2015 IEEE INTERNATIONAL WORKSHOP ON MACHINE LEARNING FOR SIGNAL PROCESSING, SEPT. 17–20, 2015, BOSTON, USA

UNSUPERVISED SEGMENTATION OF TASK ACTIVATED REGIONS IN FMRI

Rasmus E. Røge, Kristoffer H. Madsen, Mikkel N. Schmidt, Morten Mørup

Section for Cognitive Systems
Department of Applied Mathematics and Computer Science
Technical University of Denmark

ABSTRACT

Functional Magnetic Resonance Imaging has become a central measuring modality to quantify functional activiation of the brain in both task and rest. Most analysis used to quantify functional activation requires supervised approaches as employed in statistical parametric mapping (SPM) to extract maps of task induced functional activations. This requires strong knowledge and assumptions on the BOLD response as a function of activitation while smoothing in general enhances the statistical power but at the cost of spatial resolution. We propose a fully unsupervised approach for the extraction of task activated functional units in multi-subject fMRI data that exploits that regions of task activation are consistent across subjects and can be more reliably inferred than regions that are not activated. We develop a non-parametric Gaussian mixture model that apriori assumes activations are smooth using a Gaussian Process prior while assuming the segmented functional maps are the same across subjects but having individual time-courses and noise variances. To improve inference we propose an enhanced split-merge procedure. We find that our approach well extracts the induced activity of a finger tapping fMRI paradigm with maps that well corresponds to a supervised group SPM analysis. We further find interesting regions that are not activated time locked to the paradigm. Demonstrating that we in a fully unsupervised manner are able to extract the task-induced activations forms a promising framework for the analysis of task fMRI and resting-state data in general where strong knowledge of how the task induces a BOLD response is missing.

***Index Terms*—** Functional connectivity, Gaussian Mixture Model, fMRI analysis

1. INTRODUCTION

Functional Magnetic Resonance Imaging (fMRI) allows for the identification of task related brain activations by measuring the blood oxygenation level dependent (BOLD) response in each voxel of the brain. The reliable identification of these regions of activations poses a major challenge due to a massive multiple comparisons problem and due to the low level of signal to noise found in fMRI data. Traditionally, fMRI data is spatially smoothened and then analysed using statistical parametric mapping (SPM) [1] in order to identify brain regions that are significantly correlated with the expected activation time course. These univariate voxel specific tests have to be corrected for multiple comparisons as fMRI data is high-dimensional i.e. in the order of 10^5 voxels. This is commonly handled by methods for correcting for the family-wise error based on Gaussian Random Fields [2] or correcting for false discoveries [3].

This work was supported by the Lundbeck Foundation, grant no. R105-9813

An alternative procedure is to cluster the fMRI time-series [4, 5] and carry out the statistical analysis at the level of clusters. Common adopted approaches have here been based on hierarchical and k-means clustering [4] and the cluster based algorithm (CBA) proposed in [5]. Thereby the number of statistical tests reduces to the number of clusters extracted while spatial smoothing no longer is necessary as the signal to noise ratio (SNR) is improved when considering the time series of each cluster centroid [5]. As opposed to segmenting the brain into clusters, independent component analysis (ICA) [6] is a widely applied approach to identify maps of task [6] and resting state [7] activations typically assuming spatial independence of the extracted components. In these approaches activations have been established using supervised evaluation of the extracted time courses [8].

In this paper we focus on a fully unsupervised approach for functional segmentation of task related activity using clustering based on a non-parametric Mixture Model tailored to the analysis of multi-subject fMRI data. Being non-parametric our model is able to automatically learn from data the number of clusters. It further assumes subjects are normalized into a common space such that the extracted functional units are consistent across subjects with subject specific cluster time-series apriori assumed smooth by imposing a Gaussian Process prior. To account for inhomogenous noise and misaligments a separate noise parameter is estimated individually for each subject and voxel. Inference in the model is accomplished by Markov-chain Monte Carlo sampling where we propose a new efficient procedure for split-merge sampling [9] that significantly reduces the computations of the in general most occuring merge operations.

For the unsupervised extraction of task-related clusters we evaluate how correlated the cluster time-series are across subjects as well as the stability of the clusters using evidence accumulation [10, 11] hypothesizing that these consistent actitvations correspond to task induced activity. We show that our unsupervised multi-subject analysis extracts the regions expected to be activated in a finger tapping paradigm and that the maps are similar to those extracted using a standard supervised group SPM analysis. Our method further circumvents smoothing as a necessary preprocessing step.

