

## Whole brain clustering on the basis of Resting-State fMRI data

H. M. Fonteijn<sup>1</sup>, and D. G. Norris<sup>2</sup>

<sup>1</sup>Department of Computer Science, Centre for Medical Imaging and Computing, London, United Kingdom, <sup>2</sup>F.C. Donders Centre for Cognitive Neuroimaging, Nijmegen, Netherlands

### Introduction

In the last decade, there is increasing evidence that low frequency components of the BOLD response of functionally associated areas are correlated. This was first shown by Biswal et al. [1], who used functionally defined areas in the motor system for a classical correlation analysis. Another approach was developed by Salvador et al. [2], who used an anatomically defined template to define seed regions for a partial correlation analysis. Recently, more data-driven methods, such as ICA [3], have been used to distinguish the networks active in the brain. They have thus far only shown large networks of brain areas, which leaves the question open, at what level the individual parts of the brain are organized in resting state. In this paper we will approach this question using clustering algorithms on resting-state fMRI (RS fMRI) data. We will determine the number of clusters needed to represent a whole-brain data set and we will develop a measure which indicates the variance of data points within each cluster and assess how this varies from cluster to cluster.

### Theory

To perform clustering on RS fMRI data two steps are required: First a suitable preprocessing procedure must be chosen. We have investigated the need for band-pass filtering of the RS signals together with some form of data reduction and we have previously shown [4] that choosing Discrete Wavelet Transform detail coefficients of scale 4 (corresponding to data between 0.01 - 0.02 Hz) gives an optimal compromise between these two requirements. Second, a clustering algorithm must be chosen, which allows maximal flexibility in the shape of clusters and is also suitable to determine the number of clusters from. Model based clustering, and more specifically Gaussian mixture models (GMM's), satisfy these criteria. In Gaussian mixture models, each cluster is represented as N-dimensional Gaussian (N is the number of dimensions used to represent the data, in this case the number of DWT detail coefficients, provided by the preprocessing step). The mean of each cluster represents its typical time course and the covariance matrix indicates the variance around this time course within this cluster. Usually, the GMM is fitted using the expectation maximization algorithm. Because for each number of clusters, a likelihood of the data, given the parameters of a model with this number of clusters, can be calculated, which can then be used for optimal model selection using for instance the Bayesian Information Criterion. Recently however a fully Bayesian Gaussian mixture model has been proposed [5] in which the number of clusters is also treated as an unknown model parameter. This model is called the infinite mixture model, because at each stage of the estimation procedure an infinite number of clusters is assumed. However, only a finite number of clusters is 'occupied' by data, which makes the model tractable. The allocation of data points to clusters is achieved by using a Dirichlet Process, which allocates data to clusters, preferring a priori clusters with already a large number of data points, but also with an a priori finite probability of assigning data to an empty cluster. The parameters of this model were estimated using a Variational Bayes scheme, as implemented in [6].

### Methods

**Data acquisition:** After informed consent was given, eight subjects participated in a resting state study. The subjects were scanned on a 3 T Siemens Trio scanner, using the eight-channel phased array head coil. Resting-state data were acquired by using gradient-echo EPI with the following imaging parameters: TR 1400 ms, flip angle 67° to conform to the Ernst angle for this TR, TE 30 ms, 21 slices, matrix size 64×64, resolution 3.5×3.5×5.0 mm, distance factor 20%, 1030 volumes, bandwidth 1815 Hz/pixel, scan time 25 min.

