ν_ℓ , we propose summing the window-function, i.e. yielding ν_ℓ equal to the window length for the box-car window.

2.2 Variational Bayes

As with many Bayesian models evaluating the posterior, $p(\boldsymbol{\theta}|\mathbf{X})$, is intractable due to the model evidence term, $p(\mathbf{X})$, obtained from Bayes rule. We therefore turn to the variational Bayesian (VB) framework to approximate the posterior. In VB the goal is to find a distribution, $Q(\boldsymbol{\theta})$, which is "close" in the Kullback-Leibler (KL) divergence to the posterior and has a simpler form such that inference becomes tractable. We choose to use the well-known mean-field approximation in which the distribution of each parameter is assumed to factorize. Minimizing the KL-divergence between the intractable posterior and the Q -distribution is equivalent to maximizing the evidence lower-bound (ELBO), which can be formulated as,

$$\begin{aligned} \mathcal{L}(\mathbf{C}, \boldsymbol{\theta}) &= \langle \log p(\mathbf{C}|\boldsymbol{\Sigma}^{-1}, \mathbf{z}, \boldsymbol{\pi}) \rangle + \langle \log p(\boldsymbol{\Sigma}^{-1}) \rangle + \langle \log p(\eta) \rangle \\ &\quad + \langle \log p(\mathbf{z}|\boldsymbol{\pi}) \rangle + \langle \log p(\boldsymbol{\pi}) \rangle - \langle \log Q(\boldsymbol{\Sigma}^{-1}) \rangle \\ &\quad - \langle \log Q(\eta) \rangle - \langle \log Q(\mathbf{z}) \rangle - \langle \log Q(\boldsymbol{\pi}) \rangle, \end{aligned} \quad (2)$$

in which \mathcal{L} is the ELBO, \mathbf{C} is the collection of all the windowed scatter matrices, $\boldsymbol{\Sigma}^{-1}$ is the collection of all $\boldsymbol{\Sigma}^{(k)-1}$, \mathbf{z} is a vector of length L containing all z_ℓ , $\langle \cdot \rangle$ denotes expectation wrt. the Q -distribution, and $\boldsymbol{\theta}$ is the collection of all parameters in the model. Now we maximize the ELBO using coordinate ascend

variational inference (CAVI), which results in calculating the sufficient statistics of each factor (all of which are closed-form) while keeping the others fixed in a cyclic fashion. The Q-distributions along with the update rules have the form,

$$Q(\boldsymbol{\Sigma}^{-1}) = \prod_k \mathcal{W}(\boldsymbol{\Sigma}^{(k)-1} | \boldsymbol{\Omega}^{(k)}, v_k) \quad (3)$$

$$\boldsymbol{\Omega}^{(k)} = \left(\boldsymbol{\Sigma}_0^{-1} + \sum_{\ell} \langle z_{\ell k} \rangle \mathbf{C}_{\ell} \right)^{-1}, v_k = \nu_0 + \sum_{\ell} \langle z_{\ell k} \rangle \nu_{\ell}$$

$$Q(\eta) = \mathcal{G}^{-1}(\eta | \tilde{a}, \tilde{b}) \quad (4)$$

$$\tilde{a} = a_0 + \frac{\nu_0 p K}{2}, \tilde{b} = b_0 + \frac{1}{2} \sum_k \text{tr}(\langle \boldsymbol{\Sigma}^{(k)-1} \rangle)$$

$$Q(\mathbf{z}) = \prod_{\ell} \text{Cat}(\mathbf{z}_{\ell} | \mathbf{r}_{\ell}) \quad (5)$$

$$\tilde{r}_{\ell k} = \frac{\nu_{\ell} - p - 1}{2} \ln |\mathbf{C}_{\ell}| - \frac{\nu_{\ell}}{2} \ln(2) - \ln \Gamma_p\left(\frac{\nu_{\ell}}{2}\right)$$

$$- \frac{\nu_{\ell}}{2} \langle \ln |\boldsymbol{\Sigma}^{(k)}| \rangle - \frac{1}{2} \text{tr}(\langle \boldsymbol{\Sigma}^{(k)-1} \rangle \mathbf{C}_{\ell}) + \langle \ln \pi_k \rangle$$

$$r_{\ell k} = \frac{\exp(\tilde{r}_{\ell k})}{\sum_{k'} \exp(\tilde{r}_{\ell k'})}$$

$$Q(\boldsymbol{\pi}) = \text{Dir}(\boldsymbol{\pi} | \mathbf{a}), \quad a_k = \sum_{\ell} \langle z_{\ell k} \rangle + \alpha_k \quad (6)$$

In all experiments the following parameters were fixed: $\boldsymbol{\alpha} = [1, 1, \dots, 1]$ and $\nu_0 = p$. Note that if \mathbf{C}_{ℓ} does not have full rank some terms in (5) cannot be computed. These terms are however constant over k and can thus be ignored.