2. METHODS

2.1. The Infinite Gaussian Mixture model with a Gaussian Process prior (IGMMGP)

Let $\boldsymbol{X}_s$ be the $N \times T$ matrix of the N voxels with a time course with T measurements of subject s. We use $\boldsymbol{z}$ as a vector for the group assignment, such that $\boldsymbol{z}(i) = k$ if the i'th voxel is assigned to the cluster k. We thus assume that the voxels are aligned across subjects and that the clustering is shared. Our model is illustrated as

a directed graphical model in Fig. 1 and is described by the following generative process where we use the Chinese Restaurant Process (CRP) [12, 13] as a prior for partitioning the voxels into clusters.:

$$\boldsymbol{z} \sim CRP(\gamma) \qquad \text{groups}, \tag{1}$$

$$\boldsymbol{\mu}_{k,s} \sim GP(0, \boldsymbol{\Sigma}) \qquad \text{group time series}, \tag{2}$$

$$\boldsymbol{x}_{i,s} \sim \mathcal{N}(\boldsymbol{\mu}_{z(i),s}, \sigma_{i,s}^2 \boldsymbol{I}) \qquad \text{voxel time series}, \tag{3}$$

where $\boldsymbol{\Sigma}$ is the covariance structure encoding the imposed temporal dynamics, $\boldsymbol{\mu}_{k,s}$ is the time series of cluster k of subject s, $\sigma_{i,s}^2$ is the variance of voxel i of subject s, and $\boldsymbol{x}_{i,s}$ is the time series of voxel i of subject s. The model we propose is an extension of the Infinite Gaussian Mixture Model [14] in which temporal dynamics are imposed on the mixtures through the Gaussian Process (GP) prior with covariance function $\boldsymbol{\Sigma}$. A Gaussian Process prior has previously been considered in the Infinite Gaussian Mixture Models also imposing Markov Random Field constraints [15]. A benefit of our model is that it includes voxel and subject specific noise that can potentiallty account for misalignment across subjects as well as spatially varying noise levels. In the following we call our model the Infinite Gaussian Mixture model with a Gaussian Process prior (IGMMGP).

Let $\boldsymbol{X}_s$, and $\boldsymbol{\mu}_s$ be the time-series for all voxels and groups and $\boldsymbol{\sigma}_s^2$ be the noise variance for all voxels for subject s. We define by $\{\boldsymbol{X}\}$, $\{\boldsymbol{\mu}\}$, and $\left\{\boldsymbol{\sigma}^{\boldsymbol{2}}\right\}$ the collections of all subjects voxel time series, group time series and voxel variances respectively. With $\boldsymbol{x}_{\boldsymbol{z}(i)=k,s} = \{\boldsymbol{x}_{i,s} \mid \boldsymbol{z}(i) == k\}$ we denote the time series of all voxels assigned to cluster k of subject s.

According to the generative model the joint distribution of data and parameters can be expressed as

$$p(\boldsymbol{z}, \{\boldsymbol{\mu}\}, \{\boldsymbol{X}\} \mid \left\{\boldsymbol{\sigma}^2\right\}, \boldsymbol{\Sigma}, \gamma) \tag{4}$$
$$= \left[\prod_s p(X_s|\boldsymbol{\mu}_s, \boldsymbol{\sigma}_s^2) p(\boldsymbol{\mu}_s|\boldsymbol{\Sigma})\right] p(\boldsymbol{z}|\gamma).$$

Due to conjugacy it is possible to analytically integrate out the group time series from the joint distribution

$$p(\boldsymbol{z}, \{\boldsymbol{X}\} \mid \left\{\boldsymbol{\sigma}^2\right\}, \boldsymbol{\Sigma}, \gamma) = \tag{5}$$
$$= \int \left[\prod_s p(X_s|\boldsymbol{\mu}_s, \boldsymbol{\sigma}_s^2) p(\boldsymbol{\mu}_s|\boldsymbol{\Sigma}) d\boldsymbol{\mu}_s\right] p(\boldsymbol{z}|\gamma))$$
$$= \frac{\Gamma(\gamma)\gamma^K \prod_{k=1}^K \Gamma(n_k)}{\Gamma(N+\gamma)} \prod_{i,s} (2\pi\sigma_{i,s}^2)^{-T/2} \exp\left\{-\frac{1}{2\sigma_{i,s}^2} \boldsymbol{x}_{i,s}^\top \boldsymbol{x}_{i,s}\right\}$$
$$\prod_{k=1}^K (|\boldsymbol{S}_{k,s}|/|\boldsymbol{\Sigma}|)^{-1/2} \exp\left\{\frac{1}{2} \sum_{k=1}^K \bar{\boldsymbol{x}}_{k,s}^\top \boldsymbol{S}_{k,s}^{-1} \bar{\boldsymbol{x}}_{k,s}\right\},$$

where

$$\bar{\boldsymbol{x}}_{k,s} = \sum_{Z(i)=k} \frac{1}{\sigma_i^2} \boldsymbol{x}_{i,s}, \quad \boldsymbol{S}_{k,s} = \left(\boldsymbol{\Sigma}^{-1} + \sum_{\boldsymbol{z}(i)=k} \frac{1}{\sigma_{i,s}^2} \boldsymbol{I}\right),$$

and n_k is the number of voxels assigned to cluster k with N being the total number of voxels.

