ANATOMO-FUNCTIONAL DESCRIPTION OF THE BRAIN : A PROBABILISTIC APPROACH

Benjamin Thyreau, Bertrand Thirion, Guillaume Flandin and Jean-Baptiste Poline

Service Hospitalier Frédéric Joliot
Département de Recherche Médicale - CEA - DSV, Inserm U562
4, Place du Général Leclerc
91401 Orsay Cedex

ABSTRACT

The acquisition of brain images in fMRI yields rich topographic information about the functional structure of the brain. However, these descriptions are limited by strong inter-subject variability. A recent approach to represent the gross functional architecture across the population as seen in fMRI consists in automatically defining accross-subjects brain parcels. This technique yields large-scale inter-subject correspondences while allowing some spatial relaxation in the alignment of the brains. We address here the open question of an optimal parameterization (number of parcels) of brain parcellations using information theoretic criteria and cross-validation. Moreover, a finer analysis of variance components enables us to better characterize intra- and inter-subject variability sources in parcellation models.

1. INTRODUCTION

Functional Magnetic Resonance Imaging (fMRI) is a recent modality for the exploration of brain function that produces activity maps of the whole brain. Neuroscientific inference proceeds by assessing the presence of a task-related signal increase at a spatial location, in a group of subjects. To date, spatial locations are defined as voxel coordinates, which implies that a spatial realignment of the images is applied prior to data analysis. One usually applies spatial deformations to the images to coregister anatomical structures of each subject to a common template [1]. This transformation makes the strong assumption that the spatial organization matches the functional one, which disregards the large inter-subject variability and is compensated by smoothing the data (smoothing kernel typically around 8-12 mm FWHM). Yet, inter-subject variability still needs to be better modelled and understood.

In order to overcome the shortcomings of spatial normalization, Flandin et al [3] have introduced a novel method of data analysis which divides the brain in many parcels of homogeneous functional activity and spatial location across subjects. This accounts for small spatial variation in the position of the functional activity across subjects. Several algorithms are possible to delineate the parcels. For instance, a spectral clustering technique was introduced in [7] while Flandin et al resorted to Gaussian Mixture Model on pooled (spatial/functional) coordinates.

In this work, we build on the dual nature of the GMM, which is not only a clustering technique, but also a density estimation technique. In particular, it enables us to address some important modelling issues, such as the dimensionality of the model, or the homogeneity of the population.

Choosing the number K of components in the GMM is somewhat analogous to the choice of the spatial smoothing kernel generally performed before group analysis in fMRI, as pointed out in [3]. Intuitively, the more spatial variability across subjects, the greater the smoothing kernel, and the smaller the number of parcels. To choose objectively an adequate number of parcels, we use information theoretic criteria and cross validation techniques on an fMRI study of 10 subjects, and show converging evidence for $K \sim 500$.

2. FEATURE SPACE AND STATISTICAL MODEL

We define the parcels as empirically derived functional modules, i.e. regions with homogeneous functional activity and compact spatial layout. Since they are derived from clustering techniques, it is necessary to use information from both the spatial and functional domain.

The anatomical information brought in the model consists in the three spatial coordinates $\tau(v)$ of each voxel v in Talairach space (which is the standard referential).

The functional input data comes from a General Linear Model (GLM) analysis of fMRI datasets [1]. In each subject, the set of images can be rewritten as a data matrix Y, by pooling all the voxels along a unique spatial axis. A design matrix X defines the temporal effects (i.e. stimulus-related effects) under study, and the GLM proceeds by estimating

the voxel-based parameters β such that

$$Y = X\beta + E, \quad (1)$$

where E represents a residual noise, independent across voxels but temporally correlated. The GLM algorithm produces weighted least-squares estimates $\hat{\beta}(v)$ of the parameters for each voxel v, and of their covariance $\hat{\Lambda}(v)$. In our framework, we interpret the GLM as a posterior distribution of the parameters with uninformative (improper) priors [1]

$$P(\beta(v)|Y(v)) = \mathcal{N}(\hat{\beta}(v), \hat{\Lambda}(v)). \quad (2)$$

We now consider the parameters β as functional coordinates of the voxels. Assuming a spatially stationary noise model, all the matrices $\Lambda(v)$ are equal up to a multiplicative factor ($\hat{\Lambda}(v) = \hat{\sigma}(v)^2\Lambda$), which allows us to consider them as the identity by multiplying the β's with $\Lambda^{-\frac{1}{2}}$. The voxels of all datasets can be pooled in a single feature space by forming feature vectors $\mathcal{F}(v) = (\tau, \frac{\hat{\beta}}{\hat{\sigma}})(v)$. In order to balance anatomical and functional information, as well as for the sake of computation speed, it may be necessary to reduce the dimension d_F of the functional information to e.g. $d_F = 3$. We thus perform a Singular Value Decomposition (SVD) of the functional feature matrix $\{\frac{\hat{\beta}}{\hat{\sigma}}(v)\}$ pooled across subjects. The dimension d of the feature space is then the sum of the spatial and functional ones : $d = 3 + d_F$.

