



## Bayesian structure learning for dynamic brain connectivity

Andersen, Michael Riis; Winther, Ole; Hansen, Lars Kai; Poldrack, Russell; Koyejo, Oluwasanmi

*Published in:*

Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics

*Publication date:*

2018

*Document Version*

Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

*Citation (APA):*

Andersen, M. R., Winther, O., Hansen, L. K., Poldrack, R., & Koyejo, O. (2018). Bayesian structure learning for dynamic brain connectivity. In *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics* (Vol. 84, pp. 1436-1446)

---

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

---

# Bayesian Structure Learning for Dynamic Brain Connectivity

---

**Michael Riis Andersen**  
Aalto University

**Ole Winther**  
Technical University of Denmark

**Lars Kai Hansen**  
Technical University of Denmark

**Russell Poldrack**  
Stanford University

**Oluwasanmi Koyejo**  
University of Illinois at Urbana-Champaign

## Abstract

Human brain activity as measured by fMRI exhibits strong correlations between brain regions which are believed to vary over time. Importantly, dynamic connectivity has been linked to individual differences in physiology, psychology and behavior, and has shown promise as a biomarker for disease. The state of the art in computational neuroimaging is to estimate the brain networks as relatively short sliding window covariance matrices, which leads to high variance estimates, thereby resulting in high overall error. This manuscript proposes a novel Bayesian model for dynamic brain connectivity. Motivated by the underlying neuroscience, the model estimates covariances which vary smoothly over time, with an instantaneous decomposition into a collection of spatially sparse components – resulting in parsimonious and highly interpretable estimates of dynamic brain connectivity. Simulated results are presented to illustrate the performance of the model even when it is mis-specified. For real brain imaging data with unknown ground truth, in addition to qualitative evaluation, we devise a simple classification task which suggests that the estimated brain networks better capture the underlying structure.

## 1 INTRODUCTION

Human brain function is studied using a variety of techniques in psychology, neurobiology and neuro-

science. Researchers are motivated by the desire to understand how brain function relates to behavior, and to neurological disorders. Recent research in computational neuroscience has shown that human brain activity, as measured by fMRI exhibits strong correlations between brain regions (Friston et al., 1993; Biswal et al., 1995; van den Heuvel and Hulshoff Pol, 2010; Bastos and Schoffelen, 2015; Fingelkurts et al., 2005; Li et al., 2009). Such correlations, known as *brain connectivity*, are thought to represent communication between distant neural populations, which is crucial to neural information processing. A more recent paradigm shift is the observation that brain connectivity exhibits dynamics – both in response to task demands and at rest (Sakoğlu et al., 2010; Chang et al., 2016). Importantly, time varying connectivity has been linked to individual differences in physiology, psychology and behavior (Shine et al., 2016a,b), and has shown promise as a biomarker for disease (Fox and Raichle, 2007; Calhoun et al., 2014).

From the statistical point of view, the measured fMRI signal corresponds to a non-stationary multivariate time series, and time varying connectivity corresponds the sequence of associated covariances. By far the most popular technique for time-varying connectivity estimation is the sliding window approach and its variants (Hutchison et al., 2013; Damaraju et al., 2014; Calhoun et al., 2014; Gonzalez-Castillo et al., 2015). This involves dividing the time series into overlapping sliding windows, then the connectivity is computed as the covariance within each windowed temporal region. On one hand, the sliding approach is susceptible to high variance when computed with small windows, and alternately large windows will tend to smooth out dynamics. In practice, the window size is often fixed a-priori between 10 - 100 samples based on recommendations in published papers (Hindriks et al., 2016). The next step of the sliding window analysis is to cluster the covariance estimates to determine a smaller set of canonical brain states (Allen et al.,

2014; Gonzalez-Castillo et al., 2015) – thereby, implicitly assuming that the range of brain connectivity can be summarized via a small set of states combined with discrete switching dynamics. While the sliding window remains the most popular approach in the literature, several authors have shown how with or without the clustering step, sliding window connectivity is susceptible to noise, resulting in false detections of connectivity where none exists (Lindquist et al., 2014; Hindriks et al., 2016; Laumann et al., 2016). It is clear that learning time-varying functional connectivity is inherently high dimensional, as it involves estimating many more parameters than the number of samples available within a typical brain scan. For instance, data with 100 time samples from 100 brain regions requires  $\mathcal{O}(100^3)$  parameters! Thus, naïve techniques generally result in noisy and inaccurate estimates. We explore two ways to improve the quality of the dynamic connectivity estimates, namely we (i) incorporate structure via neuroscience-inspired priors, and (ii) pool data with shared information across multiple samples i.e. hierarchical modeling.

**Neuroscience-inspired statistical modeling:** Our model is inspired by a variety of neuroscientific phenomena. We provide a brief outline, with further details in the text. Our model decomposes the connectivity into a weighted combination of sparse low rank components. Our approach is based on the scientific hypothesis that brain function can be decomposed into a small number of cognitive components (Posner et al., 1988). For similar reasons, we incorporate sparse priors for each component which encourage spatial localization into distinct brain regions. This approach also follows the success of interpretable sparse decomposition methods for describing brain function (Dohmatob et al., 2016). The time-varying weights are modeled as Gaussian processes to enable complex non-linear and long range correlations, which is based on the hypothesis that brain connectivity exhibits complex but temporally smooth non-linear dynamics (Shine et al., 2016a,b), and motivated by empirical evidence which suggests that brain dynamics are continuous rather than discrete switching (Smith et al., 2012). Further, the model also encourages sparsity of the component weights to automate model selection.

**Hierarchical Modeling:** Post-hoc analysis of functional connectivity in both task data and rest data from the same individual suggests that they involve similar sub-components, albeit with individual temporal activation (Arbabshirani and Calhoun, 2011; Di et al., 2013; Schultz and Cole, 2016). Similarly, static and time-varying functional brain structures are known to be preserved across participants (Zalesky et al., 2012; Preti et al., 2016; Calhoun et al., 2014),

with shared temporal dynamics when the participants are all involved in the same task with the same task temporal design. Such information are discarded by common methods such as sliding window connectivity. Instead, our approach incorporates the shared structure to improve the quality of estimates. Taken together, these ideas result in a novel Bayesian structure learning model for dynamic brain connectivity. In summary, our main contribution is a novel hierarchical model for time-varying brain functional connectivity – inherently a high dimensional static inference, which combines neuroscience-inspired statistical modeling with hierarchical modeling and data pooling for improved estimates as compared to standard approaches. We design and implement scalable mean-field variational inference algorithm for the model and evaluate the resulting performance using both synthetic data and real functional brain imaging data. Our results show improved quantitative and qualitative performance as compared to standard approaches.