2.2. Model inference and accellerated merge steps

For inference of the clustering $\boldsymbol{z}$ we use Gibbs sampling with split-merge moves [9]. In each Gibbs move a voxel can be placed in any

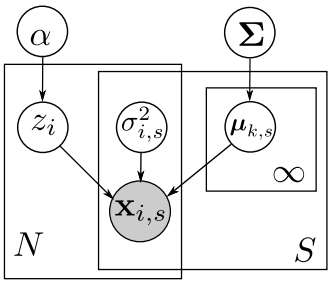

Fig. 1. Directed graphical model representation of the generative process.

of the currently occupied K clusters or create a new. This means that the expression $\boldsymbol{x}_{k,s}^\top \boldsymbol{S}_{k,s}^{-1} \bar{\boldsymbol{x}}_{k,s}$ has to be evaluated $K+1$ times. Let $\boldsymbol{\Sigma} = \boldsymbol{V}^\top \boldsymbol{D} \boldsymbol{V}$ be the eigendecomposition of $\boldsymbol{\Sigma}$. Then

$$\boldsymbol{x}_{k,s}^\top \boldsymbol{S}_{k,s}^{-1} \bar{\boldsymbol{x}}_{k,s} = \boldsymbol{x}_{k,s}^\top \left((\boldsymbol{V}^\top \boldsymbol{D} \boldsymbol{V})^{-1} + \sum_{\boldsymbol{z}(i)=k} \frac{1}{\sigma_{i,s}^2} \boldsymbol{I}\right)^{-1} \bar{\boldsymbol{x}}_{k,s}$$
$$= (\boldsymbol{V}\boldsymbol{x}_{k,s})^\top (\boldsymbol{D}^{-1} + \sum_{\boldsymbol{z}(i)=k} \frac{1}{\sigma_{i,s}^2} \boldsymbol{I})^{-1} \boldsymbol{V}\boldsymbol{x}_{k,s}.$$

Keeping $\boldsymbol{V}\boldsymbol{x}_{k,s}$ and $(\boldsymbol{D}^{-1} + \sum_{\boldsymbol{z}(i)=k} \frac{1}{\sigma_{i,s}^2} \boldsymbol{I})$ in memory reduces the computational complexity of evaluating $\boldsymbol{x}_{k,s}^\top \boldsymbol{S}_{k,s}^{-1} \bar{\boldsymbol{x}}_{k,s}$ to $O(T)$ when the projections $V\{\boldsymbol{X}\}$ and eigenvalues $\boldsymbol{D}$ have been computed. This reduces the total timecomplexity of a Gibbs sweep to $O(SNKT)$.

In the split-merge procedure two nodes are randomly sampled and if they are in the same group the group is proposed split and if they are in two different groups these two groups are proposed merged. The split/merge move is accepted according to the Metropolis-Hastings ratio

$$\alpha(\boldsymbol{z}^* \mid \boldsymbol{z}) = \min\left[1, \frac{p(\boldsymbol{z}^*, \{\boldsymbol{X}\} \mid \gamma, \left\{\boldsymbol{\sigma}^2\right\}, \boldsymbol{\Sigma}) q(\boldsymbol{z}|\boldsymbol{z}^*)}{p(\boldsymbol{z}, \{\boldsymbol{X}\} \mid \gamma, \{\boldsymbol{\sigma}^2\}, \boldsymbol{\Sigma}) q(\boldsymbol{z}^*|\boldsymbol{z})}\right], \tag{6}$$

where the transition probability $q(\boldsymbol{z}^{split}|\boldsymbol{z}^{merge})$ for splitting a cluster is calculated using Gibbs sampling restricted to the observations influenced by the move. The transition probability is calculated by first sampling a so-called launch state and keeping track of the transition probabilities from this state to the final split configuration. As the merge move is deterministic we have for the transition probability $q(\boldsymbol{z}^{merge}|\boldsymbol{z}^{split}) = 1$, see also [9].

Provided no group contain more than 50 % of the observations there will be more merge than split proposals. As merge moves are deterministic we further have for these moves that $\frac{q(\boldsymbol{z}|\boldsymbol{z}^*)}{q(\boldsymbol{z}^*|\boldsymbol{z})} \leq 1$. We can thus significantly accelerated the evaulation of these proposals if we are able to reject the move by the ratio of the joint probabilities alone, i.e. $\frac{p(\boldsymbol{z}^*, \{\boldsymbol{X}\}|\gamma, \left\{\boldsymbol{\sigma}^2\right\}, \boldsymbol{\Sigma})}{p(\boldsymbol{z}, \{\boldsymbol{X}\}|\gamma, \left\{\boldsymbol{\sigma}^2\right\}, \boldsymbol{\Sigma})}$ thereby circumventing the more computationally demanding restricted Gibbs sweeps. This leas us to propose the following accelerated merge procedure: Before computing the launch state and final configuration using restricted Gibbs sampling compute the preliminary acceptance probability for a merge step

$$\alpha_1(\boldsymbol{z}^* \mid \boldsymbol{z}) = \min\left[1, \frac{p(\boldsymbol{z}^*, \{\boldsymbol{X}\} \mid \gamma, \left\{\boldsymbol{\sigma}^2\right\}, \boldsymbol{\Sigma})}{p(\boldsymbol{z}, \{\boldsymbol{X}\} \mid \gamma, \{\boldsymbol{\sigma}^2\}, \boldsymbol{\Sigma})}\right]. \tag{7}$$

In case we cannot accept the proposal based on α_1 we will not be able to accept it based on α as $\alpha(\boldsymbol{z}^*|\boldsymbol{z}) \leq \alpha_1(\boldsymbol{z}^*|\boldsymbol{z})$.