2.3 Predictive Likelihood and Bayes Factors

It is a well-known fact that VB is prone to underestimating the posterior variance [17], and therefore we do not usually rely only on the ELBO to do model selection. Thus, we need a more conservative measure that promotes generalizable models. We thus turn to predictive likelihood on previously unseen data, \mathbf{C}^* , which is dependent on the choice of the number of states K , i.e.

$$p(\mathbf{C}^* | \mathbf{C})_K = \int p(\mathbf{C}^* | \boldsymbol{\theta})_K p(\boldsymbol{\theta} | \mathbf{C})_K d\boldsymbol{\theta} \quad (7)$$

Since we do not have access to the true posterior, we use the approximation $Q(\boldsymbol{\theta})_K \approx p(\boldsymbol{\theta} | \mathbf{C})_K$, and due to the structure of the likelihood and the Q-distribution the approximation can be calculated analytically. In the following we will run the inference for a different number of states and calculate the predictive log Bayes factor, BF_k , towards the static model (with $K = 1$),

$$BF_k = \log p(\mathbf{C}^* | \mathbf{C})_k - \log p(\mathbf{C}^* | \mathbf{C})_1 \quad (8)$$

2.4 Generating Synthetic Data

To investigate the models capabilities and to validate our implementation we ran a number of synthetic experiments. In the following sections we will refer to

data as being "synthetic" meaning that data generated by the following process. First, we generate K random upper triangular matrices \mathbf{R}_k of dimension $p \times p$ by drawing each non-zero element of \mathbf{R}_k from a standard normal distribution $\mathcal{N}(0, 1)$. In all of our experiments the number of states K was set equal to three. Now we fix a "true" window length, w_α , and for each subsegment of the synthetic data first draw a random state (i.e. a number from $1..K$) and then w_α observations from $\mathcal{N}(\mathbf{0}_p, \mathbf{R}_k^T \mathbf{R}_k)$. This yields a data matrix, $\mathbf{X}_{\text{signal}}$, of size $p \times T$. Finally, we generate white noise, $\mathbf{X}_{\text{noise}} \sim \mathcal{N}(\mathbf{0}_p, \mathbf{I}_p)$, and create a linear combination of the data and noise to control the signal-to-noise (SNR) ratio, i.e. $\mathbf{X} = \gamma \mathbf{X}_{\text{signal}} + (1 - \gamma) \mathbf{X}_{\text{noise}}$. We do this process independently for the training and test data.

2.5 Resting State Data

We analyze the single subject dataset from [18] containing resting-state fMRI sessions collected over a period of 18 months. Using SPM 12¹ revision 6685, we applied the following preprocessing steps to sessions 014-104. All resting-state sessions were coregistered to the first image of session 014. We jointly corrected all sessions for motion artefacts using a rigid-body transformation towards the mean volume. An anatomical image from session 012 (T1 weighted) was coregistered to the functional space and a tissue probability map for grey matter (GM), white matter (WM) and cerebrospinal fluid (CSF) extracted using the standard map from SPM12. Next we applied bandpass filtering [0.009-0.08] Hz, nuisance regression (motion parameters, eroded CSF and WM masks) and wavelet despiking [19]. The images were then resliced (due to a change in the number of slices after session 027) to the first session and smoothed using a FWHM 5mm Gaussian kernel. After preprocessing we ran a group ICA using the GIFT software² version 4.0a using 85 components, the ERBM algorithm and otherwise default settings.

