CHAPTER

Functional connectivity parcellation of the human brain

1

A. SchaeferSupported by a fellowship within the Postdoc-Program of the German Academic Exchange Service (DAAD)., R. Kong, B.T. Thomas Yeo

National University of Singapore, Singapore

CHAPTER OUTLINE

1.1 Introduction 3
1.2 Approaches to Connectivity-Based Brain Parcellation 5
1.3 Mixture Model 7
1.3.1 Model 7
1.3.2 Inference 9
1.4 Markov Random Field Model 12
1.4.1 Model 12
1.4.2 Inference 16
1.5 Summary 21
References 22

1.1 INTRODUCTION

Brain disorders, comprising psychiatric and neurological disorders, are seen as one of the core health challenges of the 21st century (Wittchen et al., 2011). Their prevalence in developed countries surpasses those of cardiovascular diseases and cancer (Collins et al., 2011). Furthermore, while there have been strong advances in treatment of cardiovascular diseases, which translates to saving millions of lives each year, this progress has been absent in the treatment of psychiatric disorders (Insel, 2009). In order to develop new treatments, we need a deeper understanding of brain organization and function.

The largest structure of the brain is the cerebral cortex, which has the topology of a 2-D sheet, and is responsible for many higher order functions such as consciousness, memory, attention, and language. The cerebral cortex can be subdivided into a larger number of different areas (Brodmann, 1909; Vogt and Vogt, 1919). Identifying these areas is important as it is believed that complex human behavior

Machine Learning and Medical Imaging. http://dx.doi.org/10.1016/B978-0-12-804076-8.00001-3

is mainly enabled by their interaction. For example, it has been shown that visual information is processed by distinct parallel pathways through the brain (Ungerleider, 1995). While each of the areas along these pathways are specialized, only their interplay allows a complex process such as visual recognition. To understand these complex interactions, identification of the distinct functional areas is needed. A map of all cortical areas is called a cortical parcellation.

Cortical areas are defined by their distinct microarchitecture, connectivity, topology, and function (Kaas, 1987; Felleman and Van Essen, 1991). Ideally we want to estimate all of these features in vivo as the location of cortical areas can vary between different subjects by more than 10 mm (Amunts et al., 1999; Fischl et al., 2008). As a consequence their results cannot be accurately translated to other (living) subjects. However, microarchitecture of cortical areas is commonly estimated by ex vivo methods which combine staining and chemicals to estimate myelo-, cyto-, or chemo-architectonics (Zilles and Amunts, 2010). While efforts have been made to develop noninvasive neuroimaging methods (Mackay et al., 1994; Glasser and Van Essen, 2011; Lodygensky et al., 2012), their resolution is much coarser compared to ex vivo methods. The function of different cortical locations can be approximated by in vivo task-based activation studies (Belliveau et al., 1991) or lesion studies (Rorden and Karnath, 2004; Preusser et al., 2015). These studies are often focused on a single brain location at a time and are therefore not ideally suited for identifying all cortical areas. Meta-analytic approaches allow the combination of hundreds or thousands of these studies (Eickhoff et al., 2009; Yarkoni et al., 2011; Fox et al., 2014; Gorgolewski et al., 2015), which can then be used to derive cortical maps (Eickhoff et al., 2011; Yeo et al., 2015). Topology can also be derived from functional activation studies, for example, for the visual cortex (Sereno et al., 1995; Swisher et al., 2007). Although it is not possible to solely use this feature to create a complete cortical map, it may be combined with other features.

In contrast, connectivity can be estimated for the whole brain within minutes via noninvasive imaging methods (Biswal et al., 2010; Craddock et al., 2013; Eickhoff et al., 2015). Functional connectivity is commonly defined as the synchronization of functional activation patterns (Friston, 1994; Biswal et al., 1995; Smith et al., 2011). In recent years, functional connectivity has been widely used as a feature for estimating brain parcellation (Shen et al., 2010; Yeo et al., 2011; Craddock et al., 2012; Varoquaux and Craddock, 2013). Two examples are shown in Fig. 1.1. Structural connections can be assessed by water diffusivity, which aims to identify white matter axons (Tuch et al., 2003; Hagmann et al., 2007; Johansen-Berg and Behrens, 2013). Connectivity poses some advantages compared to other features as it is not bounded to a specific task and can therefore be assessed in a single experiment.

Human brain parcellation has become an exciting field for the application of machine learning approaches as the dimensionality and diversity of the acquired data keeps increasing. With the rise of large-scale imaging initiatives (Biswal et al., 2010; Van Essen et al., 2012; Nooner et al., 2012; Zuo et al., 2014; Holmes et al., 2015), vast amounts of multimodal brain information have become publicly available, providing information about connectivity, function, and anatomy. These large datasets pose fascinating challenges for machine learning approaches.

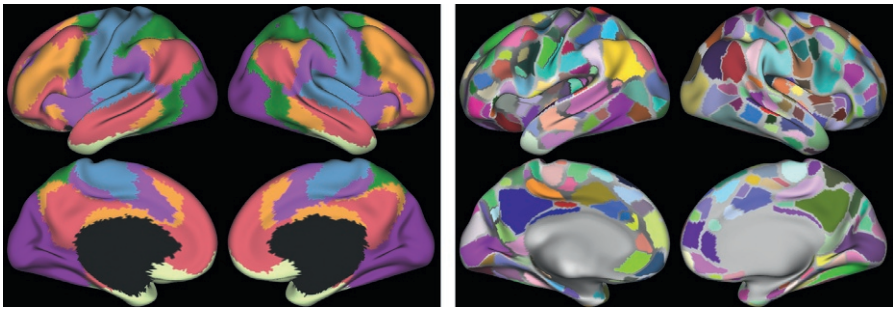

FIG. 1.1

Cortical labels based on functional connectivity. (Left) Average functional connectivity of 1000 subjects modeled with a mixture of von Mises-Fisher distributions and clustered into networks of areas with similar connectivity patterns (Yeo et al., 2011). (Right) Areas derived from estimating gradients of functional connectivity and applying a watershed transform (Gordon et al., 2014).

In this chapter, we provide an overview of current approaches to connectivity-based parcellations with particular emphasis on mixture models and Markov random fields (MRFs) in Sections 1.3 and 1.4, respectively.

1.2 APPROACHES TO CONNECTIVITY-BASED BRAIN PARCELLATION

This section provides a general overview of different approaches to brain parcellation based on connectivity features. We will cover mixture and MRF models in more details in Sections 1.3 and 1.4.

In connectivity-based brain parcellations we want to model the relationship between connectivity features and cortical labels. Let us assume N brain locations denoted by $x_1, \ldots, x_N$. Let Y_n be the connectivity at brain location x_n. We assume that each connectivity feature Y_n is D-dimensional. We further assume L cortical labels of interest with $l_n \in \{1, \ldots, L\}$. Our goal is to estimate $l_1, \ldots, l_N$ for the N brain locations $x_1, \ldots, x_N$. Often we will just write $l_{1:N}$ instead of $l_1, \ldots, l_N$. Finding an assignment $l_{1:N}$ can be seen as a segmentation problem, which is a well-studied objective in pattern recognition and machine learning.

Common approaches to brain networks estimation include independent component analyses (ICAs) (Bell and Sejnowski, 1995; Calhoun et al., 2001; Beckmann and Smith, 2004), which in the context of brain parcellations is a form of temporal demixing of spatial factors. An ICA is commonly formulated as a matrix decomposition problem. More specifically, the decomposition is

$$S = EY, \tag{1.1}$$

where Y is a $D \times N$ matrix of D-dimensional observations for each of the N brain locations, S is an $L \times N$ matrix of L hidden factors, and E is an $L \times D$ demixing matrix. Here the matrix S represents a soft estimate of the labels. An ICA maximizes the independence of the signals in S. This independence assumption is sometimes criticized for its limited biological validity (Harrison et al., 2015). An ICA is a soft "clustering" algorithm which may result in overlapping areas and is therefore not directly a solution to the parcellation problem. Manual or automatic thresholding techniques (eg, random walks) can be applied to retrieve a parcellation (Abraham et al., 2014). ICAs belong to the class of linear signal decomposition models, which includes dictionary learning. This more general form can, for example, be used to build hierarchical models that model the variability between subjects (Varoquaux et al., 2011).

K-means clustering (Lloyd, 1982; Jain, 2010) is another widely used approach for connectivity-based parcellation (Mezer et al., 2009; Bellec et al., 2010; Kim et al., 2010; Cauda et al., 2011; Zhang and Li, 2012; Mars et al., 2012). The approach performs a mapping to a preselected number of K nonoverlapping clusters. In our notation K equals L. K-means clustering is performed in an iterative fashion by first hard-assigning the labels of the N brain regions to their respective closest centers. Then the L cluster centers are recomputed. The whole process is repeated until convergence. The initial cluster centers are usually assigned randomly and the final labeling is highly dependent on this initial assignment. Consequently, the algorithm is typically repeated many times and the solution with the best cost function value is selected.

Spectral clustering (Jianbo Shi and Malik, 2000; Ng et al., 2001) is also widely used for connectivity-based parcellation (Johansen-Berg et al., 2004; Thirion et al., 2006; van den Heuvel et al., 2008; Shen et al., 2010; Craddock et al., 2012). The approach is based on the affinity between each pair of locations x_i and x_j. Affinity is typically computed by some form of similarity between the corresponding observations y_i and y_j. In our case this similarity is often the similarity of structural or functional connectedness of x_i and x_j. Based on the affinity matrix A we can compute a Laplacian (von Luxburg, 2007), for example,

$$L = DA, \tag{1.2}$$

where D is the degree matrix $D(i,i) = \sum_j A(i,j)$. The idea of spectral clustering is to decompose this matrix into the eigenvectors of L. Then the assumption is that the data points are easier to separate in this eigenspace than in the original space. A related approach identifies areas that maximize modularity (Meunier et al., 2010; He et al., 2009). Modularity is maximized by selecting areas which have high within-module connections and low between-module connections. Modularity maximization is NP-hard but can be approximated by spectral methods (Newman, 2006).

Agglomerative hierarchical clustering (Eickhoff et al., 2011; Michel et al., 2012; Blumensath et al., 2013; Orban et al., 2015; Thirion et al., 2014; Moreno-Dominguez et al., 2014) is a bottom-up approach in which initially every brain location x_n has a separate label l_n. The agglomerative clustering builds a hierarchy by iteratively

merging two labels based on a linkage criterion. This linkage criterion needs to be defined, for example, as the average similarity between all the connectivity features belonging to the two labels. The derived hierarchical organization allows a multiresolution parcellation by thresholding the hierarchy at different levels.

Gradient approaches identify rapid transitions in the connectivity pattern of adjacent regions (Cohen et al., 2008; Gordon et al., 2014; Wig et al., 2014). For example, transitions can be identified by applying a Canny edge detector (Canny, 1986) on the cortical surface. Gradient approaches have been applied on parts of the cortex (Cohen et al., 2008; Nelson et al., 2010a,b; Hirose et al., 2012), as well as the entire cerebral cortex (Wig et al., 2014; Gordon et al., 2014). A cortical parcellation (Fig. 1.1, right) can be derived by applying a watershed transform (Beucher and Lantuejoul, 1979) on the resulting gradient maps (Gordon et al., 2014). The first step of the watershed transform involves identifying local gradient minima as seed regions. The seed regions are then iteratively grown by including neighboring brain locations that have a gradient value lower than the current threshold. The threshold is iteratively increased until every location belongs to one of the seed regions corresponding to the local minima.