2.1. Statistical model

The feature data $\mathcal{F}$ is then submitted to a Gaussian Mixture Model (GMM). The data is represented by K weighted multidimensional normal densities, with parameters $\{\mu_k, \Sigma_k\}$.

$$p(\mathcal{F}(v) = F|\Theta_K, K) = \sum_{k=1}^{K} \pi_k \mathcal{N}(F; \mu_k, \Sigma_k), \quad (3)$$

where the set of parameters to estimate is:

$$\Theta_K = \{\pi_1, ..., \pi_{K-1}, \mu_1, ..., \mu_K, \Sigma_1, ..., \Sigma_K\}. \quad (4)$$

The number η_K of free parameters of the GMM is thus

$$\eta_K = \begin{cases} K - 1 + K(d(d+3)/2) & \text{for full covariance} \\ K - 1 + 2Kd & \text{for diagonal covariance} \end{cases}$$

Thereafter, we force the model to have diagonal covariance and precision matrices. This is an advantage for both computational (matrix inversibility, computation speed) and modelling (separability of the model) purposes.

The estimation of Θ_K is carried out through a standard Expectation-Maximisation (EM) algorithm [2]. We use simple numerical shortcuts given the amount of data ($N \simeq 5.10^5$ pooled voxels, with d-dimensional features). Initialization is done with a c-means on the spatial features only to obtain spatially compact parcels. We noticed empirically that this procedure avoids numerical degeneracies that are frequent with EM estimations of GMMs.

[1] Rigorously, a Student distribution should be considered, but given the typically high number of degrees of freedom ($\nu > 100$), this has very little impact on the following analysis

2.2. Selection of the number of classes

Here, we are interested in determining the value for K, using either information theoretic criteria or cross-validation. We simply look at the evolution of the fitting process when varying the number of clusters (see Fig. 1). Other approaches, based e.g. on Reversible jump Markov Chain Monte Carlo techniques [4] would be appropriate, but their prohibitive computation and storage costs make them intractable for huge datasets. A Variational Bayes approach to the GMM estimation can also yield an optimal value for K, that can be well approximated by a BIC criterion [6]. Here, we rely on two criteria to evaluate the fitting process: the Bayesian Information Criterion (BIC), and the generalizability of the model (Cross Validation). Let $P(\mathcal{F}|K)$ be the probability that the feature data $\mathcal{F}$ is generated from a model with K components, BIC reads :

$$\begin{aligned} -\frac{1}{2}BIC(K) &= \log P(\mathcal{F}|K) & (5) \\ &= \log \int P(\mathcal{F}|\Theta, K) P(\Theta|K) d\Theta & (6) \\ &= \log P(\mathcal{F}|\hat{\Theta}_K) - \frac{\eta_K}{2} log(N) + O(1), \end{aligned}$$

where $\hat{\Theta}_K$ represents the estimated, locally optimal values for Θ_K. The third equality follows from the so-called Laplace approximation of the integral in Eq. (6). Then, assuming a flat prior over K in a range $[1 \dots K_{max}]$, the likelihood of the model given the data $P(K|\mathcal{F})$ is proportional to $P(\mathcal{F}|K)$ [5].

In fMRI data analysis, inter-subject variability is prominent. Another way to validate a data model is thus to compare the goodness of fit of a model estimated on $(S-1)$ subjects on the data of subject S. More precisely, for $s \in [1, .., S]$, we form the dataset $\mathcal{F}^{(-s)}$, where the data is pooled from all subjects excepts s, and estimate the corresponding parameters $\hat{\Theta}_K^{(-s)}$. Using Eq. (3), we estimate the likelihood of the left-out dataset $\mathcal{F}^{(s)}$. When averaged over s, this yields $CV(K)$.