## 2 Related work

Related work includes sliding window methods (Hutchison et al., 2013; Damaraju et al., 2014; Calhoun et al., 2014; Gonzalez-Castillo et al., 2015), which have been described in the introduction. Furthermore, state-space models (Yang et al., 2016; Olsson and Hansen, 2006), dictionary learning models Eavani et al. (2012), and independent component analysis models (Dyrholm et al., 2007; Hyvarinen and Oja, 2000) are also popular within neuroimaging. Hidden Markov models (Rabiner, 1989) are easily adapted to the task of estimating functional connectivity. However, such models assume that the underlying latent space is discrete. As we show, while such an approach can improve performance as compared to the sliding window, it does not seem well adapted to neuroimaging applications, and achieves worse performance. Dynamic factor models, somewhat related to our dynamic connectivity approach have been explored the statistics and machine learning literature (Fox and Dunson, 2015; Kastner, 2016), but with differences in the modeling assumptions which are less well adapted to neuroscience data e.g. strong autoregressive assumptions on the temporal dynamics. Further, as we employ variational inference instead of MCMC, our model is able to scale up to realistic brain imaging parcellated time series with hundreds of regions and hundreds of time points, and where the number of regions often exceeds the time series length. To our knowledge, our proposal is the first Bayesian continuous dynamic model with such high dimensional scalability.

### 3 GENERATIVE MODEL

Let  $\mathbf{x}_t^n \in \mathbb{R}^D$  be the  $D$ -dimensional observed time series at time  $t \in \{1, 2, \dots, T\}$  for subject  $n \in \{1, 2, \dots, N\}$  and let  $\mathcal{D}$  be the collection of the all observed time series for all subjects. The sampling distribution of  $\mathbf{x}_t^n$  is assumed to be  $\mathbf{x}_t^n \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_t^n)$ , where  $\boldsymbol{\Sigma}_t^n$  is the instantaneous covariance matrix time at  $t$  for subject  $n$ . Let  $\mathcal{S} = \{\mathbf{S}_k \mid \mathbf{S}_k \in \mathbb{R}^{D \times D}\}_{k=1}^K$  represents the *dictionary of covariance matrix components*. Our model decomposes  $\boldsymbol{\Sigma}_t^n$  into a non-negative weighted sum of components as:

$$\boldsymbol{\Sigma}_t^n = \beta^{-1} \mathbf{I} + \sum_{k=1}^K \alpha_{k,t}^n \mathbf{S}_k, \quad (1)$$

where the coefficients  $\alpha_{k,t}^n \geq 0$  are a set of non-negative real mixing weights. It is clear that that  $\alpha_{k,t}^n$ , where  $\mathcal{A}^n = \{\alpha_{k,t}^n\}$  also govern the dynamics, by controlling the contribution of each  $\mathbf{S}_k$  towards the instantaneous covariance  $\boldsymbol{\Sigma}_t^n$  at time  $t$  for the  $n$ 'th subject. The parameter  $\beta^{-1} > 0$  is a positive real number controlling the amount of additive white noise.

The dictionary of covariance matrix components,  $\mathcal{S}$  is shared across both time and subjects. Loosely speaking,  $\mathcal{S}$  is a common basis of covariance matrices for all time points and all subjects, where  $(\alpha_{1,t}^n, \dots, \alpha_{K,t}^n)$  are the coordinates for the specific covariance matrix at time  $t$  for the  $n$ 'th subject. This shared representation allows us to pool data across multiple subjects to estimate each  $\mathbf{S}_k$ . As we show, estimating each covariance matrix component,  $\mathbf{S}_k$ , using data from multiple subjects simultaneously rather than estimating them independently for each subject leads to more robust estimates. From the neuroscience perspective, each element of  $\mathcal{S}$  represents the connectivity matrix of a particular *cognitive process*, so that the instantaneous brain connectivity is a combination of these processes. Thus, our model assumes that the cognitive components are shared among individuals, and individual differences are due to dynamics. That the cognitive components are shared is a basic assumption in cognitive neuroscience and is considered a weak assumption.

In this work, we choose the prior distribution for  $\mathcal{A}$  based on the following three desired properties: sparsity, temporal smoothness and non-negativity. The properties of temporal smoothness and sparsity are inspired by the neuroscience as discussed in the introduction, and also serve as a way to regularize the model. Sparsity ensures that only a small subset of elements from the dictionary  $\mathcal{S}$  contribute to the instantaneous covariance  $\boldsymbol{\Sigma}_t^n$  at any given time  $t$ . This follows the assumption that only a subset of the cognitive processes are required to represent instantaneous connectivity.

Temporal smoothness of  $\boldsymbol{\alpha}_k^n = (\alpha_{k,1}^n, \alpha_{k,2}^n, \dots, \alpha_{k,T}^n) \in \mathbb{R}^T$  implies that  $\boldsymbol{\Sigma}_t^n$  will change slowly in time, i.e. two samples  $\mathbf{x}_t^n$  and  $\mathbf{x}_{t'}^n$  are more likely to have similar second order moments, if  $t$  and  $t'$  are close in time. This encodes the assumption of smoothly varying cognitive process dynamics. Finally, non-negativity is a technical condition to ensure that the instantaneous covariance matrix  $\boldsymbol{\Sigma}_t^n$  remains positive definite for all time points and for all subjects. We find that the positive constraint also simplifies interpretability of the overall covariance as a sum of parts. We model the set of mixing coefficients using Gaussian processes (Rasmussen and Williams, 2005) as follows

$$\alpha_{k,t}^n = \max(0, a_{k,t}^n), \quad \mathbf{a}_k^n \sim \mathcal{GP}(\mathbf{m}_k^n, \mathbf{C}_k^n). \quad (2)$$

That is, we model each  $\boldsymbol{\alpha}_k^n$  as a truncated Gaussian process with prior mean  $\mathbf{m}_k^n \in \mathbb{R}^T$  and prior covariance  $\mathbf{C}_k^n \in \mathbb{R}^{T \times T}$ . Using this construction, we can explicitly control the smoothness properties of  $\boldsymbol{\alpha}_k^n$  using the prior covariance matrix  $\mathbf{C}_k^n$ . Furthermore, the marginal probability of a given weight  $\alpha_{k,t}^n$  being non-zero is given by

$$p(\alpha_{k,t}^n > 0) = p(a_{k,t}^n > 0) = \Phi\left(\frac{\mu_{k,t}^n}{\sqrt{C_{k,tt}^n}}\right), \quad (3)$$

where  $\Phi: \mathbb{R} \rightarrow (0, 1)$  is the standardized normal cumulative distribution function.