1.3 MIXTURE MODEL

In this section, we provide a more detailed introduction to the mixture model approach. We will show examples and describe a popular inference approach to learn the mixture parameters and labels.

Mixture models are a flexible way of clustering as they allow soft or weighted assignments. The value of the weight indicates the strength of the affinity to the corresponding cluster. In mixture models, this weight corresponds to the posterior probability that the data point belongs to a mixture component. A final parcellation or labeling can be gained by thresholding the posterior probabilities.

1.3.1 MODEL

A mixture model is a combination of component models which form a richer model:

$$p(Y) = \sum_{l=1}^{L} p(Y|l)p(l|\alpha_l). \tag{1.3}$$

The features Y are observed, whereas the labels l are hidden to us. The component models are given by $p(Y|l)$ and the prior probabilities of the different components are given by α_l, where α_l is non-negative and $\sum_{l=1}^{L} \alpha_l = 1$. We can interpret the above equation as a generative model to sample a data point Y. The probability of label l is α_l. We first draw a label l with $p(l|\alpha_l)$, and then draw an observation Y from $p(Y|l)$.

Let the probability distribution $p(Y|l)$ be parameterized by θ_l. Then the model becomes

$$p(Y|\Theta) = \sum_{l=1}^{L} p(Y|\theta_l)p(l|\alpha_l), \tag{1.4}$$

where $\Theta = \{\alpha_{1:L}, \theta_{1:L}\}$.

For our parcellation, we assume that each label l_n is independently drawn from a probability distribution $p(l_{1:N}) = \prod_n p(l_n)$. The features $Y_1, \ldots, Y_N$ are assumed to be independent conditioned on $\Theta = \{\alpha_{1:L}, \theta_{1:L}\}$, and so the mixture model is of the form:

$$p(Y_{1:N}|\Theta) = \prod_{n=1}^{N} p(Y_n|\Theta) = \prod_{n=1}^{N} \sum_{l_n=1}^{L} p(Y_n|\theta_{l_n})p(l_n|\alpha_{l_n}). \tag{1.5}$$

Conditioned on cortical label l_n at spatial location x_n, the observed features Y_n are assumed to be generated from the distribution $p(Y_n|\theta_{l_n})$.

In the case of a Gaussian mixture with parameter $\theta_{l_n} = \{\mu_{l_n}, \Sigma_{l_n}\}$ (Golland et al., 2007; Tucholka et al., 2008; Jbabdi et al., 2009), the distribution is set to

$$p(Y_n|\theta_{l_n}) = \mathcal{N}(Y_n|\mu_{l_n}, \Sigma_{l_n}), \tag{1.6}$$

where the Gaussian distribution $\mathcal{N}(Y_n|\mu_{l_n}, \Sigma_{l_n})$ is defined as

$$\mathcal{N}(Y_n|\mu_{l_n}, \Sigma_{l_n}) = \frac{1}{\sqrt{\det(2\pi\,\Sigma_{l_n})}} \mathrm{e}^{-\frac{1}{2}(Y_n-\mu_{l_n})^{\mathrm{T}}\Sigma_{l_n}^{-1}(Y_n-\mu_{l_n})}, \tag{1.7}$$

where μ_{l_n} is the mean and Σ_{l_n} is the covariance matrix of cluster l_n.

Another example is a mixture of von Mises-Fisher distributions with parameter $\theta_{l_n} = \{\mu_{l_n}, \kappa\}$ (Yeo et al., 2011; Ryali et al., 2013; Liu et al., 2014), where

$$p(Y_n|\theta_{l_n}) = \mathrm{vmf}(Y_n|\mu_{l_n}, \kappa). \tag{1.8}$$

The von Mises-Fisher distribution $\mathrm{vmf}(Y_n|\mu_{l_n}, \kappa)$ is defined as

$$\mathrm{vmf}(Y_n|\mu_{l_n}, \kappa) = z_D(\kappa)\mathrm{e}^{\kappa\mu_{l_n}^{\mathrm{T}}Y_n}, \tag{1.9}$$

with mean direction $||\mu_{l_n}|| = 1$, concentration parameter κ and dimensionality $D \geq 2$. The normalizing constant $z_D(\kappa)$ is given by

$$z_D(\kappa) = \frac{\kappa^{D/2-1}}{(2\pi)^{D/2} I_{D/2-1}(\kappa)}, \tag{1.10}$$

where $I_r(\cdot)$ represents the modified Bessel function of the first kind and order r. The graphical representation of the von Mises-Fisher mixture model is given in Fig. 1.2.

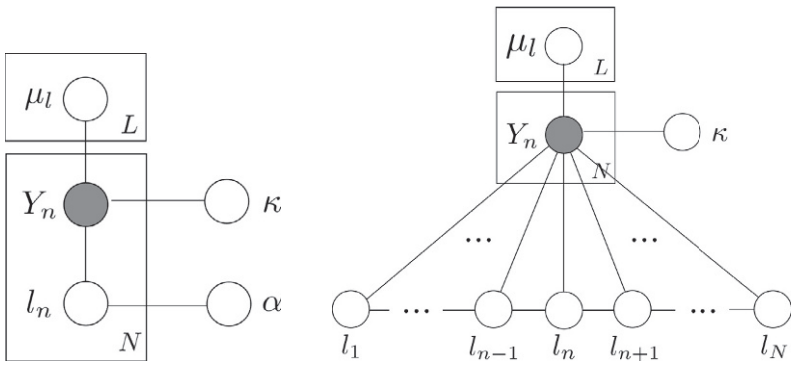

FIG. 1.2

Graphical representation of von Mises-Fisher mixture model with and without MRF. Shaded variables are observed. All other variables are hidden. Lines indicate dependencies between the variables. Any part of a graphical model within a plate with subscript X is replicated X times, where all dependencies between variables outside of the plate and variables on the same plate are preserved. There are no dependencies between elements on different replicates of a plate. (Left) Graphical model for a mixture of von Mises-Fisher distributions model (Yeo et al., 2011) (Section 1.3). Each label l_n is independently drawn from a probability distribution $p(l_{1:N}) = \prod_n p(l_n)$. The hidden label l_n indexes the mixture component. The probability of l_n is given by α. Given the label l_n and the corresponding von Mises-Fisher parameters κ and μ_{l_n}, the observations Y_n are generated via $p(Y_n|l_n, \mu_{l_n}, \kappa)$. (Right) Graphical model for a mixture of von Mises-Fisher distributions together with an MRF prior (Ryali et al., 2013) (Section 1.4). The MRF prior comprises dependencies between the labels of neighboring brain locations, for example, label l_n at brain location x_n is dependent on the labels at neighboring brain locations.

1.3.2 INFERENCE

The labels $l_{1:N}$ and parameters Θ are unknowns that need to be estimated. The optimal parameters Θ are typically estimated by maximum likelihood estimation:

$$\arg\max_{\Theta} p(Y_1, \ldots, Y_N|\Theta) = \arg\max_{\Theta} \prod_n p(Y_n|\Theta). \tag{1.11}$$

In other words we want to find the parameters that maximize the likelihood of our observations. A common approach to estimate hidden variables and parameters is expectation-maximization (EM) (Dempster et al., 1977). In EM we use an iterative two-step process. In the E-step we compute the posterior $p(l_{1:N}|Y_{1:N}, \Theta^t)$ of the labels $l_{1:N}$ given the current parameters Θ^t at iteration t. In the M-step we compute the parameters Θ^{t+1} based on the current estimate of the posterior probability $p(l_{1:N}|Y, \Theta^t)$. Notice that EM computes soft assignments instead of hard decisions.

As an example, we provide derivations for the mixture of von Mises-Fisher distributions (Lashkari et al., 2010). The likelihood of the mixture model is

$$p(Y_n|\Theta) = \prod_{n=1}^{N} \sum_{l_n=1}^{L} p(Y_n|\theta_{l_n}) p(l_n|\alpha_{l_n}), \tag{1.12}$$

where $\theta_{l_n} = \{\mu_{l_n}, \kappa\}$ and $p(Y_n|\theta_{l_n}) = z_D(\kappa) e^{\kappa \mu_{l_n}^{\mathrm{T}} Y_n}$ as in Eq. (1.9). To derive EM, we apply log to the likelihood:

$$\log p(Y_{1:N}|\Theta) = \sum_{n=1}^{N} \log p(Y_n|\Theta) = \sum_{n=1}^{N} \log \sum_{l_n=1}^{L} p(Y_n|\theta_{l_n}) p(l_n|\alpha_{l_n}). \tag{1.13}$$

Now we introduce a probability distribution $q_n(l_n)$ over the latent labels:

$$\log p(Y_{1:N}|\Theta) = \sum_{n=1}^{N} \log \sum_{l_n=1}^{L} \frac{p(Y_n|\theta_{l_n}) p(l_n|\alpha_{l_n}) q_n(l_n)}{q_n(l_n)}. \tag{1.14}$$

Using Jensen's inequality we can write

$$\log p(Y_{1:N}|\Theta) = \sum_{n=1}^{N} \log \sum_{l_n=1}^{L} \frac{p(Y_n|\theta_{l_n}) p(l_n|\alpha_{l_n}) q_n(l_n)}{q_n(l_n)} \tag{1.15}$$

$$\geq \sum_{n=1}^{N} \sum_{l_n=1}^{L} q_n(l_n) \log \frac{p(Y_n|\theta_{l_n}) p(l_n|\alpha_{l_n})}{q_n(l_n)}. \tag{1.16}$$

If $q_n(l_n)$ is equal to $p(l_n|Y_n, \Theta)$, the inequality in Eq. (1.16) becomes an equality. Eq. (1.16) is sometimes referred to as the completed log likelihood.

