Nonparametric Modeling of Dynamic Functional Connectivity in fMRI Data

Søren F. V. Nielsen[1], Kristoffer H. Madsen[2], Rasmus Røge[1], Mikkel N. Schmidt[1], and Morten Mørup[1]

[1] Section for Cognitive Systems, DTU Compute, Technical University of Denmark
Richard Petersens Plads, Building 324, DK-2800 Kgs. Lyngby
`sfvn@dtu.dk`,
[2] Danish Research Centre for Magnetic Resonance, Section 714
Copenhagen University, Hospital Hvidovre, Kettegaard Allé 30, DK-2650 Hvidovre

Abstract. Dynamic functional connectivity (FC) has in recent years become a topic of interest in the neuroimaging community. Several models and methods exist for both functional magnetic resonance imaging (fMRI) and electroencephalography (EEG), and the results point towards the conclusion that FC exhibits dynamic changes. The existing approaches modeling dynamic connectivity have primarily been based on time-windowing the data and k-means clustering. We propose a nonparametric generative model for dynamic FC in fMRI that does not rely on specifying window lengths and number of dynamic states. Rooted in Bayesian statistical modeling we use the predictive likelihood to investigate if the model can discriminate between a motor task and rest both within and across subjects. We further investigate what drives dynamic states using the model on the entire data collated across subjects and task/rest. We find that the number of states extracted are driven by subject variability and preprocessing differences while the individual states are almost purely defined by either task or rest. This questions how we in general interpret dynamic FC and points to the need for more research on what drives dynamic FC.

Keywords: dynamic functional connectivity, Bayesian nonparametric modeling, hidden Markov modeling, Wishart mixture modeling, predictive likelihood

1 Introduction

The invention of functional magnetic resonance imaging (fMRI) paved the way for non-invasive studies of neuronal activity in the human brain. Especially the correlation of activity between brain regions, *functional connectivity* (FC), has been of interest for over a decade. In recent years the term *dynamic* functional connectivity has emerged in the field trying to explain temporal changes in the FC pattern [1,2,3]. Intuitively this makes sense since we expect that the interaction between segregated brain regions changes. These changes can be due to a number of factors such as the experimental design (task and resting state),

non-experimental physical factors (fatigue, caffeine intake) and even neurological disorders [4]. An important aim of modeling FC as a dynamic concept is to run the models in a fully-unsupervised setting to achieve greater knowledge of how the brain functions and even extract biomarkers for diseases [5]. One very prominent approach to analysing dynamic FC was presented in [6]. Here windowed covariance matrices was extracted from a group-ICA representation of 405 healthy subjects resting state fMRI data. The upper triangular part of each extracted covariance matrix was then stacked into a vector yielding a vector space representation of the covariance structure for each window and subject. These vectors were then clustered using K-means clustering and the number of clusters was chosen using the Elbow-criterion. The results indicated that the resting state paradigm exhibits a temporally dynamical structure and that some of the FC patterns extracted diverge from the classical stationary results (such as the default mode network). In order to validate that the FC patterns (or states) extracted can be used to characterize resting state, we must test that the models find meaningful results in a setting where we have ground truth information available. [7] constructed an experiment where the participants were scanned while being asked to do different tasks in one continuous scan. They found that using windowed covariance matrices and k-means clustering (similar to [6]) the cluster centroids could be used to explain the tasks carried out. But the number of dynamic states k (fixed to the number of tasks) and window length had to be pre-specified whereas the model was not evaluated on independent test data.

In this paper we propose an extension of the infinite hidden Markov model (IHMM) [8] tailored for the modeling of dynamic FC states in fMRI. We furthermore present a predictive likelihood framework for validating that the extracted structure can be used to characterize a motor task from rest. Using the IHMM circumvents having to specify the number of states in the model and window length. Instead the model can be viewed as an adaptive windowing method, where states can persist on different state specific time scales and the number of states learned as part of the inference. We will use this framework to investigate what drives dynamic FC.