In order to infer the hyperparameters γ and $\{\boldsymbol{\sigma}^2\}$ we impose the non-informative and improper prior $p(\theta) = \theta^{-1}$. We use Metropolis-Hastings sampling by transforming the variable to the log-domain and use the symmetric normal distribution as proposal density. Using the eigendecomposition of $\boldsymbol{\Sigma}$ and $\boldsymbol{V}\boldsymbol{x}_{k,s}$ the cost of evaluating the joint density ratio of the Metropolis-Hastings ratio is $O(T)$. Therefore a sweep of evaluating proposals of $\{\boldsymbol{\sigma}^2\}$ is $O(SNT)$, i.e. this inference step is less computationally demanding than the Gibbs sweep.

2.3. Unsupervised extraction of consistent clusters

The posterior of the cluster time series given the data and clustering can be calculated using Bayes theorem:

$$\begin{aligned} p(\boldsymbol{\mu}_{k,s}|\boldsymbol{x}_{\boldsymbol{z}(i)=k,s}, \boldsymbol{\sigma}^2_{\boldsymbol{z}(i)=k,s}, \boldsymbol{\Sigma}) \\ = \frac{p(\boldsymbol{x}_{\boldsymbol{z}(i)=k} \mid \boldsymbol{\mu}_{k,s}, \boldsymbol{\sigma}^2_{\boldsymbol{z}(i)=k,s}) p(\boldsymbol{\mu}_{k,s} \mid \boldsymbol{\Sigma})}{\int p(\boldsymbol{x}_{\boldsymbol{z}(i)=k} \mid \boldsymbol{\mu}_{k,s}, \boldsymbol{\sigma}^2_{\boldsymbol{z}(i)=k,s}) p(\boldsymbol{\mu}_{k,s} \mid \boldsymbol{\Sigma}) d\boldsymbol{\mu}_{k,s}} \\ = \mathcal{N}(\boldsymbol{S}_k^{-1}\bar{\boldsymbol{x}}_k, \boldsymbol{S}_k). \end{aligned} \tag{8}$$

To un-supervised select clusters of relevance for the task we evaluate the consistency of the cluster time-series as well as the consistency of the clustering across separate chains based on the sample with highest value of the joint-distribution $p(\boldsymbol{z}, \{\boldsymbol{X}\} \mid \{\boldsymbol{\sigma}^2\}, \boldsymbol{\Sigma}, \gamma)$ presently denoted the MAP solution.

We evaluate the consistency of the cluster time-series across subjects by computing the posterior mean, $\boldsymbol{S}_{k,s}^{-1}\bar{\boldsymbol{x}}_{k,s}$, for all clusters and all subjects and rank the clusters according to the mean correlation (over all pairs of subjects), i.e.

$$R(k) = \frac{1}{S(S-1)/2} \sum_{s>s'} correlation(\boldsymbol{\mu}_{k,s}, \boldsymbol{\mu}_{k,s'}). \tag{9}$$

To evaluate the consistency of the clusterings we use evidence accumulation [10, 11] in order to quantify how consistent across L separate sampling chains (excluding the chain with the MAP solution) voxels are grouped together according to the following cluster specific consistency score

$$C(k) = \frac{1}{n_k(n_k-1)/2} \sum_{i>j : z_i^{\text{MAP}} = z_j^{\text{MAP}} = k} \frac{1}{L} \sum_l \mathbb{I}(z_i^{(l)} = z_j^{(l)}), \tag{10}$$

where $\mathbb{I}(a)$ is the indicator function that evaluates to 1 if a is true and 0 otherwise and n_k is the number of voxels in cluster k.

3. RESULTS AND DISCUSSION

To impose smoothness we use as kernel for the covariance of the Gaussian Process, $\boldsymbol{\Sigma}$, that is generated by the following expression:

$$k_{\text{SE}}(\boldsymbol{x}_i(t), \boldsymbol{x}_i(t')) = \exp\left(-\frac{(t-t')^2}{2l^2}\right), \tag{11}$$

with the characteristic length-scale as the optimal length-scale for modeling the hemodynamic response function as provided by the SPM12 software (SPM12, Wellcome Trust Centre for Neuroimaging, http://www.fil.ion.ucl.ac.uk/spm/software/spm12/). The length-scale was inferred by optimizing the fit of a Gaussian Process with a squared exponential kernel and was $l = 4.6s$ or 1.85 frames (with TR = 2.49). In our analysis we initialized the clustering configuration to have all voxels in the same cluster. The noise parameter $\boldsymbol{\sigma}^2$ was initialized to the variance of the data, i.e. $\sigma_i^2 = \text{var}(\boldsymbol{x}_i)$ and the CRP parameter γ was initialized to 5.