2.3. A hierarchical approach

In order to gain more insight on the modelling of inter-subject variability, we can make the distinction between intra-subject variance, which simply measures the extent of the parcels in the feature space, and inter-subject variance, which

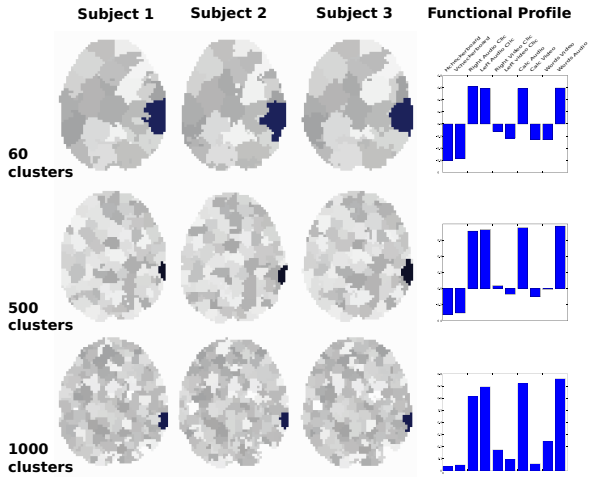
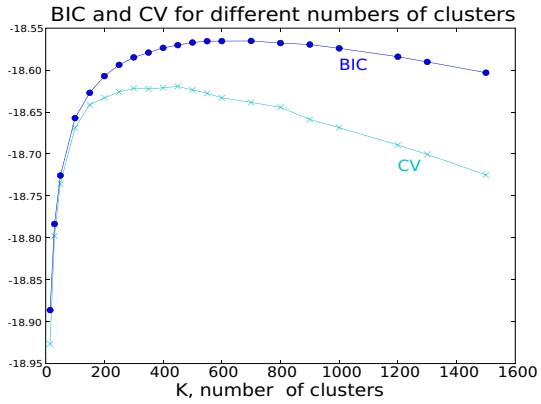


Fig. 1. Parcellation of a group of datasets with a GMM model, for different values of K. (Left) After convergence of the GMM estimation algorithm, each voxel from each subject can be assigned to a preferred cluster. This reveals different levels of detail according to K(60, 500 or 1000). (Right) The average functional activity associated with any parcel k is given by μ_k, which corresponds to a typical parcel response to the 10 experimental conditions of the paradigm. Increasing K enhances the functional profile of each parcel. In the above example, we obtain sharper responses to the four auditory stimulations. The highlighted parcels above illustrates an audio-sensitive area.

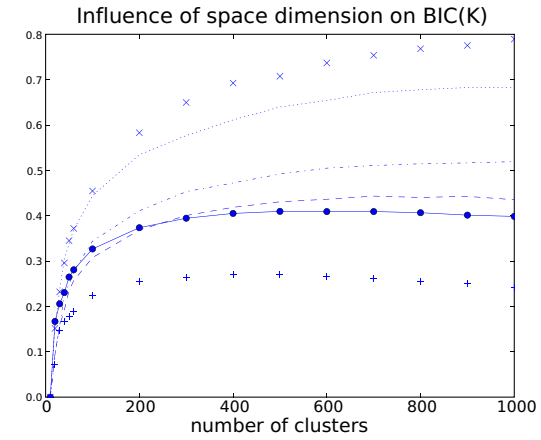
measures the mismatch of the feature data drawn from different subjects. To this end, we use an empirical Bayesian formulation of the GMM, in which the prior is the group-averaged information, and the posterior adapts to each subject's data. The implementation relies on Variational Bayes approximations, which are detailed in [6]. Based on this formulation, we compute the logarithm $\Gamma(K)$ of the determinant of the inter-subject covariance, averaged on the K parcels, and $\bar{\Sigma}(K)$, which is the parcel- and subject- averaged log-determinant of the intra-subject covariance matrices. We study these quantities as functions of K.

3. RESULTS

We tested the algorithm on a dataset comprising $S = 10$ subjects who underwent the same fMRI protocol. This protocol is intended to activate many brain areas related to several cognitive functions (motor, audio and video perception, sentence analysis and computation), thus producing some kind of functional benchmark. It is a subsample of the data used in [7], with the same pre-processing and GLM analysis procedures. We applied our model on S fMRI datasets described above, for values of K ranging from 20 to 1500, and computed the corresponding BIC and Cross-Validation criteria.



(a)



(b)

Fig. 2. (a) Evolution of -BIC/2 and the CV as functions of K, for $d_F = 3$. (b) Evolution of the -BIC/2 criterion when d_F varies from 2 (lower one) to 7 (upper one). The curves are normalized by subtracting the value obtained for $K = 10$.

3.1. Analysis with different criteria

Fig. 2(a) shows the evolution of - BIC / 2 as a function of K, with $d_F = 3$. We observe that the highest values are reached around $K \in [500, 700]$ before decreasing slowly for larger K. The function CV(K) is similar, with a maximum at $K^\star \sim 400$ Our results show that an optimal representation for this 10 subjects-group on this specific paradigm, is obtained with (~ 500) parcels of approximately $2,7cm^3$ each.