3 RESULTS

3.1 Synthetic Data

To investigate the influence of window length and SNR on the predictive framework, we conduct a synthetic study with the following fixed parameters: $w_\alpha = 10$, $p = 10$, $T = 10000$ and fixed $\eta^{-1} = 1e - 4$ in the model. We restarted each model inference 10 times and varied the number of states $K = 1..10$. Furthermore, we repeated the data generation process 10 times, and the mean BF_k over data sets (including standard deviation as error bars) can be seen in figure 1. In the noise-less case (top-left, $\gamma = 1$), we note that when the window-length is sufficiently small ($w \leq 10$) the model estimates the true number of states $K = 3$. However, we see an overestimation when the window-length becomes larger than w_α . This can be explained by the data being very inhomogeneous, and longer window lengths will be penalized more due to the mixing of different states within a window. Thus the models with longer window lengths need more states to explain the data. This effect gradually disappears as we decrease the

¹<http://www.fil.ion.ucl.ac.uk/spm/>

²<http://mialab.mrn.org/software/gift/index.html>

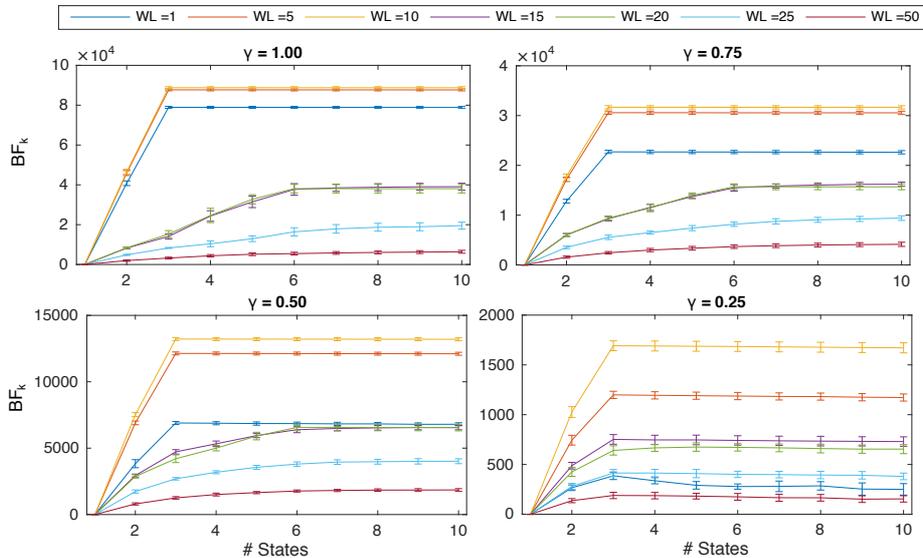


Figure 1: Synthetic experiments for the influence of window length. We generated 10 synthetic data sets with states that had a "true" window length of $w_\alpha = 10$ for different SNR levels $\gamma = [1, 0.75, 0.5, 0.25]$. Then we ran our predictive likelihood framework for different window lengths, and we plot the mean and standard deviation of the BF_k over data sets.

SNR. One thing to note for all SNR levels is that the window lengths that are shorter than or equal to w_α seem to have a larger increase in Bayes factor from the static $K = 1$ state model to the best predictive performance saturating at the $K = 3$ state model when compared to the longer window lengths. However, when we reach a certain noise level (in this case $\gamma = 0.25$) the shortest window length, $WL = 1$ has a flatter curve. Thus we find ourselves in a trade-off between window length and SNR; we want to make the window length low enough in order to not mix states together, but on the other hand not too low such that the estimation becomes unstable.

3.2 Single Subject Resting-State fMRI

To test the predictive framework on real data, we analysed a single subject resting-state fMRI data from [18]. Due to computational complexity, we ran the inference on 10 random subsets of data, each containing 5 sessions, and then calculated the predictive likelihood on the remaining sessions. It should be noted that some of the training sessions were in multiple training subsets. Each inference was restarted 10 times. Furthermore, we fixed η^{-1} during inference but varied its value in the range $[10^{-5}, 10^5]$ (sampled at ten points equidistantly in the log-domain). We choose the η^{-1} -value that yields the best predictive likelihood. The mean BF_k (over subsets) as a function of the number of states in the model is shown in figure 2. We see that the lowest window length has an almost flat curve, meaning that all number of states is equally likely, indicating that the window length is too short. The window length having the highest contrast between static and dynamic modeling, thus having the greatest increase