2 Methods

The IHMM-Wishart model: A commonly used representation of FC is to represent the connectivity between areas of the brain by the covariance matrix. In a dynamic setting we model each state as having separate covariance matrices. We thus model fMRI in terms of a latent sequence of states, z_t for $t = 1..T$ that for each time point generates a brain image according to the state specific covariance matrix $\mathbf{x}_t \sim \mathcal{N}(\mathbf{0}, \sigma_t^2 \boldsymbol{\Sigma}^{(z_t)})$, where σ_t^2 defines the magnitude of the state specific covariance structure $\boldsymbol{\Sigma}^{(k)}$ at time t allowing states to be invariant to data magnitude but defined in terms of the connectivity profile. For the modeling of transitions between states we use the infinite hidden Markov model (IHMM), first proposed in [8] and further developed in [9,10]. Completing the IHMM with a Gaussian distribution on the observed data and an inverse Wishart prior on

the covariance structure, similar to the Wishart mixture modeling considered in [11,12] and the infinite Gaussian mixture model of [13], we arrive at the following generative model,

$$\boldsymbol{\beta} \sim \text{GEM}(\gamma), \tag{1}$$

$$\boldsymbol{\pi}^{(k)} | \boldsymbol{\beta} \sim \text{DP}(\alpha, \boldsymbol{\beta}), \tag{2}$$

$$z_t | z_{t-1} \sim \text{Multinomial}(\boldsymbol{\pi}^{(z_{t-1})}), \tag{3}$$

$$\boldsymbol{\Sigma}^{(k)} \sim \mathcal{W}^{-1}(\eta \boldsymbol{\Sigma}_0, v_0), \tag{4}$$

$$\mathbf{x}_t \sim \mathcal{N}(\mathbf{0}, \sigma_t^2 \boldsymbol{\Sigma}^{(z_t)}). \tag{5}$$

GEM is the stick-breaking construction (cf. [14,15]), DP is the Dirichlet process, γ and α are positive hyper-parameters controlling the state sequence, $\boldsymbol{\pi}^{(k)}$ denotes the k'th row of the transition matrix $\boldsymbol{\pi}$, η a is (positive) scale parameter, $\boldsymbol{\Sigma}_0$ is a $p \times p$ matrix (covariance prior), and $\mathcal{W}^{-1}(\boldsymbol{\Sigma}_0, v_0)$ denotes the inverse Wishart distribution, and σ_t^2 is the time specific covariance scaling. Viewing this in light of fMRI and FC (cf. also [16,12]), each state can further be represented by a covariance matrix defining the FC where each state is invariant to magnitude of the FC due to the time-specific scaling parameter σ_t^2. We learn this parameter in the inference procedure (cf. next section) using a vague improper $1/\mathcal{X}$-prior. $\boldsymbol{\Sigma}_0$ in the inverse Wishart prior plays the role of the default connectivity. In the experiments we estimate $\boldsymbol{\Sigma}_0$ from a separate resting state fMRI scan. The parameter η is the scaling or level of this default connectivity, which is learned during the inference. The degrees of freedom, v_0, set to the number of dimensions p to make the inference well posed.

Inference: For inference we use Markov chain Monte Carlo (MCMC). Due to conjugacy we can analytically integrate the state specific covariances $\boldsymbol{\Sigma}^{(k)}$ and infer the state sequence $\boldsymbol{z}$ using Gibbs sampling as described in [10] and split-merge sampling [17,18] which has been demonstrated to improved mixing and convergence properties for non-parametric Bayesian mixture models. We learn η and σ_t^2 by Metropolis-Hastings with proposal distribution, $\eta^* = \exp(\ln \eta + z), z \sim \mathcal{N}(0, \sigma = 0.1)$ (and similarly for σ_t^2) imposing an improper and uninformative $1/\mathcal{X}$-prior on both variables. The implementation of the model was done in MATLAB building on top of Juergen Van Gael's iHMM-toolbox [19]. We used the code available online[3] for sampling state-sequence relevant hyperparameters, where vague Gamma priors are placed on α and γ and inferred via. an auxiliary variable Gibbs sampler [20]. The predictive likelihood can be calculated using a modified Viterbi algorithm [21], and analytically integrating out σ_t^2 and using parameter posterior samples from the inference.

3 Results

We used fMRI data from a population of 29 subjects, that carried out a simple finger-tapping experiment (denoted motor) [22] as well as a resting state scan

---

[3] http://mloss.org/software/view/205/

[23]. Preprocessing included normalization to MNI space and wavelet despiking [24]. We did a group ICA [25] using the GIFT toolbox[4] into 20 components using the maximum likelihood ERBM algorithm [26] with default settings. We discarded six components by visual inspection of the spatial maps representing common noise confounds. We split each subjects motor and resting-state data into a training (116 images) and a test set (120 images). Each training and test set was individually corrected for motion- and respiratory effects [27], reference signal from cerebrospinal fluid (lateral ventricles) and white matter along with high-pass filtering (1/128 s cut-off) in a single regression step. The resting state scan was twice as long as the motor experiment so the second half of the resting state scan was removed and used for estimating the prior covariance $\boldsymbol{\Sigma}_0$. For each of the 29 subjects training sets available (motor and rest) we ran ten IHMM-Wishart models and ten constant Wishart models, i.e. an IHMM-Wishart forced to be in one state.