Each  $\mathbf{S}_k$  is assumed to be sparse, symmetric and of rank one, i.e.  $\mathbf{S}_k = \mathbf{v}_k \mathbf{v}_k^T$ , where  $\mathbf{v}_k$  is a sparse vector – thus encoding the assumption that each cognitive process is "simple", and spatially localized. To encourage sparsity of  $\mathbf{v}_k$  we impose the spike-and-slab prior (Mitchell and Beauchamp, 1988) on  $\mathbf{v}_k$  as follows

$$\mathbf{v}_k = \mathbf{s}_k \circ \mathbf{u}_k, \quad (4)$$

$$\mathbf{s}_k \sim \prod_{i=1}^D \text{Bernoulli}(p_k), \quad \mathbf{u}_k \sim \prod_{i=1}^D \mathcal{N}(0, \tau_k),$$

where  $\circ$  is the element-wise Hadamard product,  $s_{k,i} \in \{0, 1\}$  is a binary support variable for  $v_{k,i}$ , and  $p_k \in (0, 1)$  is a hyperparameter controlling the degree of sparsity, i.e. the expected fraction of non-zero entries in  $\mathbf{v}_k$  is  $p_k$ .

**Adaptation for Task data:** We note that neuroimaging task data has the advantage of temporal synchronization i.e. it is reasonable to expect that all subjects will be in similar brain states at similar time-points during the experiment. Thus, we model each subject as i.i.d. observations of the same underlying spatio-temporal process. In particular, we assume that both the spatial maps and dynamic weights are

shared, i.e.  $\alpha_{k,t} = \alpha_{k,t}^n$  for all  $n$ . The simplified model becomes:

$$\Sigma_t^n = \beta^{-1} \mathbf{I} + \sum_{k=1}^K \alpha_{k,t} \mathbf{S}_k \quad \forall n. \quad (5)$$

**Inference and Learning:** We estimate the covariance matrix components  $\mathbf{S}$  and the dynamic mixing weights,  $\mathcal{A} = \{\mathcal{A}^n\}_{n=1}^N$ , simultaneously from the observed time series  $\mathcal{D}$  via mean-field variational inference (Blei et al., 2016). To estimate the quantities of interest, we use the posterior expectation of the random variables conditioned on the data, e.g.  $\hat{\mathbf{S}}_k = \mathbb{E}_Q[\mathbf{S}_k|\mathcal{D}]$ , where expectations are with respect to the variational distribution  $Q$ . A graphical representation of both the full and simplified model, and details of the inference are deferred to the Appendix.

## 4 EXPERIMENTAL EVALUATION

To study and quantify the performance of the model and the inference algorithm, we conducted a series of numerical experiments using both synthetic data and real fMRI data. In all experiments, we choose the mean function to be a constant, i.e.  $\mathbf{m}_k = m_k \mathbf{1} \in \mathbb{R}^T$ , and we choose the covariance function for  $\mathbf{C}_k$  to be the Matérn covariance (Rasmussen and Williams, 2005) function plus a scaled identity matrix so that  $\mathbf{C}_k(t, t')$  is given by:

$$c_k \left[ 1 + \frac{\sqrt{5}|t-t'|}{\ell_k} + \frac{5|t-t'|^2}{3\ell_k^2} \right] \exp \left[ -\frac{\sqrt{5}|t-t'|}{\ell_k} \right] + d_k \mathbf{I}.$$

Hyperparameters such as  $m_k$ ,  $c_k$ ,  $d_k$ , &  $\ell_k$  for each component are estimated from the training data using a maximization step i.e. the overall model is estimated via variational EM. Observe that due to the spatial sparsity on the mixing weights, our model automatically implements model selection (see figure 2c), so we only need to set an upper bound on the number of components  $K$ .

### 4.1 Simulated Data

First we investigate the performance of the model using simulated data. As we are interested in recovery of the covariance matrices (and not just likelihood), we use the Log-Euclidean Riemannian Metric (LERM) – a metric on the manifold of symmetric positive definite matrices (Vemulapalli and Jacobs, 2015; Huang et al., 2015). For a sequence of estimated covariance matrices, we compute the time-averaged LERM-distance to the ground truth sequence. Details of the LERM metric are provided in the Appendix.

#### 4.1.1 Continuous mixing data set

In the first experiment, we generated time series for a number of subjects using the model in eq. (5) assuming that all subjects share the same time-varying covariance structure. In particular, we generated a sequence of ground truth covariance matrices  $\Sigma_t$  using four components and 145 time points, i.e.  $K = 4$  and  $T = 145$ . We ran the experiment for three different dimensions,  $D = 10, 30, 50$ , respectively. The ground truth mixing weights for the four components are chosen to be a linear function, a constant function, and two sinusoidal functions with different frequencies (see Figure 1(a)).

As expected, we found that the sparsity-promoting priors are capable of pruning unnecessary covariance components. To demonstrate this, we initialized the model using  $K = 20$  random covariance components. The amount of energy that the  $k$ 'th component contributes to the total energy depends on both  $\mathbf{S}_k$  and  $\alpha_k$  and for each  $k$ , we computed the energy contribution for each component point-estimate as the expectation of  $E_k = \text{Trace}(\mathbf{S}_k) \sum_{t=1}^T \alpha_{k,t}$ . Figure 1(c) shows the energy contribution of each component along with the corresponding estimated component after fitting the model to a data set with  $N = 5$  subjects. It is seen that the model correctly identifies four true non-zero components. Figure 1(b) shows the mean of the posterior distribution for  $\alpha_k$  superimposed with 2 standard deviations (left panel) and the corresponding non-zero covariance components (right panel). There is a scaling ambiguity in the model, i.e. in eq. (5) we can scale  $\alpha_k$  with some non-zero constant and divide  $\mathbf{S}_k$  with the same constant to obtain the same covariance matrix  $\Sigma_t$ . To facilitate comparison with the ground truth values, we scale each estimate of  $\alpha_k$  such that the maximum value is 1.

We also compared the performance of the proposed model to competing methods from the neuroimaging literature. Specifically, we consider the sample covariance estimator (completely ignoring any dynamics), the sliding window estimator and the hidden Markov model using multivariate Gaussian emission distributions (HMM) (Rabiner, 1989). We considered three different window sizes for the sliding window estimator, i.e.  $L = 10, 20, 30$ , where  $L \approx 20$  is optimal w.r.t. parameter recovery. We note that the sliding window results shown reflect window sizes chosen to highlight the strongest baseline conditions. In practice, window selection is a notoriously hard problem that leads to radically different and non-reproducible results. For training of the HMMs, we choose the number of states based on log likelihood of data from one hold out subject.