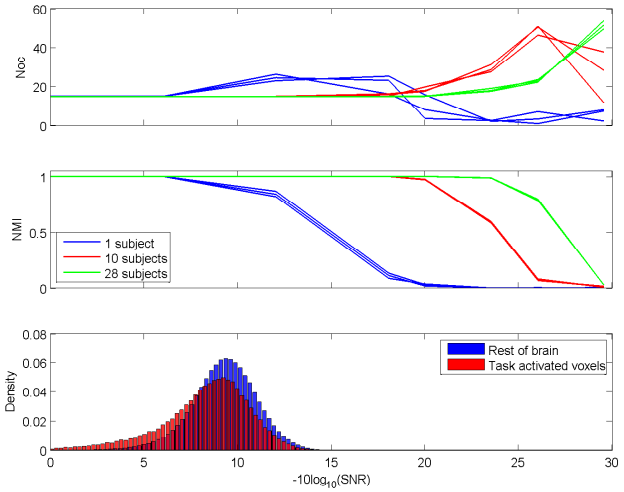

Fig. 2. Number of components extracted as a function of the SNR, NMI using a single subject, and group analysis of 10 and 25 subjects. Below is given the density of the SNR for the real multi-subject fMRI data is given in.

For the inference procedure a full sweep consisted of 1 Gibbs sweep and a number of split-merge moves defined such that the CPU time spent on the split-merge moves matched that of the Gibbs sweep. This means that the number of split-merge moves performed changed dynamically during the model inference. In each split-merge move 3 restricted Gibbs sweeps were performed. Additionally 10 sweeps of Metropolis-Hastings hyper-parameter sampling was performed for each $\sigma_{i,s}^2$ and for γ. On the fMRI dataset consisting of 28 subjects and 48799 voxels each with a time-series of 240 measurements (further details on the data is given below) this entire sampling forming one iteration took approximately one hour to complete.

To test the model on the synthetic data we performed 3 runs for each selection of noise and number of synthetic subjects. On the fMRI dataset 10 runs using the accelerated split-merge procedure and 10 runs with the standard split-merge procedure were performed to illustrate the impact on convergence of the change in the split-merge procedure.

3.1. Synthetic data

In order to investigate the level of noise for which the model can infer the correct clustering, a number of synthetic data sets were generated of varying noise. Furthermore we varied the number of subjects to illustrate how the performance of the model increases with more subjects. For each dataset we generated 15 cluster means for each subject according to Eq. (2) and (11) with a characteristic length-scale of 1.85. For each cluster we generated 400 voxels with the same temporal dimension as the fMRI dataset. This was done independently for each of the synthetic subjects such that the only thing shared was the clustering configuration. This was done 3 times for each noise level and number of subjects pair in order to verify the

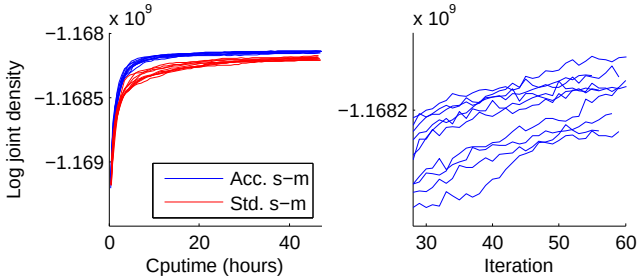

Fig. 3. Progression of the joint distribution given in equation (5) using standard split-merge sampling and using the proposed accelerated merge procedure. On the right are the last 30 iterations of the accelerated chains.

stability of the method.

To quantify the extend in which the clustering matches the ground truth we used the normalized mutual information (NMI) [16] measure as well as the number of clusters inferred as a function of the average SNR of the generated data defined by

$$\text{SNR}_{\text{DB}} = 10\log_{10}\frac{\text{P}_{\text{signal}}}{\text{P}_{\text{noise}}} = 10\log_{10}\frac{\sum_{s,k} n_k \boldsymbol{\mu}_k^\top \boldsymbol{\mu}_k}{\sum_{s,n} T\sigma_{s,n}^2}. \quad (12)$$

As seen in Fig. 2, the method is able to handle Gaussian noise until a SNR_{DB} of approximately -5. After that additional subjects are needed in order to be able to infer the correct configuration. In order to get a crude estimate of the SNR in the used fMRI data we used the following estimate: For the mean of the signal we used the 10 repetitions of the task to construct the mean, $\bar{\boldsymbol{\mu}}_v^{(s)} = \frac{1}{10}\sum_{i=1}^{10} \boldsymbol{x}_v^{(s)}$, and then estimates the SNR by:

$$\text{SNR}^2_{v,\text{estimate}} = \frac{\bar{\boldsymbol{\mu}}_v^{(s)} \bar{\boldsymbol{\mu}}_v^{(s)\top}}{\frac{1}{10-1}\sum_{i=1}^{10}(\boldsymbol{x}_v^{(s)} - \bar{\boldsymbol{\mu}}_v^{(s)})(\boldsymbol{x}_v^{(s)} - \bar{\boldsymbol{\mu}}_v^{(s)\top})}. \quad (13)$$

According to this crude estimate the model should not be able to correctly infer the correct configuration for a single subject. However, this should be possible when using 10 subjects and in our regime using 28 subjects.