First of all, we investigate if a model trained on the motor task (denoted a motor-model) predicts better on the motor-test set relative to a model trained on the resting state (denoted a rest-model). Second of all, we can test if predictive performance increases using a dynamic model, compared to a one-state constant model. For the motor data, we calculated the predictive likelihood for all ten runs, and made a paired t-test to evaluate if the predictive performance was better by the motor-model or the rest-model. The same was done for the resting state data. We saw on test data that within all 29 subjects, motor and rest was distinguishable with an IHMM-Wishart model. We furthermore tested if there was a significant difference between using a dynamic model vs. a static model. Except for one subject, this difference was not significant, which can be explained by the IHMM-Wishart finding almost exclusively one state on both motor and rest.

The above analysis can be extended to group level, i.e. to investigate if a subjects task and resting state data can be distinguished by models trained on other subjects. We looked at 4 hypothesis tests, 1) that the motor- and rest-model predict equally well on motor data, 2) that the rest- and motor-model predict equally well on resting state data, 3) that the motor-model is equal to using a one state constant model on motor data and 4) that the rest-model is equal to using a one state constant model on the resting state data. For motor hypothesis 1 was rejected on average 26.5 ± 2 out of 28 times for a subject, and for resting-state hypothesis 2 was rejected on average 26.1 ± 2 out of 28 times. Similarly as the within-subject analysis, very few dynamic models perform better on testing data than the one-state counterpart models. Hypothesis 3) and 4) could only be rejected 0.31 ± 0.5 and 0.24 ± 0.6 times respectively.

To investigate dynamic FC at group level, we collated all subjects motor and resting state data together and ran the IHMM-Wishart fully unsupervised on the whole data set. In the modeling, states are shared across subjects and tasks but we handle discontinuities in the data by restarting the chain at each block that has been preprocessed individually (4 per subject). We ran the IHMM-Wishart

---

[4] http://mialab.mrn.org/software/gift/

five times, and the sample with highest posterior probability was taken out for further analysis. The results can be seen in figure 1. In figure 1a we report the four most populated states' covariance matrices in both motor and resting-state along with the ICA components (IC). We note IC1 that well corresponds to the motor task is active in all the motor-states, and has been down-weighted in the resting-states. Inspecting the two tasks, the average number of states pr. subject in the motor-task was 3.28 ± 0.21 and for the resting state 3.60 ± 0.27. Furthermore, the states extracted are generally pertaining almost perfectly to either motor or rest (cf. figure 1b). However, from figure 1c we see that the average mutual information (MI) over posterior samples with state sequences describing the subjects, task and preprocessing shows that subjects and preprocessing drives the dynamics more than the tasks themselves.

4   Discussion

The IHMM-Wishart can distinguish between motor and rest on test data from the same subject. One confound is that the two experiments are recorded separately meaning that our ability to distinguish between motor and resting-state test data probably is caused by training and test set being recorded close in time. Since the group analysis shows that models trained on a subject can well distinguish between motor and rest of others, this confound is not present in this setting. This means that the IHMM-Wishart is able to well extract states that characterize motor from rest. In this analysis the model only found support for a single state in most of these single subject analyses carried out. The number of states found by the IHMM-Wishart is influenced though by the number of data points available and dimensionality, and this could influence the 'absence' of dynamics found due to the size of the training set (116 images). Collating data together from all subjects, we ran our model on the entire data set and saw each state were clearly pertaining to mainly either motor or rest. We see though that preprocessing has an impact on the states extracted - in an information theoretic perspective, the dynamics found are closer related to preprocessing than both to subjects and task (cf. figure 1c).