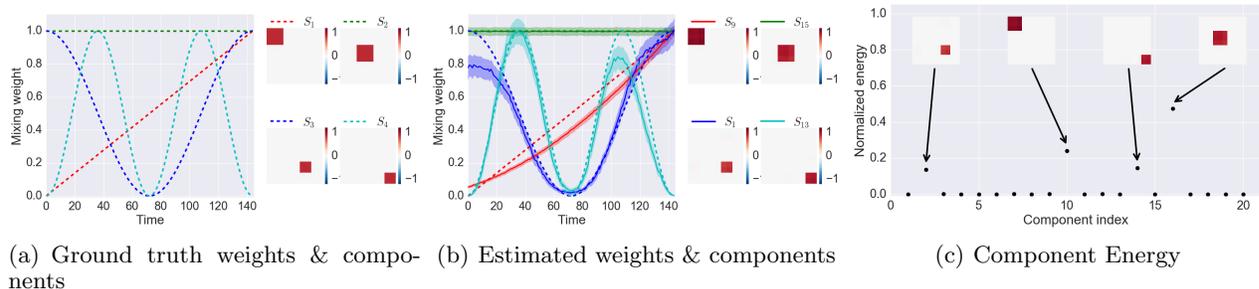


Figure 1: Ground truth and model estimates for simulated experiment using continuous mixing dynamics for  $D = 10$  and for  $N = 5$  subjects. The leftmost panel in figure (a) shows the true mixing weights, while the rightmost panel in figure (a) shows the corresponding ground truth components. The left and right panel in figure (b) show the estimated mixing weights  $\alpha_k$  and the estimated covariance components  $\mathbf{S}_k$  from the model. (c) Energy contribution for each component normalized wrt. total energy for continuous data set.

The three rows in Figure 2(a) show the performance of each method as a function of number of training subjects for three different values of number of regions, i.e.  $D = 10, 30, 50$ , averaged over  $R = 20$  realizations of the data. First, it is seen that the proposed method performs uniformly better than the reference methods. Furthermore, it is also seen that as the dimension of data  $D$  increases, the performance of the reference methods drop significantly. In particular, when the number of training subjects  $N < 10$  and  $D > 10$ , the sliding window estimators and the HMMs performs equal or worse than the sample covariance estimator.

#### 4.1.2 Simulated data with discrete switching

Inspired by the simulated experiments of Calhoun et al. (2014), we performed simulated data where the ground truth covariance switches instantaneously between a set of discrete states (rather than continuous mixing). As in (Calhoun et al., 2014), we considered four states with four different covariance matrices as shown in Figure 4. Using the same fixed state sequence of  $1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 2$  for each subject, we generated time series for each subject such that the samples within each state are drawn i.i.d. from a multivariate Gaussian distribution with a state-specific covariance matrix, i.e. the emission model of the HMM. Figure 3 shows the posterior mean for each  $\alpha_k$  for the non-zero components and the corresponding covariance components after training the model on  $N = 5$  subjects. It is seen that the model captures the dynamics and the covariance matrices for all states correctly, even though the estimated mixing weights are temporally smoothed. This is due to the fact that step functions are not well-modeled by the Gaussian process priors on  $\alpha_k$  using generic stationary kernels. The bottom-most panels in Figure 3 show how the model decomposes the four unique covariance matrices into 8 rank one components, where some of the components are used

in multiple states. For example, state 1 is decomposed into components  $\mathbf{S}_{16}$  and  $\mathbf{S}_{13}$ , while state 3 is decomposed into components  $\mathbf{S}_{13}$ ,  $\mathbf{S}_{10}$ , and  $\mathbf{S}_1$ . Thus, the samples from states 1 and 3 both contribute to the estimation of  $\mathbf{S}_{13}$  even though the complete covariance matrices for the two states are different. The decomposition of distinct states into a set of shared components aligns well with the neuroscientific hypothesis that brain function is decomposable into a set of elementary cognitive processes (Posner et al., 1988) as discussed in the introduction. The model also produces accurate estimates of the instantaneous covariance matrices as weighted sums of the estimated covariance components  $\hat{\mathbf{S}}$  as evidenced in Figure 2b. The proposed model is mis-specified in this experiment, however we observe that for  $D = 10$ , the proposed model outperforms the reference methods for  $N < 4$ , while it achieves the same level of performance as the HMM for  $N \geq 4$ . For  $D = 30, 50$ , the proposed model performs uniformly better than the reference methods. We explain this by observing that our model enjoys additional shrinkage from the priors that are not available to HMM. The effect is especially stark in high dimensions (as one expects in neuroimaging data). We conjecture that one could improve the HMM estimates by including additional shrinkage, but this is beyond the scope of this paper.

#### 4.2 Analysis of fMRI motor task data

In this experiment, we applied the proposed model to the fMRI motor task data set from the Human Connectome Project (HCP) (Van Essen et al., 2013). All models are evaluated after standard preprocessing – to simplify comparison to baseline models (Poldrack et al., 2011). The multivariate time series for each subject was parcellated into  $D = 333$  regions using the Gordon Atlas (Gordon et al., 2016). Each subject time series has length  $T = 284$ . Unfortunately, ground