In the expectation or E-step, we estimate the posterior probability of the labels given the current estimate of the parameters. Given the current estimate of $\Theta^t = \{\theta_{1:L}^t, \alpha_{1:L}^t\}$ with $\theta_l^t = \{\mu_l^t, \kappa^t\}$ at iteration t, we compute

$$q_n^{t+1}(l_n) = p(l_n|Y_n, \Theta^t) = \frac{p(Y_n|\theta_{l_n}^t) p(l_n|\alpha_{l_n}^t)}{\sum_{l'=1}^{L} p(Y_n|\theta_{l'}^t) p(l'|\alpha_{l'}^t)}, \tag{1.17}$$

where

$$p(Y_n|\theta_{l_n}^t) = z_D(\kappa^t) e^{\kappa^t \mu_{l_n}^{tT} Y_n}. \tag{1.18}$$

In the M-step, we estimate the parameters $\Theta^{t+1} = \{\theta_{1:L}^{t+1}, \alpha_{1:L}^{t+1}\}$ by maximizing the completed log likelihood (Eq. 1.16) using the current estimate of $q_n^{t+1}(l_n)$:

$$\arg\max_{\Theta} \sum_{n=1}^{N} \sum_{l_n=1}^{L} q_n^{t+1}(l_n) \log \frac{p(Y_n|\theta_{l_n}) p(l_n|\alpha_{l_n})}{q_n^{t+1}(l_n)} \tag{1.19}$$

$$= \arg\max_{\Theta} \sum_{n=1}^{N} \sum_{l_n=1}^{L} q_n^{t+1}(l_n) \log p(Y_n|\theta_{l_n}) p(l_n|\alpha_{l_n}). \tag{1.20}$$

Using Eq. (1.18), we can write the last equation as

$$\arg\max_{\Theta} \sum_{n=1}^{N} \sum_{l_n=1}^{L} q_n^{t+1}(l_n) \log p(l_n|\alpha_{l_n}) + \sum_{n=1}^{N} \sum_{l_n=1}^{L} q_n^{t+1}(l_n) \log p(Y_n|\theta_{l_n}) \tag{1.21}$$

$$= \arg\max_{\Theta} \sum_{n=1}^{N} \sum_{l_n=1}^{L} q_n^{t+1}(l_n) \log \alpha_{l_n} + \sum_{n=1}^{N} \sum_{l_n=1}^{L} q_n^{t+1}(l_n) \log z_D(\kappa) \tag{1.22}$$

$$+ \sum_{n=1}^{N} \sum_{l_n=1}^{L} q_n^{t+1}(l_n)(\kappa \mu_{l_n}^{\mathrm{T}} Y_n).$$

To estimate the parameters $\Theta = \{\mu_{1:L}, \alpha_{1:L}, \kappa\}$ we can write the lower bound (Eq. 1.22) with only the dependent terms. Together with the Lagrange multipliers η_l and β for constraints $\mu_l^{\mathrm{T}} \mu_l = 1$, $\sum_{l=1}^{L} \alpha_l = 1$, we have

$$\mathcal{L}_\mu = \sum_{n=1}^{N} \sum_{l=1}^{L} q_n^{t+1}(l_n = l)(\kappa \mu_l^{\mathrm{T}} Y_n) + \sum_{l=1}^{L} \eta_l (1 - \mu_l^{\mathrm{T}} \mu_l), \tag{1.23}$$

$$\mathcal{L}_\kappa = \sum_{n=1}^{N} \sum_{l=1}^{L} q_n^{t+1}(l_n = l) \log z_D(\kappa) + \sum_{n=1}^{N} \sum_{l=1}^{L} q_n^{t+1}(l_n = l)(\kappa \mu_l^{\mathrm{T}} Y_n), \tag{1.24}$$

$$\mathcal{L}_\alpha = \sum_{n=1}^{N} \sum_{l=1}^{L} q_n^{t+1}(l_n = l) \log \alpha_l + \beta \left(1 - \sum_{l=1}^{L} \alpha_l\right). \tag{1.25}$$

To estimate μ_l, we take the derivative of $\mathcal{L}_\mu$ and set it to zero:

$$\frac{\partial \mathcal{L}_\mu}{\partial \mu_l} = \sum_{n=1}^{N} q_n^{t+1}(l_n = l)\kappa Y_n - 2\eta_l \mu_l = 0, \tag{1.26}$$

$$\mu_l^{t+1} = \frac{\sum_{n=1}^{N} q_n^{t+1}(l_n = l)\kappa Y_n}{2\eta_l}. \tag{1.27}$$

The Lagrange multiplier η_l is determined by the fact that μ_l should be unit norm, and so we get

$$\mu_l^{t+1} = \frac{\sum_{n=1}^{N} q_n^{t+1}(l_n = l) Y_n}{\| \sum_{n=1}^{N} q_n^{t+1}(l_n = l) Y_n \|}. \tag{1.28}$$

The unknown concentration parameter κ was canceled out from the numerator and denominator in the last equation. To estimate κ, we take the derivative of $\mathcal{L}_\kappa$ from

Eq. (1.24) and set it to zero:

$$\frac{\partial \mathcal{L}_\kappa}{\partial \kappa} = \sum_{n=1}^{N}\sum_{l=1}^{L} q_n^{t+1}(l_n = l)\frac{z'_D(\kappa)}{z_D(\kappa)} + \sum_{n=1}^{N}\sum_{l=1}^{L} q_n^{t+1}(l_n = l)(\mu_l^{\mathrm{T}} Y_n) \tag{1.29}$$

$$= N\frac{z'_D(\kappa)}{z_D(\kappa)} + \sum_{n=1}^{N}\sum_{l=1}^{L} q_n^{t+1}(l_n = l)(\mu_l^{\mathrm{T}} Y_n) = 0. \tag{1.30}$$

Solving the above equation for κ is hard (Banerjee et al., 2005), but there exist several approximation methods (Banerjee et al., 2005; Lashkari et al., 2010; Sra, 2011). Lashkari et al. (2010) suggested the following:

$$\kappa^{t+1} = \frac{(D-2)\Gamma}{1-\Gamma^2} + \frac{(D-1)\Gamma}{2(D-2)}, \tag{1.31}$$

where $\Gamma = \frac{1}{N}\sum_{n=1}^{N}\delta(l_n, l)Y_n^{\mathrm{T}}\mu_l$ and $\delta(l_n, l) = 1, \exists n : x_n = l;\ 0$ otherwise. Now we estimate α by taking the derivative of $\mathcal{L}_\alpha$ from Eq. (1.25):

$$\frac{\partial \mathcal{L}_\alpha}{\partial \alpha_l} = \frac{1}{\alpha_l}\sum_{n=1}^{N} q_n^{t+1}(l_n = l) - \beta = 0, \tag{1.32}$$

$$\alpha_l^{t+1} = \frac{1}{N}\sum_{n=1}^{N} q_n^{t+1}(l_n = l). \tag{1.33}$$

More details on inference in mixture models can also be found in the books by Bishop (2006), Barber (2012), and Koller and Friedman (2009).

1.4 MARKOV RANDOM FIELD MODEL

The previous independence assumption of $p(l_{1:N}) = \prod_n p(l_n)$ might be too strong as spatially neighboring locations often belong to the same brain area. Hence it is popular to make the weaker assumption that each $p(l_n)$ is independent given its neighborhood (Liu et al., 2011, 2012, 2014; Ryali et al., 2013; Honnorat et al., 2013, 2015). This is also called the Markov assumption.

1.4.1 MODEL

The Markov assumption can be modeled by MRFs, which are a form of probabilistic graphical models. For this we need an undirected graph which can be defined as: $G = (V, E)$, with vertex set V and edge set E. The graph structure is commonly given by the brain locations and their spatial proximity.

MRFs can then be defined by local joint probabilities which model a relationship of the assigned labels $\{l_1, \ldots, l_N\}$ using a neighborhood system $\mathcal{N}$. Here, the neighborhood system $\mathcal{N}$ constitutes for every vertex $v \in V : \mathcal{N}(v) = \{u \in$

$V : (u, v) \in E\}$. Then the Markov assumption can be written as $p(l_n| - l_n) = p(l_n|l_x \text{ with } x \in \mathcal{N}(n))$, where $-l_n$ are all labels $l_{1:N}$ without l_n. However, this is only a local property. Based on the equivalence of MRFs and Gibbs fields (Besag, 1974; Geman and Geman, 1984), the Hammersly-Clifford theorem states that this local property can be transformed into a global property. A Gibbs field X takes the form of

$$p(X) = \frac{1}{\mathcal{Z}} \mathrm{e}^{-E(X)}, \tag{1.34}$$

where $\mathcal{Z}$ is a normalizing constant that guarantees that the function sums to one (ie, valid probability distribution). This involves a summation over all possible configurations of X:

$$\mathcal{Z} = \sum_{X} \mathrm{e}^{-E(X)}. \tag{1.35}$$

While $\mathcal{Z}$ is not practically tractable it might be approximated, for example, by a pseudo-likelihood (Ryali et al., 2013). The energy function $E(X)$ is the sum over all clique potentials Φ_c over all possible cliques C:

$$E(X) = \sum_{c \in C} \Phi_c(X_c), \tag{1.36}$$

where each clique potential Φ_c only depends on the variables in the corresponding cliques c. A clique is a subgraph in which every pair of vertices is connected. For example, every single vertex or every pair of connected vertices v_i, v_j with $(v_i, v_j) \in E$ are trivial examples of cliques. In the following we will focus on these single and pairwise clique potentials, although more complicated forms are possible. In our context the energy over X (Eq. 1.36) will be an energy over labelings $l_{1:N}$:

$$E(l_{1:N}) = \sum_{n=1}^{N} \Phi(l_n) + \sum_{n=1}^{N} \sum_{j \in N_n} \Phi_{\text{neigh}}(l_n, l_j). \tag{1.37}$$

One common approach is to combine a mixture model with a MRF prior (Jbabdi et al., 2009; Ryali et al., 2013; Liu et al., 2014). The singleton clique potentials $\Phi(l_n)$ often comprise a data cost term. This term gives a penalty for label assignments that fit the data poorly. For example, this can be formulated as the negative log-likelihood (Ryali et al., 2013):

$$\Phi(l_n) \triangleq \Phi_{\text{obs}}(l_n) = -\log p(Y_n|l_n), \tag{1.38}$$

where the likelihood can be given by a von Mises-Fisher distribution (Ryali et al., 2013) as in Eq. (1.9). For the pairwise potential there exist different options. A very common idea (Jbabdi et al., 2009; Ryali et al., 2013; Liu et al., 2014) is the Potts model (Potts, 1952), which includes the penalty that neighboring brain locations x_n and x_j are assigned different labels:

$$\Phi_{\text{neigh}}(l_n, l_j) = \begin{cases} 0 & \text{if } l_n = l_j \\ \gamma & \text{if } l_n \neq l_j. \end{cases}, \tag{1.39}$$

where γ is a positive parameter that is typically manually tuned. The Potts model enforces piecewise spatial consistency of the label assignment. When we combine the potentials from Eqs. (1.38) and (1.39), our energy term (Eq. 1.37) takes the following form:

$$E(l_{1:N}) = \sum_{n=1}^{N} \Phi_{\text{obs}}(l_n) + \sum_{n=1}^{N} \sum_{j \in N_n} \Phi_{\text{neigh}}(l_n, l_j). \tag{1.40}$$

This corresponds to the graph in Fig. 1.2 (right).

In the context of brain parcellations, we would often prefer each parcel to be topologically connected, rather than spatially distributed. Honnorat et al. (2015) provide an elegant MRF solution to enforce topologically connected parcels by first defining one cluster center i for each brain location x_i and comparing the data Y_n of the current location x_n with the data Y_i of all cluster centers i:

$$\Phi_{\text{obs}}(l_n = i) = -Y_i Y_n^{\text{T}}, \tag{1.41}$$

where we assume that each data vector Y_n has been normalized to zero mean and unit variance. We assume each brain location could potentially be a cluster center, that is, $L = N$. Honnorat et al. (2015) then utilized star shape priors to enforce connectedness of each cluster using a distance metric (Veksler, 2008). First, a distance metric between neighboring brain locations is defined as

$$d(n, m) = 1 - Y_n Y_m^{\text{T}}, \tag{1.42}$$

where $m \in \mathcal{N}(n)$. Based on this metric, the distance between any two brain locations x_i and x_j can then be computed by summing over the distances on the shortest path between x_i and x_j.