Our predictive framework demonstrates that the IHMM-Wishart can be used to characterize task vs. rest. This is a very controlled setting and should also be possible to achieve for simpler models than the one we are proposing here but evidences the utility of the IHMM-Wishart in characterizing fMRI data. Importantly, the IHMM-Wishart enables us to infer the number of dynamic states from data which for the collated data turned out to be *more* than just two states, i.e. one for rest and one for motor. We here observed multiple task specific dynamic states driven by subject variability and preprocessing, which is aligned with the conclusion in [12]. This tells us that care must be taken when interpreting dynamics at a group level as preprocessing even of blocks of data from the same subject evidences the presence of multiple dynamic functional states.
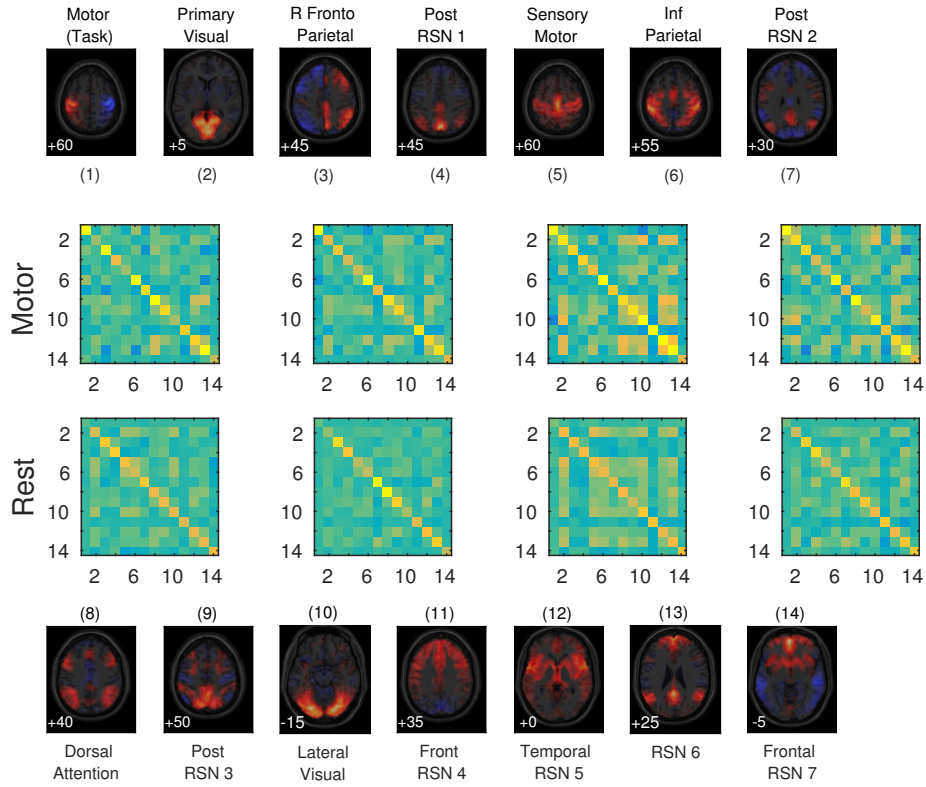
(a) Covariance estimate for each of the most populated states in the motor task (top row) and resting state (bottom row), along with the ICA component maps.
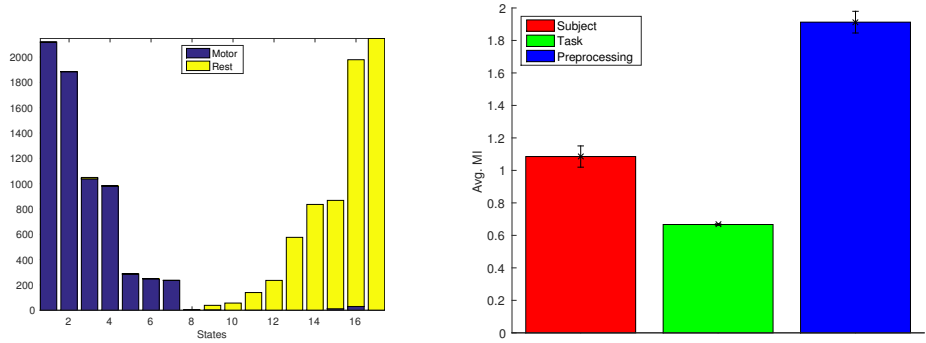
(b) A bar plot displaying how frequently each state is populated. The states have been sorted by the difference in frequency in motor and rest-state.

(c) Average mutual information (MI) over samples with three state sequences, 1) subject (red), 2) task (green) and 3) preprocessing (blue)

Fig. 1: We collated all subjects motor and resting state experiments together and ran 5 chains of the IHMM-Wishart model. We show how the different states are populated, and investigate what drives the dynamics; subjects, tasks or preprocessing, for the sample with highest posterior probability.

Acknowledgements

References

1. 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, et al. Dynamic functional connectivity: promise, issues, and interpretations. *Neuroimage*, 80:360–378, 2013.

2. Andrew Zalesky, Alex Fornito, Luca Cocchi, Leonardo L Gollo, and Michael Breakspear. Time-resolved resting-state brain networks. *Proceedings of the National Academy of Sciences*, 111(28):10341–10346, 2014.

3. 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.