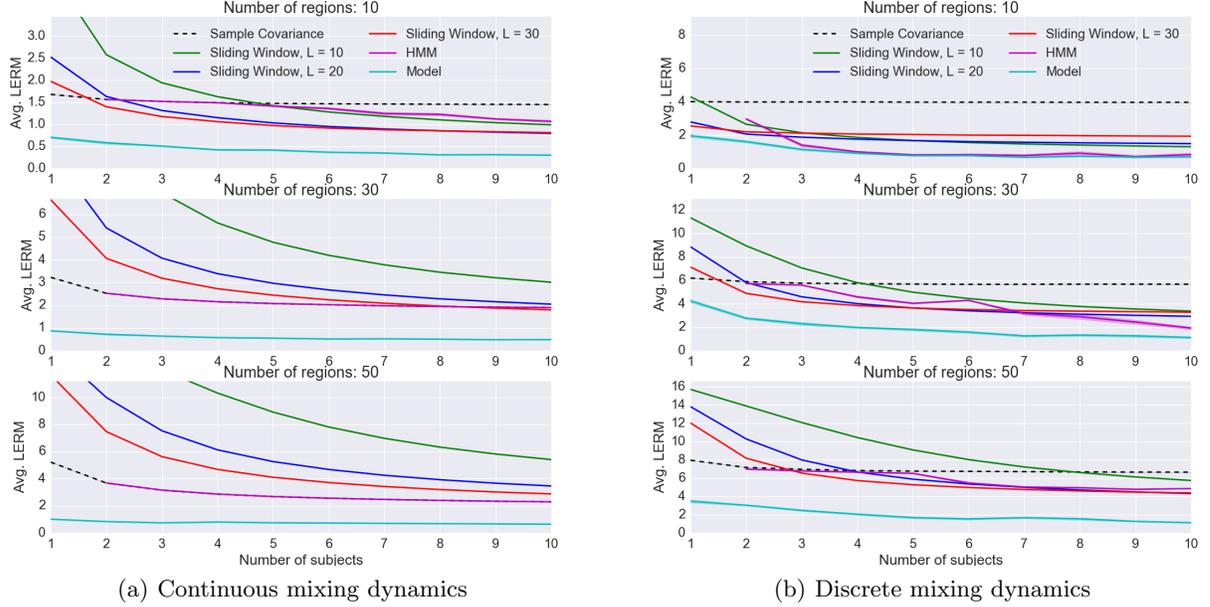


Figure 2: Estimator performance as a function of the number of training subjects for two different simulated data sets. The left-most panel shows the results for the data sets with continuous mixing, while the right-most panel shows the results for the data sets with discrete switching. Furthermore, the three rows show the results using  $D = 10, 30, 50$  number of regions, respectively. The proposed model outperforms baselines even though it is mis-specified with discrete switching data.

truth covariance estimates are unknown for real data. Instead we devise a simple classification test based on task data. In particular, we validate our approach by showing that the task labels can be predicted using the sequences of estimated covariance matrices. We wish to emphasize that the point of evaluating task data is that it provides us with the closest thing we have to ground truth for real data. We present an evaluation of the resting state version for our model (i.e. eq. (1)) on simulated data in the Appendix. Additional experiments evaluating resting state data are necessarily more subjective, and are left for an extended version of this manuscript tailored to a neuroscience audience.

We analyzed data from subjects participating in a motor task experiment. Here, the subjects were asked to perform 5 different motor tasks: left/right hand tapping, left/right foot tapping and tongue wagging at different time points. We divided the data set into a training set and a test set with  $N_{\text{training}} = N_{\text{test}} = 50$  subjects. We fitted the model to the training set and for each time point, we used the posterior expectation,  $\hat{\Sigma}_t = \mathbb{E}_Q[\Sigma_t | \mathcal{D}]$ , as an estimate of the instantaneous covariance matrix at time  $t$ . Next, we computed the time-averaged covariance matrix for each task

$$\hat{\Sigma}_{\text{task } i} = \frac{1}{|\mathcal{T}_i|} \sum_{t \in \mathcal{T}_i} \hat{\Sigma}_t, \quad (6)$$

where  $\mathcal{T}_i$  is the set of time points for the  $i$ 'th task

and  $|\mathcal{T}_i|$  is the number of volumes within task  $i$ . This way, we obtain a covariance matrix estimate,  $\hat{\Sigma}_{\text{left hand tapping}}, \dots, \hat{\Sigma}_{\text{tongue wagging}}$ , for each task. Using a flat prior for the task label,  $p(\text{task } i)$ , we classified the label of each task block of each held-out test subject using Bayes' rule

$$p(\text{task } i | \mathbf{X}^*) \propto \prod_{t=1}^{T'} \mathcal{N}(\mathbf{x}_t^* | \mathbf{0}, \hat{\Sigma}_{\text{task } i}), \quad (7)$$

where  $\mathbf{X}^* \in \mathbb{R}^{D \times T'}$  is the block of data from the test subject to be classified and  $\mathcal{N}(\mathbf{x}_t^* | \mathbf{0}, \hat{\Sigma}_{\text{task } i})$  is the likelihood of task  $i$  for the  $t$ 'th sample  $\mathbf{x}_t^*$ . We compared the proposed method with three reference methods: a regularized covariance matrix estimator, the sliding window approach and random guessing (uniformly). The regularized covariance matrix estimator denoted the *shrunk covariance estimator* given by

$$\mathbf{C}_{\text{task } i}^{\text{shrunk}} = (1 - \gamma) \hat{\mathbf{S}}_{\text{task } i} + \gamma \frac{\text{Tr}[\hat{\mathbf{S}}_{\text{task } i}]}{D} \mathbf{I}, \quad (8)$$

where  $\hat{\mathbf{S}}_{\text{task } i}$  is the sample covariance matrix of the samples within task  $i$  across all training subjects and  $\gamma \in [0, 1]$  is a shrinkage parameter. This estimator is basically a (regularized) sample covariance matrix of all data point belonging to a given task across all training subjects. In addition to providing shrinkage,

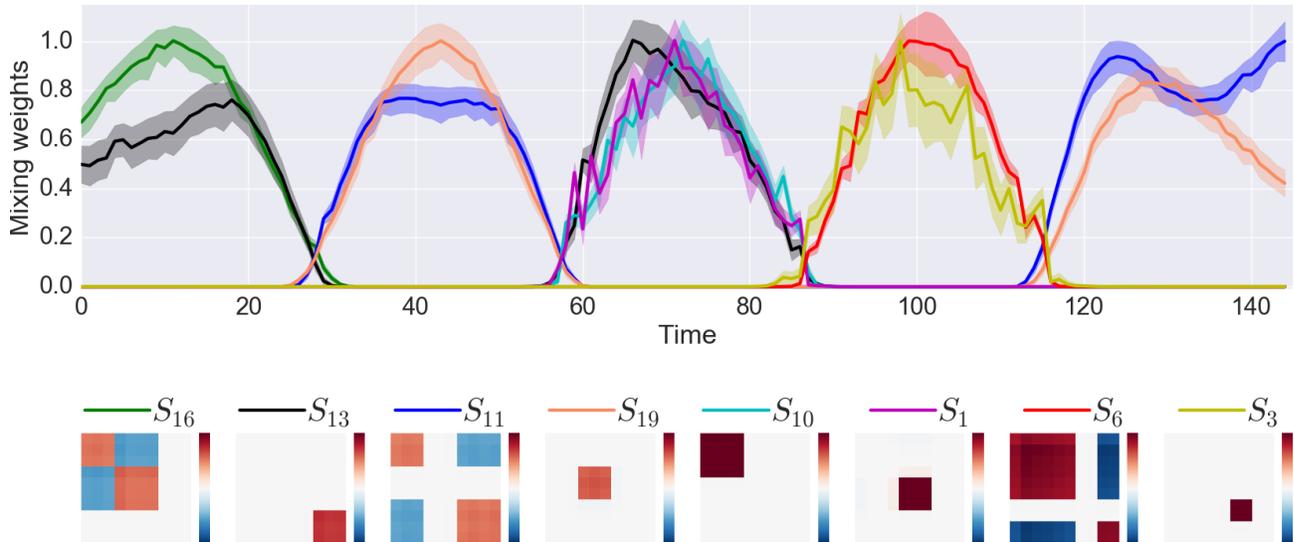


Figure 3: The topmost panel shows estimated mixing weights  $\alpha_k$  for each of the non-zero components and the bottommost panel shows the corresponding estimated covariance components  $S_k$ .