3.2. Multi-subject fMRI

To validate our proposed method we used a fMRI finger tapping data set consisting of 28 healthy subjects scanned in a Siemens 3T scanner. The dataset has previously been described in [17, 18]. The finger tapping paradigm consisted of two paced motor conditions each lasting 20 s, first right handed finger tapping followed by left handed finger tapping. Both conditions were paced by a blinking colored circle and were followed by 9.88 s rest. The stimulation cycle was repeated 10 times and 240 scans was acquired in total. Data was preprocessed using a default strategy in the SPM8 software package that comprised the following steps: (1) Rigid body realignment, (2) co-registration, (3) spatial-normalization to the MNI 152 template, (4) re-slicing of images into MNI space at 3 mm isotropic voxels. For the SPM analysis a spatial smoothing was further applied using an isotropic Gaussian filter (6 mm FWHM). Finally a rough grey matter mask (48799 voxels) was applied.

Fig. 3 shows the logarithm of the joint distribution for the 10 different runs. It is clear that the accelerated split-merge procedure using enhanced merge steps significantly improves on the convergence. We also observe that even using this enhanced inference procedure the model has not converged.

In order to show how many clusters are task relevant we sorted the clusters according to the cluster specific mean correlation $R(k)$ computed in Eq. (9). As seen from Fig. 4 it is clear that two clusters show a much higher degree of correlation across subjects whereas 8 clusters show a correlation higher than 0.3. Of these 8 clusters 7 show a consensus score (i.e., $C(k)$) higher than 0.6. These 7 clusters are colored in shades of red and are the clusters of similar color in the consensus score plot in Fig. 4. The cluster that shows a high level of correlation but a low consensus score is colored yellow. The 8 clusters having high correlation are also visualized in figure A)-H) of Fig. 6 and shown in descending order according to their correlation score colored as in Fig 4. According to the consensus score we also selected the 7 clusters with the highest consensus score, shown in shades of green. These 7 clusters are also shown in figure I)-O) of Fig. 6.

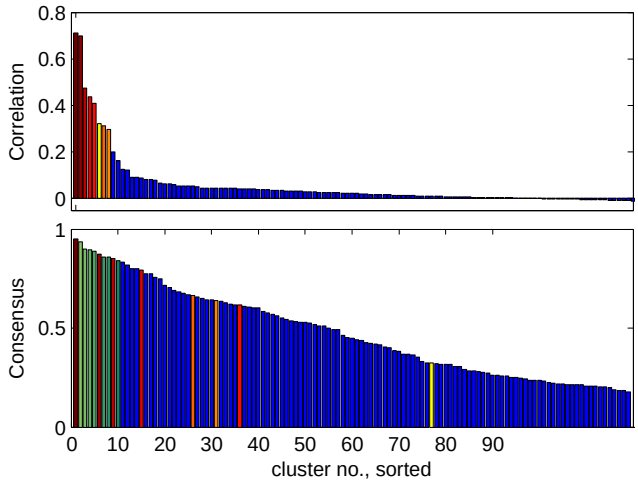

Fig. 4. Average correlation across the estimated cluster specific time-series for each subject and stability as quantified using evidence accumulation. In shades of red is given clusters that have corelation above 0.3. In green the remaining 7 clusters that are in top 10 of the consensus score.

We performed a standard multisubject SPM analysis on the smoothened data with left-right and right-left contrast maps for a comparison with the regions extracted by our method. The activation maps of the SPM analysis are thresholded and compared with the two clusters of average correlation higher than 0.7 in Fig. 5. From the figure it is clear that there is a very high degree of correspondance but also that the two top correlated maps are more localized. The SPM maps are also included in the top of Fig. 6 where it can be seen that the 10 most correlated clusters well correspond to subparcellations of the SPM maps of regions that are task activated whereas the regions with a relative high consensus score but relatively low correlation do not delineate regions that are extracted in the SPM analysis but different cortical regions that robustly group together.