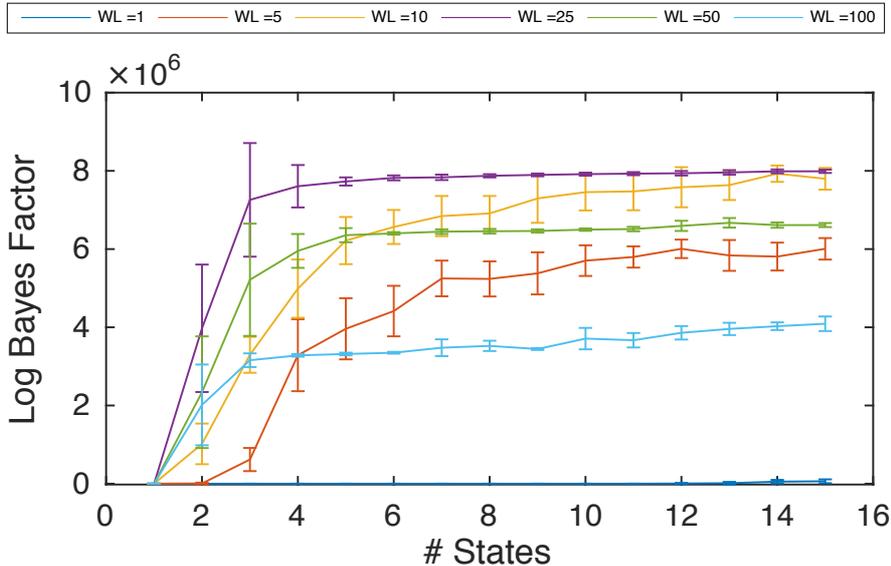


Figure 2: Experiment on resting-state fMRI data from a single subject. We randomly sampled 5 resting-state sessions, trained the VB-WMM model with a different number of states $K = 1..15$ and calculated the predictive likelihood on the remaining sessions. We repeated this process 10 times with new random training subsets. In the figure the mean BF_k (over training splits) on the held out sessions is plotted along with one standard deviation as errorbars. The entire analysis was done for different window lengths (WL) in TRs, $WL = [1, 5, 10, 25, 50, 100]$.

in log Bayes factor before hitting a plateau is $WL = 25$ (i.e., around 30 seconds).

4 DISCUSSION & CONCLUSION

We have proposed the Wishart mixture model (WMM) as a probabilistic extension of windowed k-means, to model dynamic functional connectivity in fMRI. As a way to quantify the number of states best accounting for dFC we use the predictive likelihood. We further proposed a heuristic based on contrasting the predictive likelihood of dFC to the predictive likelihood of the corresponding static model containing only one state in order to quantify a suitable window length for characterizing dFC. On synthetic data we found that this heuristic correctly indicated the correct level of complexity. On real single subject resting state data we found support for dynamic modeling for all considered window lengths except ($WL = 1$ where the static model ($K = 1$) was not outperformed by dynamic models ($K > 1$)) and using the heuristic of highest contrast in predictive likelihood between static and dynamic modeling we found $WL = 30$ most suitable for characterizing dFC.

We would like to emphasize that the proposed procedure for quantifying window-length is a heuristic that we find useful to quantify trade-offs between SNR and issues mixing dynamic states but that predictive performance using different window lengths cannot be directly compared as they are based on test

data having different properties. It should also be noted that as we increase the window length, we have fewer and fewer data points for training the WMM, which could influence the results. One could look into using overlapping windows to mitigate the effect of mixing states together, which is an avenue to pursue in future work. Also, in this work we have used Gram matrices to "stay true" to the likelihood function we are using in the WMM. However, there might be differences in using covariance matrices or even correlation matrices, which should be investigated further. In the real data there could be a pitfall caused by noisy ICA components, i.e. we have not done any post-selection. However, if states were driven by noise components they are not likely to generalize well to new sessions, and the predictive likelihood should in theory take care of this. However, to really confirm this would require an in-depth analysis. Thus, the proposed heuristic needs to be further validated. In the long run, we would like to replicate the results on larger portions of the resting state data, which would require a faster implementation. This should be possible using massively parallel architectures such as graphical processing units, as there are steps in the algorithm that are parallelizable.