Recall that every brain location could be a potential cluster center. Accordingly, there exists as many potential cluster centers as brain locations. Then for each brain location x_j the neighboring brain location x_k along the shortest path towards the cluster center i is enforced to have the same label as the center by an infinite weight:

$$\Phi_{i,j,k}(l_j, l_k) = \begin{cases} \infty & \text{if } l_j = i \text{ and } l_k \neq i, \\ 0 & \text{otherwise.} \end{cases} \tag{1.43}$$

Here, in contrast to the Potts model, the pairwise potentials vary spatially.

Assigning a separate label l_n for each brain location x_n would maximize the data cost term. To prevent this overfit, Honnorat et al. (2015) introduce an additional cost (Delong et al., 2010) for the number of labels. The label cost balances a potential overfit by placing a penalty on the number of labels:

$$\Phi_{\text{label}}(l_{1:N}) = c \sum_{l=1}^{L} \delta_l(l), \tag{1.44}$$

where $\delta_l(l) = 1, \exists n : x_n = l;\ 0$ otherwise, and c controls the amount of overall label cost. Therefore, Eq. (1.44) semi-automatically estimates the final number of labels. Parcellations for varying label costs are displayed in Fig. 1.3.

Hence, the overall energy function (Honnorat et al., 2015) can be written as

$$E(l_{1:N}) = \sum_{n=1}^{N} \Phi_{\text{obs}}(l_n) + \Phi_{\text{label}}(l_{1:N}) + \sum_{n=1}^{N} \sum_{j=1}^{N} \Phi_{n,j,k}(l_j, l_k), \tag{1.45}$$

where k is the neighbor of j that is closest to n.

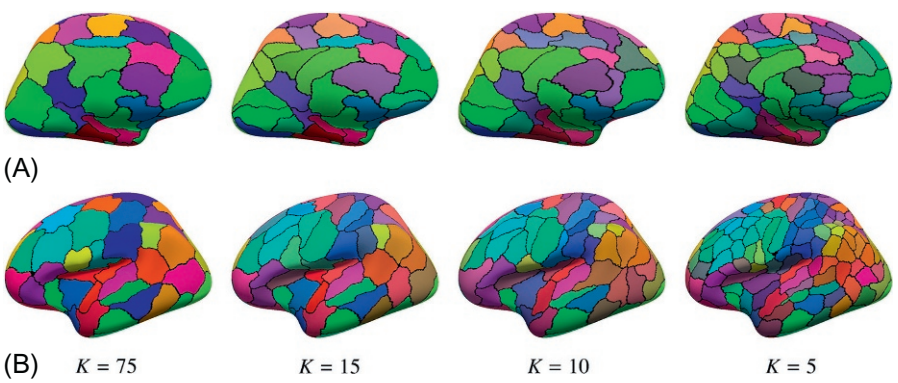

FIG. 1.3

Cortical labels based on functional connectivity from 859 fMRI scans modeled with a MRF model (Honnorat et al., 2015). Increasing number of labels for decreasing label cost K (Eq. 1.44). (A) Medial surface of the left hemisphere. (B) Lateral surface of the left hemisphere.

Source: Reprinted from Honnorat, N., Eavani, H., Satterthwaite, T., Gur, R., Gur, R., Davatzikos, C., 2015. GraSP: geodesic graph-based segmentation with shape priors for the functional parcellation of the cortex. NeuroImage 106, 207—221, with permission from Elsevier.

Most brain parcellation approaches assume that there is a single connectivity dataset (van den Heuvel et al., 2008; Yeo et al., 2011; Craddock et al., 2012) obtained by averaging connectivity over many subjects. However, it is known that there exists a large amount of variability betweens subjects (Amunts et al., 1999; Fischl et al., 2008). This variability can be modeled using an MRF approach. Liu et al. (2014) introduce a set of graphs $\{G_j = (V_j, E_j)\}$, one for each subject $j \in \{1, \ldots, J\}$, as well as one group graph $G_G = (V_G, E_G)$. Then the Potts model $\Phi_{\text{neigh}}(l_n, l_m)$ from Eq. (1.39) was used to specify a hierarchical MRF prior (Liu et al., 2014):

$$E_V(l_{1:N}) = \sum_{s,r \in V_G} \beta \Phi_{\text{neigh}}(l_s, l_r) \tag{1.46}$$

$$+ \sum_{j=1}^{J} \left(\sum_{s \in V_G, \hat{s} \in V_j} \alpha \Phi_{\text{neigh}}(l_s, l_{\hat{s}}) + \sum_{s,r \in V_j} \beta \Phi_{\text{neigh}}(l_s, l_r) \right), \tag{1.47}$$

where $\sum_{s,r \in V_G} \beta \Phi_{\text{neigh}}(l_s, l_r)$ and $\sum_{s,r \in V_j} \beta \Phi_{\text{neigh}}(l_s, l_r)$ enforce piecewise continuous labels on the group and subject level, respectively. $\alpha \Phi_{\text{neigh}}(l_s, l_{\hat{s}})$ penalizes different labels between parcellations at the subject and group levels. The parameters α and β control the strength of the respective potentials.

1.4.2 INFERENCE

Given our model and observed connectivity data $Y_{1:N}$, we want to perform inference of the unknown labels $l_{1:N}$. The complexity of the inference depends on the graph structure (Fig. 1.2), more precisely on the presence of loops in the graph. While exact inference in tree-like graphs is feasible, for general graphs we can often only approximate the inference (Bishop, 2006; Koller and Friedman, 2009; Barber, 2012).

To give an intuition of why exact inference is not possible in practice, the search space is huge as there are L^N possible solutions. Because of the interaction at each brain location x_n with its neighborhood the inference cannot be factorized.

Approximate inference can be categorized into two forms. The first form are stochastic approaches (Jbabdi et al., 2009) that, given an infinite amount of computational resources, produce exact results. The approximation arises from the natural limit of computational resources (Bishop, 2006), which can be problematic even for small instances. The second form are deterministic approximation schemes (Woolrich and Behrens, 2006; Tucholka et al., 2008; Ryali et al., 2013), which scale well, even on larger instances. Popular approximation schemes include graph cuts (Boykov et al., 2001; Delong et al., 2010), linear programming (Komodakis and Tziritas, 2007; Komodakis et al., 2011), and variational approaches, such as variational EM (Wainwright and Jordan, 2007). Here, we will give an example of deterministic approximation using variational EM. Variational EM analytically approximates the posterior probabilities by additional independence assumptions. As such they can never generate exact results even with infinite computational resources.

Given observed data $Y_{1:N}$, hidden labels $l_{1:N}$ and model parameters Θ, we aim to maximize the log likelihood. The log likelihood can be written as the marginal over the observed data $Y_{1:N}$ in terms of the sum over the joint distribution of hidden labels $l_{1:N}$ and observed data $Y_{1:N}$:

$$\log p(Y_{1:N}|\Theta) = \log \sum_{l_{1:N}} p(Y_{1:N}, l_{1:N}|\Theta). \tag{1.48}$$

Using Jensen's inequality, we can define a lower bound on the log likelihood:

$$\log p(Y_{1:N}|\Theta) = \log \sum_{l_{1:N}} \frac{p(Y_{1:N}, l_{1:N}|\Theta) q(l_{1:N})}{q(l_{1:N})} \tag{1.49}$$

$$\geq \sum_{l_{1:N}} q(l_{1:N}) \log \frac{p(Y_{1:N}, l_{1:N}|\Theta)}{q(l_{1:N})} \tag{1.50}$$

$$= \mathcal{L}(q, \Theta). \tag{1.51}$$

The difference between the log likelihood $\log p(Y_{1:N}|\Theta)$ and the lower bound $\mathcal{L}(q, \Theta)$ can be expressed by the Kullback-Leibler (KL) divergence:

$$\log p(Y_{1:N}|\Theta) - \mathcal{L}(q, \Theta) \tag{1.52}$$

$$= \log p(Y_{1:N}|\Theta) - \sum_{l_{1:N}} q(l_{1:N}) \log \frac{p(Y_{1:N}, l_{1:N}|\Theta)}{q(l_{1:N})} \tag{1.53}$$

$$= \log p(Y_{1:N}|\Theta) - \sum_{l_{1:N}} q(l_{1:N}) \log \frac{p(l_{1:N}|Y_{1:N}, \Theta) p(Y_{1:N}|\Theta)}{q(l_{1:N})} \tag{1.54}$$

$$= -\sum_{l_{1:N}} q(l_{1:N}) \log \frac{p(l_{1:N}|Y_{1:N}, \Theta)}{q(l_{1:N})} \tag{1.55}$$

$$= \mathrm{KL}(q(l_{1:N}) || p(Y_{1:N}|l_{1:N}, \Theta)). \tag{1.56}$$

As a consequence, we can decompose the log likelihood as

$$\log p(Y_{1:N}|\Theta) = \mathcal{L}(q, \Theta) + \mathrm{KL}(q(l_{1:N}) || p(Y_{1:N}|l_{1:N}, \Theta)). \tag{1.57}$$

As the KL divergence $\mathrm{KL}(q(l_{1:N}) || p(Y_{1:N}|l_{1:N}, \Theta))$ is always non-negative, instead of directly maximizing $\log p(Y_{1:N}|\Theta)$, we can maximize its lower bound $\mathcal{L}(q, \Theta)$. For our model with an MRF prior, a common approximation (Bishop, 2006) is the mean-field approximation, where $q(l_{1:N})$ are assumed to be factorizable:

$$q(l_{1:N}) = \prod_{n=1}^{N} q_n(l_n), \tag{1.58}$$

where each l_n is a single variable in $l_{1:N}$. To maximize the lower bound $\mathcal{L}(q, \Theta)$, in the E-step we fix the model parameters Θ and make use of the factorization in Eq. (1.58) to optimize $q(l_{1:N})$. Therefore the E-step is equivalent to minimizing the

KL divergence $\mathrm{KL}(q(l_{1:N})||p(Y_{1:N}|l_{1:N},\Theta))$, which aims to find the $q(l_{1:N})$ closest to the exact posterior $p(l_{1:N}|Y_{1:N},\Theta)$:

$$q^{t+1}(l_{1:N}) = \arg\max_{q(l_{1:N})} \mathcal{L}(q,\Theta^t) = \arg\min_{q(l_{1:N})} \mathrm{KL}(q(l_{1:N})||p(Y_{1:N}|l_{1:N},\Theta^t)). \tag{1.59}$$

In the M-step, we optimize the model parameters Θ to maximize the lower bound $\mathcal{L}(q,\Theta)$ using $q_n(l_n)$ updated from the E-step:

$$\Theta^{t+1} = \arg\max_{\Theta} \mathcal{L}(q^{t+1},\Theta). \tag{1.60}$$

Here, we provide further detailed derivations in the context of a von Mises-Fisher MRF model (Fig. 1.2, right). The MRF prior corresponds to the example we have given in Eq. (1.39). Conditioned on the segmentation label l_n, we assume the connectivity features Y_n are generated from the von Mises-Fisher distribution with parameters $\theta_{l_n} = \{\mu_{l_n},\kappa\}$, and therefore

$$\Phi_{\mathrm{obs}}(l_n) = -\log p(Y_n|\theta_{l_n}) = -\log\left(z_D(\kappa)\mathrm{e}^{\kappa\mu_{l_n}^{\mathrm{T}}Y_n}\right). \tag{1.61}$$