**Image analysis:** The first six volumes of the resting state scans were discarded to compensate for T1 relaxation effects. The remaining volumes were aligned and then normalized to the EPI template in standard space. We masked out white matter by applying a segmentation algorithm on the mean normalized EPI image, which results in a crude estimate of gray and white matter and csf. For each voxel in grey matter the DWT scale 4 detail coefficients were calculated using a 'sym4' wavelet basis function, as implemented by the Matlab Wavelet toolbox. This analysis results in 70 coefficients per voxel, which is equivalent to a 15-fold data reduction. The resulting time courses were then normalized to have zero mean and unit variance and were then entered in the clustering algorithm. The multi-subject data set was formed by concatenating all the data sets in the time direction and by then performing a principal component analysis. The 70 components which represented the highest variance were retained, this formed a data set with the same dimensionality as the single subject data sets. To study the variance of each cluster, we applied an eigenvector/eigenvalue decomposition on the cluster covariance matrices of the clusters from the multi-subject results. The mean eigenvalue of each cluster gives an indication of the variability between individual time-courses within the cluster and between subjects.

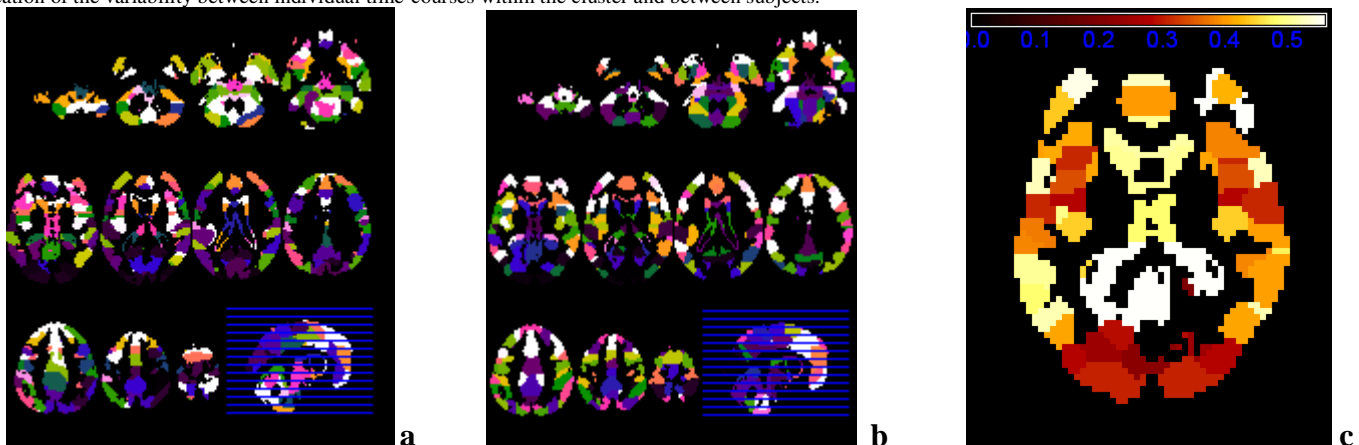


Figure 1. a Results of the clustering algorithm in a single subjects. 90 Clusters are found. b. results of the clustering algorithm on the multi-subject data set. 110 Clusters are found. Please note that similarity in color does not necessarily mean that these clusters are identical. c. Mean eigenvalues for each cluster in the multi-subject analyses. The mean eigenvalues are especially low around the primary cortices.

**Results and Conclusion:** As can be seen in figure 1, RS clustering leads to highly symmetric clusters throughout the cortex and results in a similar number of clusters than for instance the AAL atlas. These results are even stronger in the multi-subject results, especially in terms of symmetry. Figure 1c indicates that the variance within the clusters are relatively low around the primary cortices and higher in prefrontal and parietal cortex. In conclusion, a feasible subdivision of the human brain has been shown on the basis of resting-state data.

References: 1. Biswal, B. et al. *MRM* 1995, **37**(4): p.537-41. 2. Salvador, R. et al. *Cer Cor* 2003, **15**(9): p. 1332-42. 3. Beckmann, C.F. et al. *Phil Trans R Soc B* 2005, **360**: p. 1001-13. 4. Fonteijn, H. Et al. *ISMRM* 2008. 5. Blei, D.M. and Jordan, M.I. *Bayesian Analysis* 2005, **1**: p. 121-144. 6. <http://kenichi.kurihara.googlepages.com/vdpmog.html>