4. CONCLUSION

When analyzing fMRI the data is traditionally smoothened and voxels of brain activation extracted in a supervised manner for instance

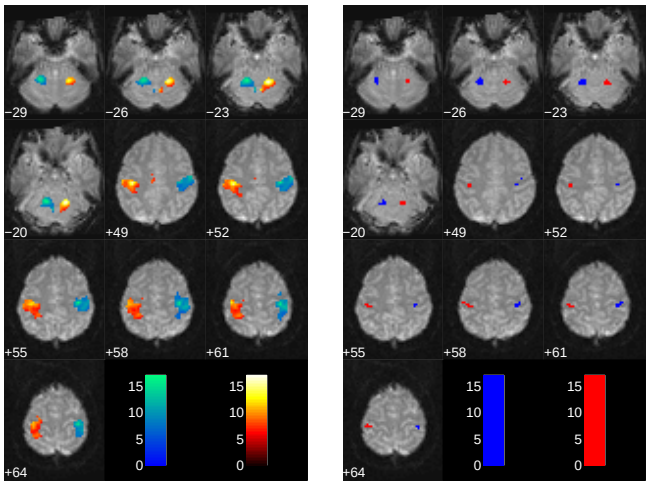

Fig. 5. On the left are SPM heatmaps of activated voxels at selected axial slices. On the right are the inferred clusters of highest cross subject correlation.

using a SPM analysis that identifies voxels with the expected temporal evolution as defined by the imposed HRF convolved with the design matrix. In contrast, our proposed approach is fully unsupervised and automatically groups voxels circumventing the need for smoothing data. It uses the temporal consistency across subjects as well as reliability over separate chains of the sampler in order to infer regions of interests. We find that regions with high temporal consistency well correspond to those derived by a standard SPM analysis whereas regions that are only reliable across chains of the sampler correspond to cortical regions that are neither identified by SPM nor our measure of correlation.

We succesfully demonstrated on a simple finger tapping paradigm that our completely unsupervised approach is able to extract the task-induced activations. This we believe forms a promising framework for the analysis of taskdata in general where there is no good knowledge of how given tasks induce BOLD responses. We also find that there is generally correspondence between regions that correlate across subjects and regions that are robustly identifiifed by our inference procedure. This indicates that consistency of the clustering by itself can be used to idenfity task relevant regions and can thereby be used to quantify activated regions when information on task is unavailable such as in the analysis of resting state fMRI.

5. REFERENCES

[1] Steven D Forman, Jonathan D Cohen, Mark Fitzgerald, William F Eddy, Mark A Mintun, and Douglas C Noll, "Improved assessment of significant activation in functional magnetic resonance imaging (fmri): use of a cluster-size threshold," *Magnetic Resonance in medicine*, vol. 33, no. 5, pp. 636–647, 1995.

[2] K J Worsley, S Marrett, P Neelin, A C Vandal, K J Friston, and A C Evans, "A unified statistical approach for determining significant voxels in images of cerebral activation," *Human Brain Mapping*, vol. 4, pp. 58–73, 1996.

[3] J Chumbley, Keith Worsley, Guillaume Flandin, and K Friston, "Topological fdr for neuroimaging," *Neuroimage*, vol. 49, no. 4, pp. 3057–3064, 2010.

[4] Cyril Goutte, Peter Toft, Egill Rostrup, Finn Å Nielsen, and Lars Kai Hansen, "On clustering fmri time series," *NeuroImage*, vol. 9, no. 3, pp. 298–310, 1999.

[5] Ruth Heller, Damian Stanley, Daniel Yekutieli, Nava Rubin, and Yoav Benjamini, "Cluster-based analysis of fmri data," *NeuroImage*, vol. 33, no. 2, pp. 599–608, 2006.

[6] M J McKeown, T P Jung, S Makeig, G Brown, S S Kindermann, T W Lee, and T J Sejnowski, "Spatially independent activity patterns in functional MRI data during the stroop color-naming task," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 95, pp. 803–810, Feb. 1998.

[7] J S Damoiseaux, S.A.R.B. Rombouts, F Barkhof, P Scheltens, C J Stam, S M Smith, and C F Beckmann, "Consistent resting-state networks across healthy subjects," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 37, pp. 13848–13853, Sept. 2006.

[8] VD Calhoun, T Adali, VB McGinty, JJ Pekar, TD Watson, and GD Pearlson, "fmri activation in a visual-perception task: network of areas detected using the general linear model and independent components analysis," *NeuroImage*, vol. 14, no. 5, pp. 1080–1088, 2001.

[9] 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*, vol. 13, no. 1, 2004.

[10] Ana LN Fred and Anil K Jain, "Combining multiple clusterings using evidence accumulation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 27, no. 6, pp. 835–850, 2005.

[11] Srikanth Ryali, Tianwen Chen, Aarthi Padmanabhan, Weidong Cai, and Vinod Menon, "Development and validation of consensus clustering-based framework for brain segmentation using resting fmri," *Journal of neuroscience methods*, vol. 240, pp. 128–140, 2015.

[12] David J Aldous, *Exchangeability and related topics*, Springer, 1985.

[13] Jim Pitman et al., "Combinatorial stochastic processes," Tech. Rep., Technical Report 621, Dept. Statistics, UC Berkeley, 2002. Lecture notes for St. Flour course, 2002.

[14] Carl Edward Rasmussen, "The infinite gaussian mixture model.," in *NIPS*, 1999, vol. 12, pp. 554–560.