References

- [1] R. M. Hutchison et al., "Dynamic functional connectivity: promise, issues, and interpretations," *Neuroimage*, vol. 80, pp. 360–378, 15 Oct. 2013.
- [2] N. Leonardi and D. Van De Ville, "On spurious and real fluctuations of dynamic functional connectivity during rest," *Neuroimage*, vol. 104, pp. 430–436, 1 Jan. 2015.
- [3] M. Breakspear, "Dynamic models of large-scale brain activity," *Nat. Neurosci.*, vol. 20, pp. 340–352, 23 Feb. 2017.
- [4] J. Ou, L. Xie, C. Jin, X. Li, D. Zhu, R. Jiang, Y. Chen, J. Zhang, L. Li, and T. Liu, "Characterizing and differentiating brain state dynamics via hidden markov models," *Brain Topogr.*, vol. 28, pp. 666–679, Sept. 2015.
- [5] Y. Du et al., "Identifying dynamic functional connectivity biomarkers using GIG-ICA: Application to schizophrenia, schizoaffective disorder, and psychotic bipolar disorder," *Hum. Brain Mapp.*, 10 Mar. 2017.
- [6] E. A. Allen, E. Damaraju, S. M. Plis, E. B. Erhardt, T. Eichele, and V. D. Calhoun, "Tracking whole-brain connectivity dynamics in the resting state," *Cereb. Cortex*, vol. 24, pp. 663–676, Mar. 2014.
- [7] U. Sakoğlu, G. D. Pearlson, K. A. Kiehl, Y. M. Wang, A. M. Michael, and V. D. Calhoun, "A method for evaluating dynamic functional network connectivity and task-modulation: application to schizophrenia," *MAGMA*, vol. 23, pp. 351–366, Dec. 2010.
- [8] R. Hindriks, M. H. Adhikari, Y. Murayama, M. Ganzetti, D. Mantini, N. K. Logothetis, and G. Deco, "Can sliding-window correlations reveal dynamic functional connectivity in resting-state fMRI?," *Neuroimage*, vol. 127, pp. 242–256, 15 Feb. 2016.

- [9] S. Shakil, C.-H. Lee, and S. D. Keilholz, “Evaluation of sliding window correlation performance for characterizing dynamic functional connectivity and brain states,” *Neuroimage*, vol. 133, pp. 111–128, 4 Mar. 2016.
- [10] T. O. Laumann et al., “On the stability of BOLD fMRI correlations,” *Cereb. Cortex*, 2 Sept. 2016.
- [11] D. Vidaurre, A. J. Quinn, A. P. Baker, D. Dupret, A. Tejero-Cantero, and M. W. Woolrich, “Spectrally resolved fast transient brain states in electrophysiological data,” *Neuroimage*, vol. 126, pp. 81–95, 26 Nov. 2015.
- [12] S. Ryali, K. Supekar, T. Chen, J. Kochalka, W. Cai, J. Nicholas, A. Padmanabhan, and V. Menon, “Temporal dynamics and developmental maturation of salience, default and Central-Executive network interactions revealed by variational bayes hidden markov modeling,” *PLoS Comput. Biol.*, vol. 12, p. e1005138, Dec. 2016.
- [13] S. F. V. Nielsen, K. H. Madsen, R. Røge, M. N. Schmidt, and M. Mørup, “Nonparametric modeling of dynamic functional connectivity in fMRI data,” 4 Jan. 2016.
- [14] S. Hidot and C. Saint-Jean, “An Expectation–Maximization algorithm for the wishart mixture model: Application to movement clustering,” *Pattern Recognit. Lett.*, vol. 31, pp. 2318–2324, 15 Oct. 2010.
- [15] J. Korzen, K. H. Madsen, and M. Mørup, “Quantifying temporal states in rs-fMRI data using bayesian nonparametrics.” Poster presentation at Human Brain Mapping 2014, 2014.
- [16] A. Cherian, V. Morellas, and N. Papanikolopoulos, “Bayesian nonparametric clustering for positive definite matrices,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, pp. 862–874, May 2016.
- [17] C. M. Bishop, *Pattern recognition and machine learning*. Springer, 2006.
- [18] R. A. Poldrack et al., “Long-term neural and physiological phenotyping of a single human,” *Nat. Commun.*, vol. 6, p. 8885, 9 Dec. 2015.
- [19] A. X. Patel, P. Kundu, M. Rubinov, P. S. Jones, P. E. Vértes, K. D. Ersche, J. Suckling, and E. T. Bullmore, “A wavelet method for modeling and despiking motion artifacts from resting-state fMRI time series,” *Neuroimage*, vol. 95, pp. 287–304, 15 July 2014.