Therefore the energy function over $l_{1:N}$ is

$$E(l_{1:N}) = \sum_{n=1}^{N}\Phi_{\mathrm{obs}}(l_n) + \sum_{n=1}^{N}\sum_{j\in\mathcal{N}_n}\Phi_{\mathrm{neigh}}(l_n,l_j) \tag{1.62}$$

$$= -\sum_{n=1}^{N}\log p(Y_n|\theta_{l_n}) + \sum_{n=1}^{N}\sum_{j\in\mathcal{N}_n}\Phi_{\mathrm{neigh}}(l_n,l_j). \tag{1.63}$$

Then the joint probability of hidden labels and observed data is

$$p(Y_{1:N},l_{1:N}|\Theta) = \frac{1}{\mathcal{Z}}\mathrm{e}^{-E(l_{1:N})} \tag{1.64}$$

$$= \frac{1}{\mathcal{Z}}\mathrm{e}^{\log\sum_{n=1}^{N}p(Y_n|\theta_{l_n}) - \sum_{n=1}^{N}\sum_{j\in\mathcal{N}_n}\Phi_{\mathrm{neigh}}(l_n,l_j)}, \tag{1.65}$$

where model parameters $\Theta = \{\mu_{1:L},\kappa\}$, and Φ_{neigh} are defined as in Eq. (1.39). Given the observed data $Y_{1:N}$, hidden labels $l_{1:N}$ and model parameters $\Theta = \{\mu_{1:L},\kappa\}$, we maximize the log likelihood:

$$\arg\max_{\Theta}\log p(Y_{1:N}|\Theta) \tag{1.66}$$

$$= \arg\max_{\Theta}\log\sum_{l_{1:N}} p(Y_{1:N},l_{1:N}|\Theta) \tag{1.67}$$

$$= \arg\max_{\Theta}\log\sum_{l_{1:N}}\frac{1}{\mathcal{Z}}\mathrm{e}^{\sum_{n=1}^{N}p(Y_n|\theta_{l_n}) - \sum_{n=1}^{N}\sum_{j\in\mathcal{N}_n}\Phi_{\mathrm{neigh}}(l_n,l_j)}. \tag{1.68}$$

Optimizing Eq. (1.68) is computationally intractable. We use variational EM to iteratively estimate Eq. (1.68), where we alternate between estimating the parameters from the von Mises-Fisher distributions $\Theta = \{\mu_{1:L}, \kappa\}$ and the hidden labels $l_{1:N}$.

To maximize Eq. (1.68), we maximize the lower bound:

$$\mathcal{L}(q, \Theta) = \sum_{l_{1:N}} q(l_{1:N}) \log \frac{p(Y_{1:N}, l_{1:N}|\Theta)}{q(l_{1:N})} \tag{1.69}$$

$$= \sum_{l_{1:N}} q(l_{1:N}) \log p(Y_{1:N}, l_{1:N}|\Theta) - \sum_{l_{1:N}} q(l_{1:N}) \log q(l_{1:N}), \tag{1.70}$$

where

$$\log p(Y_{1:N}, l_{1:N}|\Theta) = -\log \mathcal{Z} + \log \sum_{n=1}^{N} p(Y_n|\theta_{l_n}) - \sum_{n=1}^{N} \sum_{j \in \mathcal{N}_n} \Phi_{\text{neigh}}(l_n, l_j). \tag{1.71}$$

After expanding the terms, the lower bound $\mathcal{L}(q, \Theta)$ can be expressed as

$$\mathcal{L}(q, \Theta) = \sum_{n=1}^{N} \sum_{l_n=1}^{L} q_n(l_n) p(Y_n|\theta_{l_n}) - \sum_{n=1}^{N} \sum_{l_n=1}^{L} \sum_{j \in \mathcal{N}_n} \sum_{l_j=1}^{L} q_n(l_n) q_j(l_j) \Phi_{\text{neigh}}(l_n, l_j)$$
$$- \sum_{n=1}^{N} \sum_{l_n=1}^{L} q_n(l_n) \log q_n(l_n). \tag{1.72}$$

To simplify the notation in Eq. (1.72), let $\lambda_{n,l_n} = q_n(l_n)$, so the lower bound can be written as

$$\mathcal{L}(\lambda, \Theta) = \sum_{n=1}^{N} \sum_{l_n=1}^{L} \lambda_{n,l_n} \log p(Y_n|\theta_{l_n}) - \sum_{n=1}^{N} \sum_{l_n=1}^{L} \sum_{j \in \mathcal{N}_n} \sum_{l_j=1}^{L} \lambda_{n,l_n} \lambda_{j,l_j} \Phi_{\text{neigh}}(l_n, l_j)$$
$$- \sum_{n=1}^{N} \sum_{l_n=1}^{L} \lambda_{n,l_n} \log \lambda_{n,l_n}. \tag{1.73}$$

Variational EM proceeds as follows. We begin with the variational E-step where we estimate λ_{n,l_n} to maximize the lower bound $\mathcal{L}$. We add a Lagrange multiplier η_n to ensure that $\sum_{l_n} \lambda_{n,l_n} = 1$ for all n. We denote the current estimate of our parameters as Θ^t. The lower bound can be written as

$$\mathcal{L}(\lambda, \Theta^t) = \sum_{n=1}^{N} \sum_{l_n=1}^{L} \lambda_{n,l_n} \log p(Y_n|\theta_{l_n}^t) - \sum_{n=1}^{N} \sum_{l_n=1}^{L} \sum_{j \in N_n} \sum_{l_j=1}^{L} \lambda_{n,l_n} \lambda_{j,l_j} \Phi_{\text{neigh}}(l_n, l_j)$$
$$- \sum_{n=1}^{N} \sum_{l_n=1}^{L} \lambda_{n,l_n} \log \lambda_{n,l_n} + \sum_{n=1}^{N} \eta_n \left(\sum_{l_n=1}^{L} \lambda_{n,l_n} - 1 \right). \tag{1.74}$$

We will estimate $\lambda_{m,k}$ for each brain region $m \in \{1, \ldots, N\}$ and label $k \in \{1, \ldots, L\}$ by differentiating Eq. (1.74) and setting the derivative to 0:

$$\frac{\partial \mathcal{L}(\lambda, \Theta^t)}{\partial \lambda_{m,k}} = \log p(Y_m|\theta_k^t) - 2 \sum_{j \in N_m} \sum_{l_j=1}^{L} \lambda_{j,l_j} \Phi_{\text{neigh}}(l_m = k, l_j) - 1 - \log \lambda_{m,k} + \eta_m = 0. \tag{1.75}$$

Note the factor of 2 in the second term, which arises because we assume that if x_n is a neighbor of x_m, then x_m is also a neighbor of x_n and because of the symmetric nature of the Potts model. Rearranging the above equation, we get

$$\log \lambda_{m,k}^{t+1} \propto \log p(Y_m|\theta_k^t) - 2 \sum_{j \in N_m} \sum_{l_j=1}^{L} \lambda_{j,l_j} \Phi_{\text{neigh}}(l_m = k, l_j), \tag{1.76}$$

$$\lambda_{m,k}^{t+1} \propto p(Y_m|\mu_k^t, \kappa^t) \mathrm{e}^{-2 \sum_{j \in N_m} \sum_{l_j=1}^{L} \lambda_{j,l_j} \Phi_{\text{neigh}}(l_m = k, l_j)}. \tag{1.77}$$

Since the update for $\lambda_{m,k}^{t+1}$ depends on λ_{j,l_j}, we estimate $\lambda_{m,k}^{t+1}$ via fixed point iterations using Eq. (1.77), and normalizing λs in each iteration so that $\sum_{l_n} \lambda_{n,l_n} = 1$ for all n.

In the variational M-step, we compute the parameters Θ^{t+1} with the lower bound (Eq. 1.73) based on the current estimate $\lambda_{1:N,1:L}^{t+1}$. By dropping the terms that do not contain the parameters Θ, we get

$$\Theta^{t+1} = \arg\max_{\Theta} \mathcal{L}(\lambda^{t+1}, \Theta) = \arg\max_{\mu_{1:L}, \kappa} \sum_{n=1}^{N} \sum_{l_n} \lambda_{n,l_n}^{t+1} \log p(Y_n|\theta_{l_n}). \tag{1.78}$$

Using the Lagrange multiplier β_l for constraints $\mu_l^{\mathrm{T}} \mu_l = 1$, we can write the above lower bound as separate optimizations over $\mu_{1:L}$ and κ:

$$\mathcal{L}_\mu = \sum_{n=1}^{N} \sum_{l=1}^{L} \lambda_{n,l}^{t+1} (\kappa \mu_l^{\mathrm{T}} Y_n) + \sum_{l=1}^{L} \beta_l (1 - \mu_l^{\mathrm{T}} \mu_l), \tag{1.79}$$

$$\mathcal{L}_\kappa = \sum_{n=1}^{N} \sum_{l=1}^{L} \lambda_{n,l}^{t+1} \log z_D(\kappa) + \sum_{n=1}^{N} \sum_{l=1}^{L} \lambda_{n,l}^{t+1} (\kappa \mu_l^{\mathrm{T}} Y_n). \tag{1.80}$$

To compute μ_l^{t+1} we take the derivative of $\mathcal{L}_\mu$:

$$\frac{\partial \mathcal{L}_\mu}{\partial \mu_l} = \sum_{n=1}^{N} \lambda_{n,l}^{t+1} \kappa Y_n - 2\beta_l \mu_l = 0, \tag{1.81}$$

$$\mu_l^{t+1} = \frac{\sum_{n=1}^{N} \lambda_{n,l}^{t+1} \kappa Y_n}{2\beta_l}. \tag{1.82}$$

Just like in the case of the mixture model (Section 1.3), the Lagrange multiplier β_l is determined by the fact that μ_l should be unit norm, and so we get

$$\mu_l^{t+1} = \frac{\sum_{n=1}^{N} \lambda_{n,l}^{t+1} Y_n}{\| \sum_{n=1}^{N} \lambda_{n,l}^{t+1} Y_n \|}. \tag{1.83}$$

To estimate κ^{t+1}, we set the derivative of $\mathcal{L}_\kappa$ to 0:

$$\frac{\partial \mathcal{L}_\kappa}{\partial \kappa} = \sum_{n=1}^{N} \sum_{l=1}^{L} \lambda_{n,l}^{t+1} \frac{z_D'(\kappa)}{z_D(\kappa)} + \sum_{n=1}^{N} \sum_{l=1}^{L} \lambda_{n,l}^{t+1} \mu_l^{\mathrm{T}} Y_n \tag{1.84}$$

$$= N \frac{z_D'(\kappa)}{z_D(\kappa)} + \sum_{n=1}^{N} \sum_{l=1}^{L} \lambda_{n,l}^{t+1} \mu_l^{\mathrm{T}} Y_n = 0. \tag{1.85}$$

We can again use the approximation given by Lashkari et al. (2010):

$$\kappa^{t+1} \approx \frac{(D-2)\Gamma}{1-\Gamma^2} + \frac{(D-1)\Gamma}{2(D-2)}, \tag{1.86}$$

where $\Gamma = \frac{1}{N} \sum_{n=1}^{N} \delta(l_n, l) Y_n^{\mathrm{T}} \mu_l$ and $\delta(l_n, l) = 1, \exists n : x_n = l;\ 0$ otherwise.