3.2. Influence of the input data dimension

The probability distribution of the features $\mathcal{F}$ is compound, with anatomical and functional subspaces endowed with dif-

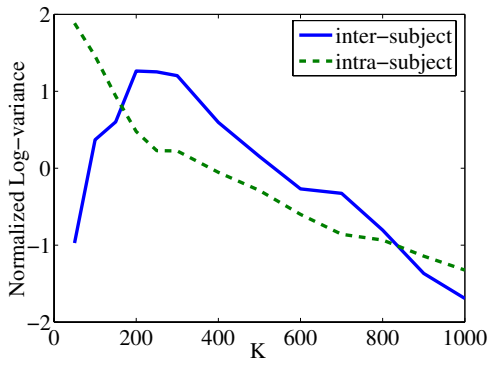


Fig. 3. Evolution of the inter- and intra- subject log-variance when K varies ($d_F = 3$). For the sake of clarity, the curves have been normalized. Intra-subject variance steadily decreases with K, while inter subject variance has a maximum for $K \simeq 250$.

ferent statistical characteristics. It is thus important to assess the impact of the relative amount of anatomical and functional information on the choice of K. Here d_F controls the amount of functional information in the model. In Fig. 2(b), we present the BIC curves obtained when letting d_F vary from 2 to 7.

As a result, increasing d_F yields an increase of the optimal $K^\star$ value. This indicates that the optimal description level depends upon the amount of information available for the characterization of different modules.

3.3. Analysis of the variance components

The quantities $\Gamma(K)$ and $\bar{\Sigma}(K)$ are shown in Fig. 3 for various values of K, and $d_F = 3$.

This shows that $\bar{\Sigma}(K)$ steadily decreases with K, which reflects the fact that the spread of the components in the feature space is reduced when K increases. More interestingly, the inter-subject variance first increases, reaches a maximum for $K \simeq 250$, and decreases afterwards. This result, which is reproducible across different parametric settings of the VB estimation algorithm (not shown), implies that situations where $K < 200$ are suboptimal from a modelling point of view: using low K, partial volume effects dominate, and inter-subject variability is under-estimated. The fact that $\Gamma(K)$ decreases for $K > 250$ indicates that finer across-subjects correspondences can always be found when one models the data with higher precision, which was not guaranteed. However, this effect is not sufficient to compensate the penalty terms of the BIC criterion, nor to guarantee generalizability of the parcellation model to new subjects.

4. DISCUSSION AND CONCLUSION

We addressed here the open problem of most adequate spatial level of description for fMRI data in group analysis. From our results $K \simeq 500$ can be viewed as an adequate level of description for inter-subject analyzes in fMRI. At the current state of the art, it cannot be interpreted as a number of intrinsic brain modules. In particular, it depends on the experimental paradigm, on the MRI scanner properties and image processing choices, as well as the amount of functional information available for characterizing brain regions (see the dependence on d_F).

Although based on different hypotheses, BIC and CV criteria yield similar results for the optimal value of K. The difficulty of making definitive assertion about the optimal K is illustrated by the flatness of the curves, especially when related to inter-subject variability. CV might be slightly more conservative, which is logical since it measures the generalizability of the parcel description rather than the penalized goodness of fit. Last, the analysis of variance components with a hierarchical GMM, reveals that choosing $K > 200$ is preferable, since the parcels are too large otherwise, so that across-subject variance is dominated by intra-parcel variability.

Knowing the optimal value for K will hopefully enable us to more accurately apply clustering techniques in functional data analysis, helping us to make more sensitive inference and more informative connectivity studies. Last, the probabilistic point of view of parcellation might be further extended to study population homogeneity and to perform subject classification.

5. REFERENCES

[1] J. Ashburner, K.J. Friston, and W. Penny, editors. *Human Brain Function, 2nd Edition*. Academic press, 2004.

[2] A.P. Dempster, A.P. Laird, and D.B. Rubin. Maximum likelihood from incomplete data via the em algorithm (with discussion). *Journal of the Royal Statistical Society, Series B*, 39:1–38, 1977.

[3] G. Flandin, F. Kherif, X. Pennec, G. Malandain, N. Ayache, and J.-B. Poline. Improved detection sensitivity of functional MRI data using a brain parcellation technique. In *Proc. 5th MICCAI*, LNCS 2488 (Part I), pages 467–474, Tokyo, Japan, September 2002. Springer Verlag.

[4] Peter J. Green. Reversible jump MCMC computation and Bayesian model determination. *Biometrika*, 82:711–732, 1995.

[5] T. Hastie, R. Tibshirani, and J. Friedman, editors. Springer Series in Statistics, 2001.

[6] W.D. Penny. Variational bayes for d-dimensional gaussian mixture models. Technical report, Wellcome Department of Cognitive Neurology, University College London, 2001.

[7] B. Thirion, G. Flandin, P. Pinel, A. Roche, P. Ciuciu, and J.-B. Poline. Dealing with the shortcomings of spatial normalization: Multi-subject parcellation of fmri datasets. *Hum. Brain Mapp.*, in Press.