4. Vince D Calhoun, Tom Eichele, and Godfrey Pearlson. Functional brain networks in schizophrenia: a review. *Frontiers in human neuroscience*, 3, 2009.

5. Archana Venkataraman, Thomas J Whitford, Carl-Fredrik Westin, Polina Golland, and Marek Kubicki. Whole brain resting state functional connectivity abnormalities in schizophrenia. *Schizophrenia research*, 139(1):7–12, 2012.

6. 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. *Cerebral cortex*, page bhs352, 2012.

7. 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.

8. Matthew J Beal, Zoubin Ghahramani, and Carl E Rasmussen. The infinite hidden markov model. In *Advances in neural information processing systems*, pages 577–584, 2001.

9. Alan S Willsky, Erik B Sudderth, Michael I Jordan, and Emily B Fox. Nonparametric bayesian learning of switching linear dynamical systems. In *Advances in Neural Information Processing Systems*, pages 457–464, 2009.

10. Jurgen Van Gael. *Bayesian Nonparametric Hidden Markov Models*. PhD thesis, University of Cambridge, 2012.

11. Sullivan Hidot and Christophe Saint-Jean. An expectation–maximization algorithm for the wishart mixture model: Application to movement clustering. *Pattern Recognition Letters*, 31(14):2318–2324, 2010.

12. Josefine Korzen, Kristoffer H Madsen, and Morten Mørup. Quantifying temporal states in rs-fmri data using bayesian nonparametrics. In *HBM 2014, Poster Number: 1726*. Organization for Human Brain Mapping, 2014.

13. Carl Edward Rasmussen. The infinite gaussian mixture model. In *NIPS*, volume 12, pages 554–560, 1999.

14. J. Sethuraman. A constructive definition of Dirichlet priors. *Statistica Sinica*, 4:639–650, 1994.

15. Jim Pitman. Poisson–dirichlet and gem invariant distributions for split-and-merge transformations of an interval partition. *Combinatorics, Probability & Computing*, 11(05):501–514, 2002.

16. Srikanth Ryali, Kaustubh Supekar, Tianwen Chen, Weidong Cai, and Vinod Menon. A variational bayes hidden markov model for discovering dynamical functional brain networks. In *HBM 2015, Abstract 3749*. Organization for Human Brain Mapping, 2015.

17. Sonia Jain and Radford M Neal. A split-merge markov chain monte carlo procedure for the dirichlet process mixture model. *Journal of Computational and Graphical Statistics*, 13(1), 2004.

18. Michael C Hughes, Erik B Sudderth, and Emily B Fox. Effective split-merge monte carlo methods for nonparametric models of sequential data. In *Advances in Neural Information Processing Systems*, pages 1295–1303, 2012.

19. J. Van Gael. The infinite hidden markov model, 2010. `http://mloss.org/software/view/205/`.

20. Yee Whye Teh, Michael I Jordan, Matthew J Beal, and David M Blei. Hierarchical dirichlet processes. *Journal of the american statistical association*, 101(476), 2006.

21. Andrew J Viterbi. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *Information Theory, IEEE Transactions on*, 13(2):260–269, 1967.

22. Peter M Rasmussen, Lars K Hansen, Kristoffer H Madsen, Nathan W Churchill, and Stephen C Strother. Model sparsity and brain pattern interpretation of classification models in neuroimaging. *Pattern Recognition*, 45(6):2085–2100, 2012.

23. Kasper Winther Andersen, Kristoffer H Madsen, Hartwig Roman Siebner, Mikkel N Schmidt, Morten Mørup, and Lars Kai Hansen. Non-parametric bayesian graph models reveal community structure in resting state fmri. *NeuroImage*, 100:301–315, 2014.

24. Ameera X Patel, Prantik Kundu, Mikail Rubinov, P Simon Jones, Petra E Vértes, Karen D Ersche, John Suckling, and Edward T Bullmore. A wavelet method for modeling and despiking motion artifacts from resting-state fMRI time series. *Neuroimage*, 95:287–304, 15 July 2014.

25. VD Calhoun, T Adali, GD Pearlson, and JJ Pekar. A method for making group inferences from functional mri data using independent component analysis. *Human brain mapping*, 14(3):140–151, 2001.

26. Xi-Lin Li and Tülay Adali. Blind spatiotemporal separation of second and/or higher-order correlated sources by entropy rate minimization. In *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, pages 1934–1937. IEEE, 2010.

27. K J Friston, S Williams, R Howard, R S Frackowiak, and R Turner. Movement-related effects in fMRI time-series. *Magn. Reson. Med.*, 35(3):346–355, March 1996.