We iterate between estimating the parameters $\Theta = \{\mu_{1:L}, \kappa\}$ (using Eqs. 1.83 and 1.86), and λ_{n,l_n} for each location x_n and label l_n (using Eq. 1.78). To generate a final parcellation, for each brain location n, the label l_n with highest λ_{n,l_n} can be chosen.

1.5 SUMMARY

Human brain parcellation is one of the major challenges in systems neuroscience and key for understanding complex human behavior. Machine learning has been and will continue to be a central element in deriving human brain parcellations as the underlying datasets become larger and more diverse. Here, we have focused on mixture and MRFs, which can be easily combined and extended to match a wide range of applications. These models can, for example, be used to create personalized brain parcellations while using population priors to increase stability (Jbabdi et al., 2009; Liu et al., 2011, 2014; Harrison et al., 2015). The resulting single subject parcellations can address the strong intersubject variability in brain organization and therefore improve sensitivity in clinical applications.

REFERENCES

Abraham, A., Dohmatob, E., Thirion, B., Samaras, D., Varoquaux, G., 2014. Region segmentation for sparse decompositions: better brain parcellations from rest fMRI, pp. 1–8, arXiv 1412.3925.

Amunts, K., Schleicher, A., Buergel, U., Mohlberg, H., Uylings, H.B., Zilles, K., 1999. Broca's region revisited: cytoarchitecture and intersubject variability. J. Comp. Neurol. 412 (2), 319–341. http://dx.doi.org/10.1002/(SICI)1096-9861(19990920)412:2<319::AID-CNE10>3.0.CO;2-7.

Banerjee, A., Dhillon, I.S., Ghosh, J., Sra, S., Ridgeway, G., 2005. Clustering on the unit hypersphere using von Mises-Fisher distributions. J. Mach. Learn. Res. 6 (9), 1345–1382.

Barber, D., 2012. Bayesian Reasoning and Machine Learning. Cambridge University Press, Cambridge, UK.

Beckmann, C.F., Smith, S.M., 2004. Probabilistic independent component analysis for functional magnetic resonance imaging. IEEE Trans. Med. Imaging 23 (2), 137–152. http://dx.doi.org/10.1109/TMI.2003.822821.

Bell, A.J., Sejnowski, T.J., 1995. An information-maximization approach to blind separation and blind deconvolution. Neural Comput. 7 (6), 1129–1159.

Bellec, P., Rosa-Neto, P., Lyttelton, O.C., Benali, H., Evans, A.C., 2010. Multi-level bootstrap analysis of stable clusters in resting-state fMRI. NeuroImage 51 (3), 1126–1139. http://dx.doi.org/10.1016/j.neuroimage.2010.02.082.

Belliveau, J.W., Kennedy, D.N., McKinstry, R.C., Buchbinder, B.R., Weisskoff, R.M., Cohen, M.S., Vevea, J.M., Brady, T.J., Rosen, B.R., 1991. Functional mapping of the human visual cortex by magnetic resonance imaging. Science 254 (5032), 716–719.

Besag, J., 1974. Spatial interaction and the statistical analysis of lattice systems. J. R. Stat. Soc. B 36 (2), 192–236. http://dx.doi.org/10.2307/2984812.

Beucher, S., Lantuejoul, C., 1979. Use of watersheds in contour detection. In: International Workshop on Image Processing: Real-Time Edge and Motion Detection/Estimation, Rennes, France.

Bishop, C.M., 2006. Pattern Recognition and Machine Learning. Springer, Berlin.

Biswal, B., Yetkin, F.Z., Haughton, V.M., Hyde, J.S., 1995. Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. Magn. Reson. Med. 34 (4), 537–541.

Biswal, B.B., Mennes, M., Zuo, X.N., Gohel, S., Kelly, C., Smith, S.M., Beckmann, C.F., Adelstein, J.S., Buckner, R.L., Colcombe, S., Dogonowski, A.M., Ernst, M., Fair, D., Hampson, M., Hoptman, M.J., Hyde, J.S., Kiviniemi, V.J., Kötter, R., Li, S.J., Lin, C.P., Lowe, M.J., Mackay, C., Madden, D.J., Madsen, K.H., Margulies, D.S., Mayberg, H.S., McMahon, K., Monk, C.S., Mostofsky, S.H., Nagel, B.J., Pekar, J.J., Peltier, S.J., Petersen, S.E., Riedl, V., Rombouts, S.A.R.B., Rypma, B., Schlaggar, B.L., Schmidt, S., Seidler, R.D., Siegle, G.J., Sorg, C., Teng, G.J., Veijola, J., Villringer, A., Walter, M., Wang, L., Weng, X.C., Whitfield-Gabrieli, S., Williamson, P., Windischberger, C., Zang, Y.F., Zhang, H.Y., Castellanos, F.X., Milham, M.P., 2010. Toward discovery science of human brain function. Proc. Natl. Acad. Sci. USA 107 (10), 4734–4739. http://dx.doi.org/10.1073/pnas.0911855107.

Blumensath, T., Jbabdi, S., Glasser, M.F., Van Essen, D.C., Ugurbil, K., Behrens, T.E., Smith, S.M., 2013. Spatially constrained hierarchical parcellation of the brain with resting-state fMRI. NeuroImage 76, 313–324. http://dx.doi.org/10.1016/j.neuroimage.2013.03.024.

Boykov, Y., Veksler, O., Zabih, R., 2001. Fast approximate energy minimization via graph cuts. IEEE Trans. Pattern Anal. Mach. Intell. 23 (11), 1222–1239. http://dx.doi.org/10.1109/34.969114.

Brodmann, K., 1909. Vergleichende Lokalisationslehre der Grosshirnrinde: in ihren Prinzipien dargestellt auf Grund des Zellenbaues. JA Barth, Leipzig.

Calhoun, V., Adali, T., Pearlson, G., Pekar, J., 2001. A method for making group inferences from functional MRI data using independent component analysis. Hum. Brain Map. 14 (3), 140–151. http://dx.doi.org/10.1002/hbm.1048.

Canny, J., 1986. A computational approach to edge detection. IEEE Trans. Pattern Anal. Mach. Intell. 8 (6), 679–698. http://dx.doi.org/10.1109/TPAMI.1986.4767851.

Cauda, F., D'Agata, F., Sacco, K., Duca, S., Geminiani, G., Vercelli, A., 2011. Functional connectivity of the insula in the resting brain. NeuroImage 55 (1), 8–23. http://dx.doi.org/10.1016/j.neuroimage.2010.11.049.

Cohen, A.L., Fair, D.A., Dosenbach, N.U.F., Miezin, F.M., Dierker, D., Van Essen, D.C., Schlaggar, B.L., Petersen, S.E., 2008. Defining functional areas in individual human brains using resting functional connectivity MRI. NeuroImage 41 (1), 45–57. http://dx.doi.org/10.1016/j.neuroimage.2008.01.066.

Collins, P.Y., Patel, V., Joestl, S.S., March, D., Insel, T.R., Daar, A.S., Bordin, I.A., Costello, E.J., Durkin, M., Fairburn, C., Glass, R.I., Hall, W., Huang, Y., Hyman, S.E., Jamison, K., Kaaya, S., Kapur, S., Kleinman, A., Ogunniyi, A., Otero-Ojeda, A., Poo, M.M., Ravindranath, V., Sahakian, B.J., Saxena, S., Singer, P.A., Stein, D.J., Anderson, W., Dhansay, M.A., Ewart, W., Phillips, A., Shurin, S., Walport, M., 2011. Grand challenges in global mental health. Nature 475 (7354), 27–30. http://dx.doi.org/10.1038/475027a.

Craddock, R.C., James, G.A., Holtzheimer, P.E., Hu, X.P., Mayberg, H.S., 2012. A whole brain fMRI atlas generated via spatially constrained spectral clustering. Hum. Brain Map. 33 (8), 1914–1928. http://dx.doi.org/10.1002/hbm.21333.

Craddock, R.C., Jbabdi, S., Yan, C.g., Vogelstein, J.T., Castellanos, F.X., Di Martino, A., Kelly, C., Heberlein, K., Colcombe, S., Milham, M.P., 2013. Imaging human connectomes at the macroscale. Nat. Meth. 10 (6), 524–539. http://dx.doi.org/10.1038/nmeth.2482.

Delong, A., Osokin, A., Isack, H.N., Boykov, Y., 2010. Fast approximate energy minimization with label costs. In: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE, Piscataway, NJ, pp. 2173–2180.

Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood from incomplete data via the EM algorithm. J. R. Stat. Soc. B 39 (1), 1–38.

Eickhoff, S.B., Laird, A.R., Grefkes, C., Wang, L.E., Zilles, K., Fox, P.T., 2009. Coordinate-based activation likelihood estimation meta-analysis of neuroimaging data: a random-effects approach based on empirical estimates of spatial uncertainty. Hum. Brain Map. 30 (9), 2907–2926. http://dx.doi.org/10.1002/hbm.20718.

Eickhoff, S.B., Bzdok, D., Laird, A.R., Roski, C., Caspers, S., Zilles, K., Fox, P.T., United States, 2011. Co-activation patterns distinguish cortical modules, their connectivity and functional differentiation. NeuroImage 57 (3), 938–949. http://dx.doi.org/10.1016/j.neuroimage.2011.05.021.

Eickhoff, S.B., Thirion, B., Varoquaux, G., Bzdok, D., 2015. Connectivity-based parcellation: critique and implications. Hum. Brain Map. 36, 4771–4792. http://dx.doi.org/10.1002/hbm.22933.

Felleman, D.J., Van Essen, D.C., 1991. Distributed hierarchical processing in the primate cerebral cortex. Cerebral Cortex (NY) 1 (1), 1–47. http://dx.doi.org/10.1093/cercor/1.1.1.

Fischl, B., Rajendran, N., Busa, E., Augustinack, J., Hinds, O., Yeo, B.T.T., Mohlberg, H., Amunts, K., Zilles, K., 2008. Cortical folding patterns and predicting cytoarchitecture. Cerebral Cortex 18 (8), 1973–1980. http://dx.doi.org/10.1093/cercor/bhm225.

Fox, P.T., Lancaster, J.L., Laird, A.R., Eickhoff, S.B., 2014. Meta-analysis in human neuroimaging: computational modeling of large-scale databases. Ann. Rev. Neurosci. 37, 409–434. http://dx.doi.org/10.1146/annurev-neuro-062012-170320.

Friston, K.J., 1994. Functional and effective connectivity in neuroimaging: a synthesis. Hum. Brain Map. 2 (1–2), 56–78. http://dx.doi.org/10.1002/hbm.460020107.

Geman, S., Geman, D., 1984. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE Trans. Pattern Anal. Mach. Intell. 6 (6), 721–741. http://dx.doi.org/10.1109/TPAMI.1984.4767596.