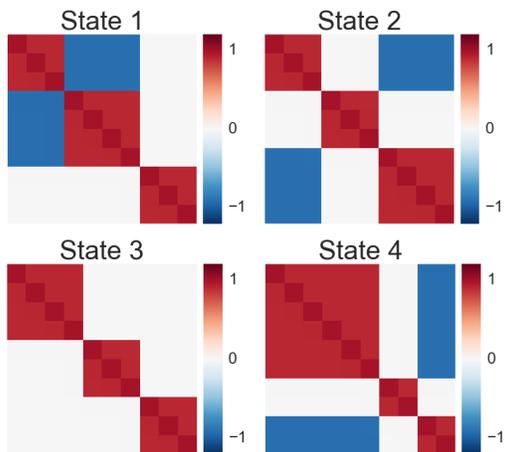


Figure 4: True states for simulated experiment with instantaneous switching dynamics. The proposed model accurately recovers the dictionary elements and mixing weights.

the regularized estimator enable invertibility of the covariance matrix estimate, required in order to evaluate the classification likelihoods in eq. (7), since the number of regions is larger than the number of time points within each task. Furthermore, we also consider the sliding window estimator (also based on the above shrunk estimator rather than the sample estimator) with window sizes of 20, 30, & 40. We use a fixed value for the shrinkage parameter  $\gamma = 0.85$  for both  $C_{\text{task}_i}^{\text{shrunk}}$  and  $C_{\text{task}_i}^{\text{sliding}}$  and for the proposed model, we fixed the upper bound on the number of compo-

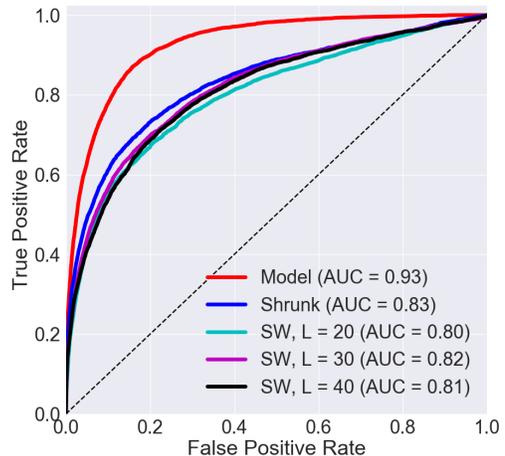


Figure 5: Macro (across all tasks) ROC curves for motor task classification for  $N_{\text{train}} = N_{\text{test}} = 50$  subjects. The results indicate improved performance of the proposed model compared to all baselines.

nents to  $K = 25$  as it was found to be sufficient based on a separate preliminary experiment using a subset of 3 subjects not part of the evaluation. Figure 5 shows the macro ROC curves (Sokolova and Lapalme, 2009) for the multi-class classification problem with  $N = 50$ . The fact that all methods perform better than random suggests that covariance structure of the data are indeed time-varying and contain information about the task labels. Visualizations of the estimated task covariance matrices as well as confusion matrices for the classification problem are included in the supplement

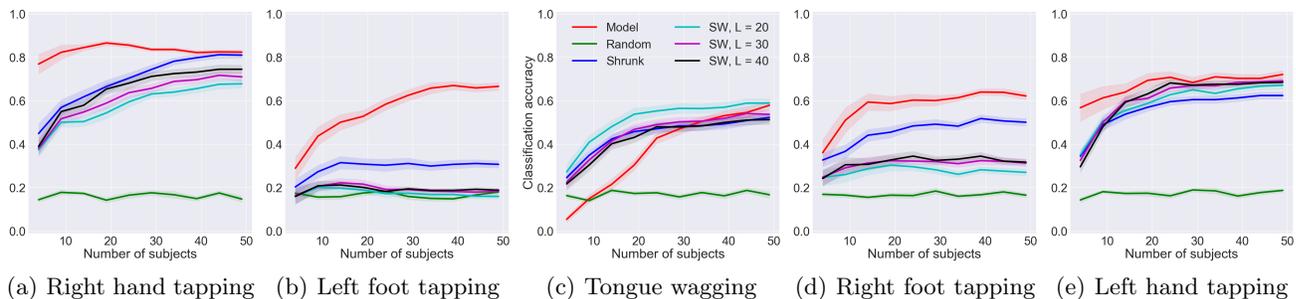


Figure 6: Classification accuracies for each task as a function of the number of training subjects averaged over 20 random splits of the data.

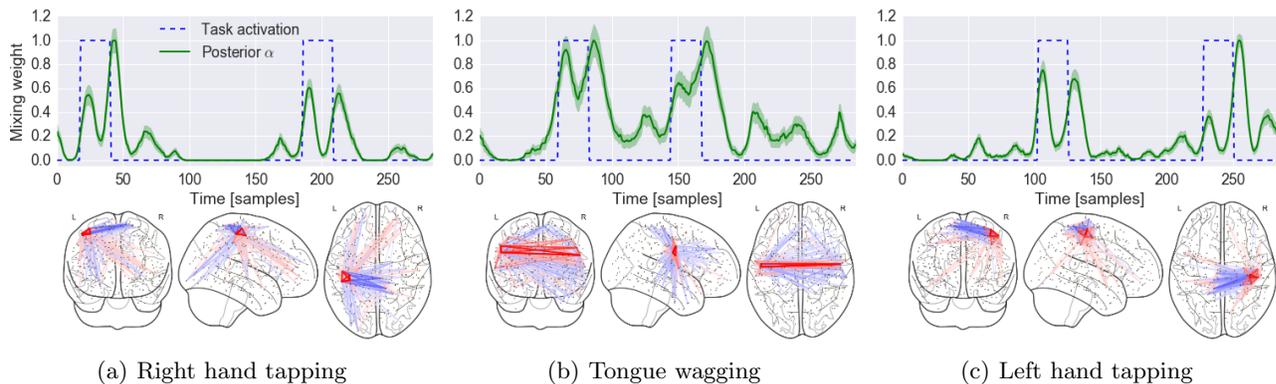


Figure 7: Top panels: Dynamic mixing weights superimposed with task activation pattern. Bottom panels: Visualization of corresponding covariance matrix component. All plots are extracted from a random split with 50 subjects. Estimated connectivity matrices closely match known motor regions associated with each subtask.

tary material. To mimic the fact that the vast majority of fMRI datasets are acquired from only a few of subjects, Figure 6 shows the classification accuracy for the individual tasks as a function of the number of training subjects. To compensate for the small sample size, the results are averaged over  $R = 20$  random splits of the data. It is seen that for 4 out of 5 tasks, the proposed method performs as well or better than the reference methods and that the performance in general increases as a function of number of training subjects as expected. Furthermore, the classification accuracies for the two finger tapping tasks are in general higher than accuracies of the remaining three tasks. However, the performance for the proposed model for the tongue wagging task is worse than the reference method.