[15] James Ross and Jennifer Dy, "Nonparametric mixture of gaussian processes with constraints," in *Proceedings of the 30th International Conference on Machine Learning (ICML-13)*, 2013, pp. 1346–1354.

[16] LNF Ana and Anil K Jain, "Robust data clustering," in *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*. IEEE, 2003, vol. 2, pp. II–128.

[17] 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*, vol. 45, no. 6, pp. 2085–2100, 2012.

[18] Peter Mondrup Rasmussen, Trine Julie Abrahamsen, Kristoffer Hougaard Madsen, and Lars Kai Hansen, "Nonlinear denoising and analysis of neuroimages with kernel principal component analysis and pre-image estimation," *NeuroImage*, vol. 60, no. 3, pp. 1807–1818, 2012.

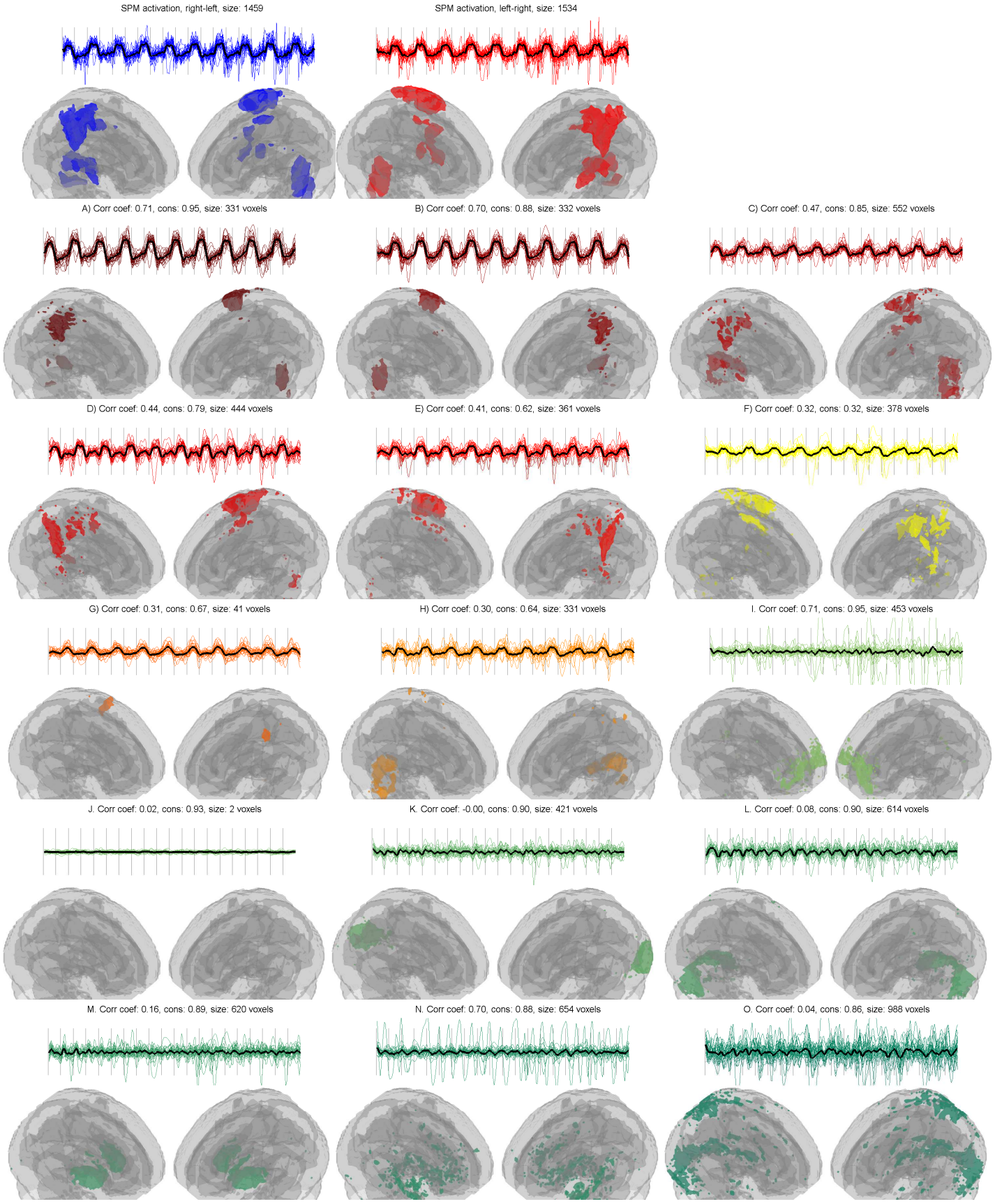

Fig. 6. The activation maps for the two conditions using group SPM analysis and the average of the time-series of the activated voxels. In A)-O) are the extraced functional activation maps using our fully unsupervised approach. A)-H) are the clusters of high cross subject correlation, and only F) has a consensus score lower than 0.6. In I)-O) are the reliable maps that are not highly correlated in time across subjects.