Glasser, M.F., Van Essen, D.C., 2011. Mapping human cortical areas in vivo based on myelin content as revealed by T1- and T2-weighted MRI. J. Neurosci. 31 (32), 11597–11616. http://dx.doi.org/10.1523/JNEUROSCI.2180-11.2011.

Golland, P., Golland, Y., Malach, R., 2007. Detection of spatial activation patterns as unsupervised segmentation of fMRI data. In: Medical Image Computing and Computer-Assisted Intervention—MICCAI 2007. Springer, Berlin, pp. 110–118.

Gordon, E.M., Laumann, T.O., Adeyemo, B., Huckins, J.F., Kelley, W.M., Petersen, S.E., 2014. Generation and evaluation of a cortical area parcellation from resting-state correlations. Cerebral Cortex 26 (1), 288–303.

Gorgolewski, K.J., Varoquaux, G., Rivera, G., Schwarz, Y., Ghosh, S.S., Maumet, C., Sochat, V.V., Nichols, T.E., Poldrack, R.A., Poline, J.B., Yarkoni, T., Margulies, D.S., 2015. NeuroVault.org: a web-based repository for collecting and sharing unthresholded statistical maps of the human brain. Front. Neuroinformat. 9. http://dx.doi.org/10.3389/fninf.2015.00008.

Hagmann, P., Kurant, M., Gigandet, X., Thiran, P., Wedeen, V.J., Meuli, R., Thiran, J.P., 2007. Mapping human whole-brain structural networks with diffusion MRI. PLoS ONE 2 (7), e597. http://dx.doi.org/10.1371/journal.pone.0000597.

Harrison, S.J., Woolrich, M.W., Robinson, E.C., Glasser, M.F., Beckmann, C.F., Jenkinson, M., Smith, S.M., 2015. Large-scale probabilistic functional modes from resting state fMRI. NeuroImage 109, 217–231. http://dx.doi.org/10.1016/j.neuroimage.2015.01.013.

He, Y., Wang, J., Wang, L., Chen, Z.J., Yan, C., Yang, H., Tang, H., Zhu, C., Gong, Q., Zang, Y., Evans, A.C., 2009. Uncovering intrinsic modular organization of spontaneous brain activity in humans. PLoS ONE 4 (4), e5226. http://dx.doi.org/10.1371/journal.pone.0005226.

Hirose, S., Watanabe, T., Jimura, K., Katsura, M., Kunimatsu, A., Abe, O., Ohtomo, K., Miyashita, Y., Konishi, S., 2012. Local signal time-series during rest used for areal boundary mapping in individual human brains. PLoS ONE 7 (5), e36496. http://dx.doi.org/10.1371/journal.pone.0036496.

Holmes, A.J., Hollinshead, M.O., O'Keefe, T.M., Petrov, V.I., Fariello, G.R., Wald, L.L., Fischl, B., Rosen, B.R., Mair, R.W., Roffman, J.L., Smoller, J.W., Buckner, R.L., 2015. Brain Genomics Superstruct Project initial data release with structural, functional, and behavioral measures. Sci. Data 2, 150031. http://dx.doi.org/10.1038/sdata.2015.31.

Honnorat, N., Eavani, H., Satterthwaite, T.D., Davatzikos, C., 2013. A graph-based brain parcellation method extracting sparse networks. In: 2013 International Workshop on Pattern Recognition in Neuroimaging. IEEE, Piscataway, NJ, pp. 157–160.

Honnorat, N., Eavani, H., Satterthwaite, T., Gur, R., Gur, R., Davatzikos, C., 2015. GraSP: Geodesic graph-based segmentation with shape priors for the functional parcellation of the cortex. NeuroImage 106, 207–221. http://dx.doi.org/10.1016/j.neuroimage.2014.11.008.

Insel, T.R., 2009. Translating scientific opportunity into public health impact. Arch. Gen. Psychiat. 66 (2), 128. http://dx.doi.org/10.1001/archgenpsychiatry.2008.540.

Jain, A.K., 2010. Data clustering: 50 years beyond K-means. Pattern Recogn. Lett. 31 (8), 651–666. http://dx.doi.org/10.1016/j.patrec.2009.09.011. 0402594v3.

Jbabdi, S., Woolrich, M.W., Behrens, T.E.J., 2009. Multiple-subjects connectivity-based parcellation using hierarchical Dirichlet process mixture models. NeuroImage 44 (2), 373–384. http://dx.doi.org/10.1016/j.neuroimage.2008.08.044.

Jianbo Shi, Malik, J., 2000. Normalized cuts and image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 22 (8), 888–905. http://dx.doi.org/10.1109/34.868688.

Johansen-Berg, H., Behrens, T.E.J. (Eds.), 2013. Diffusion MRI: From Quantitative Measurement to In Vivo Neuroanatomy. Academic Press, London.

Johansen-Berg, H., Behrens, T.E.J., Robson, M.D., Drobnjak, I., Rushworth, M.F.S., Brady, J.M., Smith, S.M., Higham, D.J., Matthews, P.M., 2004. Changes in connectivity profiles define functionally distinct regions in human medial frontal cortex. Proc. Natl. Acad. Sci. USA 101 (36), 13335–13340. http://dx.doi.org/10.1073/pnas.0403743101.

Kaas, J., 1987. The organization of neocortex in mammals: implications for theories of brain function. Ann. Rev. Psychol. 38 (1), 129–151. http://dx.doi.org/10.1146/annurev.psych.38.1.129.

Kim, J.H., Lee, J.M., Jo, H.J., Kim, S.H., Lee, J.H., Kim, S.T., Seo, S.W., Cox, R.W., Na, D.L., Kim, S.I., Saad, Z.S., 2010. Defining functional SMA and pre-SMA subregions in human MFC using resting state fMRI: functional connectivity-based parcellation method. NeuroImage 49 (3), 2375–2386. http://dx.doi.org/10.1016/j.neuroimage.2009.10.016.

Koller, D., Friedman, N., 2009. Probabilistic Graphical Models: Principles and Techniques. MIT Press, Cambridge, MA.

Komodakis, N., Tziritas, G., 2007. Approximate labeling via graph cuts based on linear programming. IEEE Trans. Pattern Anal. Mach. Intell. 29 (8), 1436–1453. http://dx.doi.org/10.1109/TPAMI.2007.1061.

Komodakis, N., Paragios, N., Tziritas, G., 2011. MRF energy minimization and beyond via dual decomposition. IEEE Trans. Pattern Anal. Mach. Intell. 33 (3), 531–552. http://dx.doi.org/10.1109/TPAMI.2010.108.

Lashkari, D., Vul, E., Kanwisher, N., Golland, P., 2010. Discovering structure in the space of fMRI selectivity profiles. NeuroImage 50 (3), 1085–1098. http://dx.doi.org/10.1016/j.neuroimage.2009.12.106.

Liu, W., Awate, S.P., Anderson, J.S., Yurgelun-Todd, D., Fletcher, P.T., 2011. Monte Carlo expectation maximization with hidden Markov models to detect functional networks in resting-state fMRI. In: MICCAI Workshop on Machine Learning in Medical Imaging. Springer, New York, pp. 59–66.

Liu, W., Awate, S.P., Fletcher, P.T., 2012. Group analysis of resting-state fMRI by hierarchical Markov random fields. Med. Image Comput. Comput. Assis. Interven. 15 (3), 189–196. http://dx.doi.org/10.1126/science.3749875.

Liu, W., Awate, S.P., Anderson, J.S., Fletcher, P.T., 2014. A functional network estimation method of resting-state fMRI using a hierarchical Markov random field. NeuroImage 100, 520–534. http://dx.doi.org/10.1016/j.neuroimage.2014.06.001.

Lloyd, S.P., 1982. Least squares quantization in PCM. IEEE Trans. Inform. Theory 28 (2), 129–137.

Lodygensky, G.A., Marques, J.P., Maddage, R., Perroud, E., Sizonenko, S.V., Hüppi, P.S., Gruetter, R., 2012. In vivo assessment of myelination by phase imaging at high magnetic field. NeuroImage 59 (3), 1979–1987. http://dx.doi.org/10.1016/j.neuroimage.2011.09.057.

Mackay, A., Whittall, K., Adler, J., Li, D., Paty, D., Graeb, D., 1994. In vivo visualization of myelin water in brain by magnetic resonance. Magn. Reson. Med. 31 (6), 673–677. http://dx.doi.org/10.1002/mrm.1910310614.

Mars, R.B., Sallet, J., Schuffelgen, U., Jbabdi, S., Toni, I., Rushworth, M.F.S., 2012. Connectivity-based subdivisions of the human right "temporoparietal junction area": evidence for different areas participating in different cortical networks. Cerebral Cortex 22 (8), 1894–1903. http://dx.doi.org/10.1093/cercor/bhr268.

Meunier, D., Lambiotte, R., Bullmore, ET., 2010. Modular and hierarchically modular organization of brain networks. Front. Neurosci. 4, 200. http://dx.doi.org/10.3389/fnins.2010.00200.

Mezer, A., Yovel, Y., Pasternak, O., Gorfine, T., Assaf, Y., 2009. Cluster analysis of resting-state fMRI time series. NeuroImage 45 (4), 1117–1125. http://dx.doi.org/10.1016/j.neuroimage.2008.12.015.

Michel, V., Gramfort, A., Varoquaux, G., Eger, E., Keribin, C., Thirion, B., 2012. A supervised clustering approach for fMRI-based inference of brain states. Pattern Recogn. 45 (6), 2041–2049. http://dx.doi.org/10.1016/j.patcog.2011.04.006.

Moreno-Dominguez, D., Anwander, A., Knösche, T.R., 2014. A hierarchical method for whole-brain connectivity-based parcellation. Hum. Brain Map. 35 (10), 5000–5025. http://dx.doi.org/10.1002/hbm.22528.

Nelson, S.M., Cohen, A.L., Power, J.D., Wig, G.S., Miezin, F.M., Wheeler, M.E., Velanova, K., Donaldson, D.I., Phillips, J.S., Schlaggar, B.L., Petersen, S.E., 2010a, A parcellation scheme for human left lateral parietal cortex. Neuron 67 (1), 156–170. http://dx.doi.org/10.1016/j.neuron.2010.05.025.

Nelson, S.M., Dosenbach, N.U.F., Cohen, A.L., Wheeler, M.E., Schlaggar, B.L., Petersen, S.E., 2010b. Role of the anterior insula in task-level control and focal attention. Brain Struct. Funct. 214 (5–6), 669–680. http://dx.doi.org/10.1007/s00429-010-0260-2.

Newman, M.E.J., 2006. Modularity and community structure in networks. Proc. Natl. Acad. Sci. USA 103 (23), 8577–8582. http://dx.doi.org/10.1073/pnas.0601602103.

Ng, A.Y., Jordan, M.I., Weiss, Y., 2001. On spectral clustering: analysis and an algorithm. Adv. Neural Inform. Process. Syst. 2, 849–856.