Finally, we also compared the set of estimated mixing weights to the task onset sequences of the experimental paradigm. The top-most panels in Figure 7 show the task activation sequences for three different tasks superimposed with the posterior mean of the  $\alpha_k$  that matches the support of the specific task activation. The bottom-most panels visualizes the corresponding covariance matrix components. Interestingly, the results show that the networks expected to be associated

with left and right hand movements (right and left motor cortex, respectively) have indeed non-zero mixing weights near the task-onsets of the relevant tasks. Additional results provided in the supplement show that the recovered connectivity matrices closely match known motor regions associated with each subtask.

## 5 CONCLUSION

We have described a Bayesian model for dynamic functional connectivity estimation, where the dynamics are captured by a set of Gaussian processes, and the covariance components are regularized to be simple and sparse – both properties informed by the neuroscience. The proposed model was evaluated on simulated data and shown to accurately recover model components even when mis-specified, and while outperforming baseline approaches. While ground truth is unknown for real fMRI data, a simple classification test using task fMRI data also indicate the efficacy of the model. Future work will include extending the model to capture additional prior knowledge of spatial dependencies using structured prior spike and slab priors (Smith and Fahrmeir, 2007; Andersen et al., 2014).

## References

- Elena A Allen, Eswar Damaraju, Sergey M Plis, Erik B Erhardt, Tom Eichele, and Vince D Calhoun. Tracking whole-brain connectivity dynamics in the resting state. *Cereb. Cortex*, 24(3):663–676, March 2014.
- Michael R Andersen, Ole Winther, and Lars K Hansen. Bayesian inference for structured spike and slab priors. In Z Ghahramani, M Welling, C Cortes, N D Lawrence, and K Q Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 1745–1753. Curran Associates, Inc., 2014.
- Mohammad R Arbabshirani and Vince D Calhoun. Functional network connectivity during rest and task: comparison of healthy controls and schizophrenic patients. In *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*, pages 4418–4421. IEEE, 2011.
- André M Bastos and Jan-Mathijs Schoffelen. A tutorial review of functional connectivity analysis methods and their interpretational pitfalls. *Front. Syst. Neurosci.*, 9:175, 2015.
- Bharat Biswal, F Zerrin Yetkin, Victor M Haughton, and James S Hyde. Functional connectivity in the motor cortex of resting human brain using echo-planar mri. *Magnetic resonance in medicine*, 34(4):537–541, 1995.
- David M. Blei, Alp Kucukelbir, and Jon D. McAuliffe. Variational Inference: A Review for Statisticians. November 2016. URL <http://arxiv.org/abs/1601.00670>.
- Vince D Calhoun, Robyn Miller, Godfrey Pearlson, and Tulay Adalı. The chronnectome: time-varying connectivity networks as the next frontier in fMRI data discovery. *Neuron*, 84(2):262–274, 22 October 2014.
- Catie Chang, David A Leopold, Marieke Louise Schölvinck, Hendrik Mandelkow, Dante Picchioni, Xiao Liu, Q Ye Frank, Janita N Turchi, and Jeff H Duyn. Tracking brain arousal fluctuations with fmri. *Proceedings of the National Academy of Sciences*, page 201520613, 2016.
- E Damaraju, E A Allen, A Belger, J M Ford, S McEwen, D H Mathalon, B A Mueller, G D Pearlson, S G Potkin, A Preda, J A Turner, J G Vaidya, T G van Erp, and V D Calhoun. Dynamic functional connectivity analysis reveals transient states of dysconnectivity in schizophrenia. *Neuroimage Clin*, 5:298–308, 24 July 2014.
- Xin Di, Suril Gohel, Eun Kim, and Bharat Biswal. Task vs. rest-different network configurations between the coactivation and the resting-state brain networks. *Frontiers in Human Neuroscience*, 7:493, 2013. ISSN 1662-5161. doi: 10.3389/fnhum.2013.00493.
- Elvis Dohmatob, Arthur Mensch, Gael Varoquaux, and Bertrand Thirion. Learning brain regions via large-scale online structured sparse dictionary learning. In *Advances In Neural Information Processing Systems*, pages 4610–4618, 2016.
- Mads Dyrholm, Scott Makeig, and Lars Kai Hansen. Convolutional ICA for spatio-temporal analysis of eeg. *Neural Computation*, 19:934–955, 2007. ISSN 1530888x, 08997667.
- Harini Eavani, Roman Filipovych, Christos Davatzikos, Theodore D Satterthwaite, Raquel E Gur, and Ruben C Gur. Sparse dictionary learning of resting state fMRI networks. *Int Workshop Pattern Recognit Neuroimaging*, pages 73–76, 2 July 2012.
- A.A. Fingelkurts, A.A. Fingelkurts, and S Kahkonen. Functional connectivity in the brain - is it an elusive concept? *Neuroscience and Biobehavioral Reviews*, 28(8):827–836, 2005. ISSN 18737528, 01497634. doi: 10.1016/j.neubiorev.2004.10.009.
- Emily B Fox and David B Dunson. Bayesian nonparametric covariance regression. *Journal of Machine Learning Research*, 16:2501–2542, 2015.
- Michael D Fox and Marcus E Raichle. Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nature Reviews Neuroscience*, 8(9):700–711, 2007.
- K J Friston, C D Frith, P F Liddle, and R S Frackowiak. Functional connectivity: the principal-component analysis of large (PET) data sets. *J. Cereb. Blood Flow Metab.*, 13(1):5–14, January 1993.
- Javier Gonzalez-Castillo, Colin W Hoy, Daniel A Handwerker, Meghan E Robinson, Laura C Buchanan, Ziad S Saad, and Peter A Bandettini. Tracking ongoing cognition in individuals using brief, whole-brain functional connectivity patterns. *Proc. Natl. Acad. Sci. U. S. A.*, 112(28):8762–8767, 14 July 2015.
- Evan M Gordon, Timothy O Laumann, Babatunde Adeyemo, Jeremy F Huckins, William M Kelley, and Steven E Petersen. Generation and evaluation of a cortical area parcellation from Resting-State correlations. 26(1):288–303, January 2016.
- Rikkert Hindriks, Mohit H Adhikari, Yusuke Murayama, Marco Ganzetti, Dante Mantini, Nikos K Logothetis, and Gustavo Deco. Can sliding-window correlations reveal dynamic functional connectivity in resting-state fmri? *Neuroimage*, 127:242–256, 2016.