Nooner, K.B., Colcombe, S.J., Tobe, R.H., Mennes, M., Benedict, M.M., Moreno, A.L., Panek, L.J., Brown, S., Zavitz, S.T., Li, Q., Sikka, S., Gutman, D., Bangaru, S., Schlachter, R.T., Kamiel, S.M., Anwar, A.R., Hinz, C.M., Kaplan, M.S., Rachlin, A.B., Adelsberg, S., Cheung, B., Khanuja, R., Yan, C., Craddock, C.C., Calhoun, V., Courtney, W., King, M., Wood, D., Cox, C.L., Kelly, A.M., Di Martino, A., Petkova, E., Reiss, P.T., Duan, N., Thomsen, D., Biswal, B., Coffey, B., Hoptman, M.J., Javitt, D.C., Pomara, N., Sidtis, J.J., Koplewicz, H.S., Castellanos, F.X., Leventhal, B.L., Milham, M.P., 2012. The NKI-Rockland sample: a model for accelerating the pace of discovery science in psychiatry. Front. Neurosci. 6, 152. http://dx.doi.org/10.3389/fnins.2012.00152.

Orban, P., Doyon, J., Petrides, M., Mennes, M., Hoge, R., Bellec, P., 2015. The richness of task-evoked hemodynamic responses defines a pseudohierarchy

of functionally meaningful brain networks. Cerebral Cortex 25 (9), 2658–2669. http://dx.doi.org/10.1093/cercor/bhu064.

Potts, R.B., 1952. Some generalized order-disorder transformations. Math. Proc. Cambridge Philos. Soc., vol. 48 (01), pp. 106–109.

Preusser, S., Thiel, S.D., Rook, C., Roggenhofer, E., Kosatschek, A., Draganski, B., Blankenburg, F., Driver, J., Villringer, A., Pleger, B., 2015. The perception of touch and the ventral somatosensory pathway. Brain 138 (3), 540–548. http://dx.doi.org/10.1093/brain/awu370.

Rorden, C., Karnath, H.O., 2004. Opinion: using human brain lesions to infer function: a relic from a past era in the fMRI age? Nat. Rev. Neurosci. 5 (10), 812–819. http://dx.doi.org/10.1038/nrn1521.

Ryali, S., Chen, T., Supekar, K., Menon, V., 2013. A parcellation scheme based on von Mises-Fisher distributions and Markov random fields for segmenting brain regions using resting-state fMRI. NeuroImage 65, 83–96. http://dx.doi.org/10.1016/j.neuroimage.2012.09.067.

Sereno, M.I., Dale, A.M., Reppas, J.B., Kwong, K.K., Belliveau, J.W., Brady, T.J., Rosen, B.R., Tootell, R.B., 1995. Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. Science (New York) 268 (5212), 889–893.

Shen, X., Papademetris, X., Constable, R.T., United States, 2010. Graph-theory based parcellation of functional subunits in the brain from resting-state fMRI data. NeuroImage 50 (3), 1027–1035. http://dx.doi.org/10.1016/j.neuroimage.2009.12.119.

Smith, S.M., Miller, K.L., Salimi-Khorshidi, G., Webster, M., Beckmann, C.F., Nichols, T.E., Ramsey, J.D., Woolrich, M.W., 2011. Network modelling methods for fMRI. NeuroImage 54 (2), 875–891. http://dx.doi.org/10.1016/j.neuroimage.2010.08.063.

Sra, S., 2011. A short note on parameter approximation for von Mises-Fisher distributions and a fast implementation of $I_s(x)$. Comput. Stat. 27 (1), 177–190. http://dx.doi.org/10.1007/s00180-011-0232-x.

Swisher, J.D., Halko, M.A., Merabet, L.B., McMains, S.A., Somers, D.C., 2007. Visual topography of human intraparietal sulcus. J. Neurosci. 27 (20), 5326–5337. http://dx.doi.org/10.1523/JNEUROSCI.0991-07.2007.

Thirion, B., Flandin, G., Pinel, P., Roche, A., Ciuciu, P., Poline, J.B., 2006. Dealing with the shortcomings of spatial normalization: multi-subject parcellation of fMRI datasets. Hum. Brain Map. 27 (8), 678–693. http://dx.doi.org/10.1002/hbm.20210.

Thirion, B., Varoquaux, G., Dohmatob, E., Poline, J.B., 2014. Which fMRI clustering gives good brain parcellations? Front. Neurosci. 8 (July), 1–13. http://dx.doi.org/10.3389/fnins.2014.00167.

Tuch, D.S., Reese, T.G., Wiegell, M.R., Van J. Wedeen, 2003. Diffusion MRI of complex neural architecture. Neuron 40 (5), 885–895. http://dx.doi.org/10.1016/S0896-6273(03)00758-X.

Tucholka, A., Thirion, B., Perrot, M., Pinel, P., Mangin, J.F., Poline, J.B., 2008. Probabilistic anatomo-functional parcellation of the cortex: how many regions? In: Medical Image Computing and Computer-Assisted Intervention—MICCAI 2008. Springer, Berlin, Heidelberg, pp. 399–406.

Ungerleider, L.G., 1995. Functional brain imaging studies of cortical mechanisms for memory. Science (New York) 270 (5237), 769–775. http://dx.doi.org/10.1126/science.270.5237.769.

van den Heuvel, M., Mandl, R., Hulshoff Pol, H., 2008. Normalized cut group clustering of resting-state fMRI data. PLoS ONE 3 (4), e2001. http://dx.doi.org/10.1371/journal.pone.0002001.

Van Essen, D., Ugurbil, K., Auerbach, E., Barch, D., Behrens, T., Bucholz, R., Chang, A., Chen, L., Corbetta, M., Curtiss, S., Della Penna, S., Feinberg, D., Glasser, M., Harel, N., Heath, A., Larson-Prior, L., Marcus, D., Michalareas, G., Moeller, S., Oostenveld, R., Petersen, S., Prior, F., Schlaggar, B., Smith, S., Snyder, A., Xu, J., Yacoub, E., 2012. The Human Connectome Project: a data acquisition perspective. NeuroImage 62 (4), 2222–2231. http://dx.doi.org/10.1016/j.neuroimage.2012.02.018.

Varoquaux, G., Craddock, R.C., 2013. Learning and comparing functional connectomes across subjects. NeuroImage 80, 405–415. http://dx.doi.org/10.1016/j.neuroimage.2013.04.007. 1304.3880.

Varoquaux, G., Gramfort, A., Pedregosa, F., Michel, V., Thirion, B., 2011. Multi-subject dictionary learning to segment an atlas of brain spontaneous activity. In: Information Processing in Medical Imaging, pp. 562–573.

Veksler, O., 2008. Star shape prior for graph-cut image segmentation. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), LNCS vol. 5304 (PART 3), pp. 454–467. http://dx.doi.org/10.1007/978-3-540-88690-7-34.

Vogt, C., Vogt, O., 1919. Allgemeine Ergebnisse Unserer Hirnforschung, vol. 21. JA Barth, Leipzig.

von Luxburg, U., 2007. A tutorial on spectral clustering, pp. 1–32. arXiv 0711.0189.

Wainwright, M.J., Jordan, M.I., 2007. Graphical models, exponential families, and variational inference. Foundations and Trends in Machine Learning 1 (1–2), 1–305. http://dx.doi.org/10.1561/2200000001.

Wig, G.S., Laumann, T.O., Petersen, S.E., 2014. An approach for parcellating human cortical areas using resting-state correlations. NeuroImage 93 (2), 276–291. http://dx.doi.org/10.1016/j.neuroimage.2013.07.035.

Wittchen, H.U., Jacobi, F., Rehm, J., Gustavsson, A., Svensson, M., Jönsson, B., Olesen, J., Allgulander, C., Alonso, J., Faravelli, C., Fratiglioni, L., Jennum, P., Lieb, R., Maercker, A., van Os, J., Preisig, M., Salvador-Carulla, L., Simon, R., Steinhausen, H.C., 2011. The size and burden of mental disorders and other disorders of the brain in Europe 2010. Eur. Neuropsychopharmacol. 21 (9), 655–679. http://dx.doi.org/10.1016/j.euroneuro.2011.07.018.

Woolrich, M., Behrens, T., 2006. Variational Bayes inference of spatial mixture models for segmentation. IEEE Trans. Med. Imaging 25 (10), 1380–1391. http://dx.doi.org/10.1109/TMI.2006.880682.

Yarkoni, T., Poldrack, R.A., Nichols, T.E., Van Essen, D.C., Wager, T.D., 2011. Large-scale automated synthesis of human functional neuroimaging data. Nat. Meth. 8 (8), 665–670. http://dx.doi.org/10.1038/nmeth.1635.

Yeo, B.T.T., Krienen, F.M., Sepulcre, J., Sabuncu, M.R., Lashkari, D., Hollinshead, M., Roffman, J.L., Smoller, J.W., Zöllei, L., Polimeni, J.R., Fischl, B., Liu, H., Buckner, R.L., 2011. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. J. Neurophysiol. 106 (3), 1125–1165. http://dx.doi.org/10.1152/jn.00338.2011.

Yeo, B.T.T., Krienen, F.M., Eickhoff, S.B., Yaakub, S.N., Fox, P.T., Buckner, R.L., Asplund, C.L., Chee, M.W., 2015. Functional specialization and flexibility in human association cortex. Cerebral Cortex 25 (10), 3654–3672. http://dx.doi.org/10.1093/cercor/bhu217.

Zhang, S., Li, C.s.R., 2012. Functional connectivity mapping of the human precuneus by resting state fMRI. NeuroImage 59 (4), 3548–3562. http://dx.doi.org/10.1016/j.neuroimage.2011.11.023.

Zilles, K., Amunts, K., 2010. Centenary of Brodmann's map—conception and fate. Nat. Rev. Neurosci. 11 (2), 139–145. http://dx.doi.org/10.1038/nrn2776.

Zuo, X.N., Anderson, J.S., Bellec, P., Birn, R.M., Biswal, B.B., Blautzik, J., Breitner, J.C.S., Buckner, R.L., Calhoun, V.D., Castellanos, F.X., Chen, A., Chen, B., Chen, J., Chen, X., Colcombe, S.J., Courtney, W., Craddock, R.C., Di Martino, A., Dong, H.M., Fu, X., Gong, Q., Gorgolewski, K.J., Han, Y., He, Y., He, Y., Ho, E., Holmes, A., Hou, X.H., Huckins, J., Jiang, T., Jiang, Y., Kelley, W., Kelly, C., King, M., LaConte, S.M., Lainhart, J.E., Lei, X., Li, H.J., Li, K., Li, K., Lin, Q., Liu, D., Liu, J., Liu, X., Liu, Y., Lu, G., Lu, J., Luna, B., Luo, J., Lurie, D., Mao, Y., Margulies, D.S., Mayer, A.R., Meindl, T., Meyerand, M.E., Nan, W., Nielsen, J.A., O'Connor, D., Paulsen, D., Prabhakaran, V., Qi, Z., Qiu, J., Shao, C., Shehzad, Z., Tang, W., Villringer, A., Wang, H., Wang, K., Wei, D., Wei, G.X., Weng, X.C., Wu, X., Xu, T., Yang, N., Yang, Z., Zang, Y.F., Zhang, L., Zhang, Q., Zhang, Z., Zhang, Z., Zhao, K., Zhen, Z., Zhou, Y., Zhu, X.T., Milham, M.P., 2014. An open science resource for establishing reliability and reproducibility in functional connectomics. Sci. Data 1, 140049. http://dx.doi.org/10.1038/sdata.2014.49.