- Zhiwu Huang, Ruiping Wang, Shiguang Shan, Xianqiu Li, and Xilin Chen. Log-Euclidean metric learning on symmetric positive definite manifold with application to image set classification. In *Proceedings of The 32nd International Conference on Machine Learning*, pages 720–729, 2015.
- R Matthew Hutchison, Thilo Womelsdorf, Elena A Allen, Peter A Bandettini, Vince D Calhoun, Maurizio Corbetta, Stefania Della Penna, Jeff H Duyn, Gary H Glover, Javier Gonzalez-Castillo, Daniel A Handwerker, Shella Keilholz, Vesa Kiviniemi, David A Leopold, Francesco de Pasquale, Olaf Sporns, Martin Walter, and Catie Chang. Dynamic functional connectivity: promise, issues, and interpretations. *Neuroimage*, 80:360–378, 15 October 2013.
- A Hyvarinen and E Oja. Independent component analysis: algorithms and applications. *Neural Networks*, 13(4-5):411–430, 2000. ISSN 18792782, 08936080. doi: 10.1016/S0893-6080(00)00026-5.
- Gregor Kastner. Sparse bayesian time-varying covariance estimation in many dimensions. *arXiv preprint arXiv:1608.08468*, 2016.
- Timothy O Laumann, Abraham Z Snyder, Anish Mitra, Evan M Gordon, Caterina Gratton, Babatunde Adeyemo, Adrian W Gilmore, Steven M Nelson, Jeff J Berg, Deanna J Greene, et al. On the stability of bold fmri correlations. *Cerebral Cortex*, 2016.
- Kaiming Li, Lei Guo, Jingxin Nie, Gang Li, and Tianming Liu. Review of methods for functional brain connectivity detection using fMRI. *Comput. Med. Imaging Graph.*, 33(2):131–139, March 2009.
- Martin A Lindquist, Yuting Xu, Mary Beth Nebel, and Brain S Caffo. Evaluating dynamic bivariate correlations in resting-state fmri: A comparison study and a new approach. *Neuroimage*, 101:531–546, 2014.
- T J Mitchell and J J Beauchamp. Bayesian variable selection in linear regression. *J. Am. Stat. Assoc.*, 83(404):1023–1032, 1988.
- Rasmus Kongsgaard Olsson and Lars Kai Hansen. Linear state-space models for blind source separation. *Journal of Machine Learning Research*, 7: 2585–2602, 2006. ISSN 15337928, 15324435.
- Russell A Poldrack, Jeanette A Mumford, and Thomas E Nichols. *Handbook of functional MRI data analysis*. Cambridge University Press, 2011.
- Michael I. Posner, Steven E. Petersen, Peter T. Fox, and Marcus E. Raichle. Localization of cognitive operations in the human brain. *Science*, 240(4859): 1627–1631, 1988. ISSN 10959203, 00368075. doi: 10.2307/1701013, 10.2307/1701013.
- Maria Giulia Preti, Thomas AW Bolton, and Dimitri Van De Ville. The dynamic functional connectome: State-of-the-art and perspectives. *NeuroImage*, 2016.
- L R Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proc. IEEE*, 77(2):257–286, February 1989.
- Carl Edward Rasmussen and Christopher K I Williams. *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press, 2005.
- Ünal Sakoğlu, Godfrey D Pearlson, Kent A Kiehl, Y Michelle Wang, Andrew M Michael, and Vince D Calhoun. A method for evaluating dynamic functional network connectivity and task-modulation: application to schizophrenia. *Magnetic Resonance Materials in Physics, Biology and Medicine*, 23(5-6):351–366, 2010.
- Douglas H Schultz and Michael W Cole. Higher intelligence is associated with less task-related brain network reconfiguration. *Journal of Neuroscience*, 36(33):8551–8561, 2016.
- James M. Shine, Patrick G. Bissett, Peter T. Bell, Oluwasanmi Koyejo, Joshua H. Balsters, Krzysztof J. Gorgolewski, Craig A. Moodie, and Russell A. Poldrack. The dynamics of functional brain networks: Integrated network states during cognitive task performance. *Neuron*, 2016a.
- James M Shine, Oluwasanmi Koyejo, and Russell A Poldrack. Temporal metastates are associated with differential patterns of time-resolved connectivity, network topology, and attention. *Proceedings of the National Academy of Sciences*, 2016b.
- Michael Smith and Ludwig Fahrmeir. Spatial bayesian variable selection with application to functional magnetic resonance imaging. *J. Am. Stat. Assoc.*, 102(478):417–431, 1 June 2007.
- Stephen M Smith, Karla L Miller, Steen Moeller, Junqian Xu, Edward J Auerbach, Mark W Woolrich, Christian F Beckmann, Mark Jenkinson, Jesper Andersson, Matthew F Glasser, David C Van Essen, David A Feinberg, Essa S Yacoub, and Kamil Ugurbil. Temporally-independent functional modes of spontaneous brain activity. *Proc. Natl. Acad. Sci. U. S. A.*, 109(8):3131–3136, 21 February 2012.
- Marina Sokolova and Guy Lapalme. A systematic analysis of performance measures for classification tasks. *Inf. Process. Manag.*, 45(4):427–437, 2009.
- Martijn P van den Heuvel and Hilleke E Hulshoff Pol. Exploring the brain network: a review on resting-state fMRI functional connectivity. *Eur. Neuropsychopharmacol.*, 20(8):519–534, August 2010.

David C Van Essen, Stephen M Smith, Deanna M Barch, Timothy E J Behrens, Essa Yacoub, Kamil Ugurbil, and WU-Minn HCP Consortium. The WU-Minn human connectome project: an overview. *Neuroimage*, 80:62–79, 15 October 2013.

Raviteja Vemulapalli and David W. Jacobs. Riemannian metric learning for symmetric positive definite matrices. *CoRR*, abs/1501.02393, 2015. URL <http://arxiv.org/abs/1501.02393>.

Ying Yang, Elissa Aminoff, Michael Tarr, and Kass E Robert. A state-space model of cross-region dynamic connectivity in meg/eeg. In D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems 29*, pages 1234–1242. Curran Associates, Inc., 2016.

Andrew Zalesky, Alex Fornito, and Ed Bullmore. On the use of correlation as a measure of network connectivity. *Neuroimage*, 60(4):2096–2106, 2012.