

Parcellation Schemes and Statistical Tests to Detect Active Regions on the Cortical Surface

Bertrand Thirion^{1,2}, Alan Tucholka^{1,2}, and Jean-Baptiste Poline^{1,2}

¹ Parietal team, INRIA Saclay-le-de-France, Saclay, France
CEA Saclay, Bâtiment 145, 91191, Gif-sur-Yvette, France
bertrand.thirion@inria.fr,

² CEA, DSV, I²BM, Neurospin,
CEA Saclay, Bâtiment 145, 91191, Gif-sur-Yvette, France

Abstract. Activation detection in functional Magnetic Resonance Imaging (fMRI) datasets is usually performed by thresholding activation maps in the brain volume or, better, on the cortical surface. However, basing the analysis on a site-by-site statistical decision may be detrimental both to the interpretation of the results and to the sensitivity of the analysis, because a perfect point-to-point correspondence of brain surfaces from multiple subjects cannot be guaranteed in practice. In this paper, we propose a new approach that first defines anatomical regions such as cortical gyri outlined on the cortical surface, and then segments these regions into functionally homogeneous structures using a parcellation procedure that includes an explicit between-subject variability model, i.e. random effects. We show that random effects inference can be performed in this framework. Our procedure allows an exact control of the specificity using permutation techniques, and we show that the sensitivity of this approach is higher than the sensitivity of voxel- or cluster-level random effects tests performed on the cortical surface.

Keywords: statistical testing, EM algorithm, spatial models, neuroimaging

1 Introduction

In neuroimaging, brain activation detection is traditionally performed through the thresholding of statistical maps. In contrast with standard volume-based analyses, cortical surface mapping (CSM) consists in detecting brain activations on the cortical surface, after projection of the fMRI volume-based data onto the surface (Fischl et al. (1999), Andrade et al. (2001)). This offers the advantage of positioning functional activations in the two-dimensional space where they are indeed generated, as well as a better sensitivity/specificity compromise due to the limitation of the statistical tests to grey matter only. Although it has been suggested that CSM could be more sensitive in group studies than traditional volume-based studies (see e.g. Fischl et al. (1999)), inter-subject analyses have been limited by the problem of defining properly brain location on the surface in the absence of a standard coordinate system. A meaningful solution to this problem consists in defining intermediate representations, such as gyri, that represent a delineation of the main regions on

the cortical surface, as obtained from an atlas, see Fischl et al. 2004. Then, one still has to decide how to take into account such gyral parcellations when making statistical tests, or how to test region-specific hypotheses based on the available data. Because gyri represent the cortex at a very coarse resolution, functional information should be used to test more precise regions.

A second aspect of the problem is that it is not possible to require a perfect match between brain meshes since they have different shapes, as seen with various sulci or gyrification indexes. But, even assuming that a perfect anatomical match can be obtained between the brains of several individuals, it is not clear that functional regions would be matched perfectly. A promising solution consists in introducing an intermediate representation between mesh vertices and gyri, for instance through the concept of brain parcellation (Tucholka et al. (2008)). Conceptually, such parcellations are defined through the use of both anatomical and functional information. Although this may provide meaningful entities, it is not clear how these structures can be used to infer active regions across subjects, i.e. how to make random effects analyses. In this work, we address this particular question by introducing a new probabilistic parcellation framework that includes random effects, and finally allows the test of some contrasts of interest. An unbiased assessment of these tests using permutations is possible thanks to the relatively mild computation cost of the proposed method.

In Section 2, we develop the random-effects anatomo-functional permutation model, then we described the validation procedure and give some results on a real dataset in Section 3.

2 Model

2.1 Inputs and notations

Let $X^s = \{x_i^s\}_{i=1..I^s}$ be a set of pre-defined coordinates that represent the position of cortical sites in a subject $s \in \{1, \dots, S\}$ in a certain gyrus $g \in \{1, \dots, G\}$ (in Sections 2.1-2.4, we omit the dependence on the gyrus to keep notations simple). These coordinates are assumed to yield an approximate correspondence across individuals. Let $Y^s = \{y_i^s\}_{i=1..I^s}$ be n_f -dimensional vectors that represent the functional activity related to these sites in subjects $s \in \{1, \dots, S\}$. In this work, we use $n_f = 1$. Let $K > 0$ be the number of components of the probabilistic parcellation.

Let $(w_{ik}^s)_{i=1..I^s, k=1..K}$ denote the probability that the site i belongs to component $k \in \{1, \dots, K\}$, i.e. $p(z_i^s = k)$, given its position. w_{ik}^s are function of the positions x_i^s and a set of two-dimensional coordinates $\mathcal{T} = (\tau_k)_{k=1..K}$ that describe the position of the clusters on the cortex, and a spatial variance parameter γ :

$$w_{ik}^s(\mathcal{T}) = \frac{\exp(-\frac{\|x_i^s - \tau_k\|^2}{2\gamma^2})}{\sum_{l=1}^K \exp(-\frac{\|x_i^s - \tau_l\|^2}{2\gamma^2})} \quad (1)$$

Note that the position variables \mathcal{T} are shared across subjects.

2.2 Hierarchical parcellations

The parameters $\Theta = (\theta_k)_{k=1..K}$ used to model the functional information $(Y^s)_{s=1..S}$ are part of a hierarchical model that includes group-level and subject-specific activation maps in the chosen gyrus. The activation is assumed to be normally distributed in the population, and then normally distributed in each subject given the parameters of this subject: $\forall k \in \{1..K\}$, let (μ_k, Σ_k) be the population parameters, and $((\mu_k^s, \Sigma_k^s)_{s=1..S})$ be the individual parameters.

$$p(\mu_k^s | \mu_k, \Sigma_k) = \mathcal{N}(\mu_k^s; \mu_k, \Sigma_k) \quad (2)$$

$$p(y_i^s | z_i^s = k, \mu_k^s, \Sigma_k^s) = \mathcal{N}(y_i^s; \mu_k^s, \Sigma_k^s) \quad (3)$$

The parameters of the model are thus $\Theta = (\mu_k, \Sigma_k, (\Sigma_k^s)_{s=1..S})_{k=1..K}$, and the log-likelihood of the data can be written:

$$\mathcal{L}(Y|\Theta, \mathcal{T}) = \sum_{s=1}^S \sum_{i=1}^{I_s} \log \left(\sum_{k=1}^K w_{ik}^s(\mathcal{T}) \mathcal{N}(y_i^s; \mu_k, \Sigma_k + \Sigma_k^s) \right) \quad (4)$$

This assumes conditional independence of the functional information given the parcel parameters, as is classically done for mixture models. The model is summarized in Fig. 1. The important aspect with the random effects model is that it allows second level inference: let c be a certain contrast of experimental conditions; if we define second level statistics as

$$t_{RFX}(k) = \frac{c^T \mu_k}{\sqrt{c^T \Sigma_k c}} \sqrt{S-1} \quad (5)$$

This statistic is readily computed for each parcel in each gyrus. Note the distribution of this quantity cannot be assumed as known under the null hypothesis, but a corrected threshold can be derived through statistical resampling procedures (see Sec. 2.5).

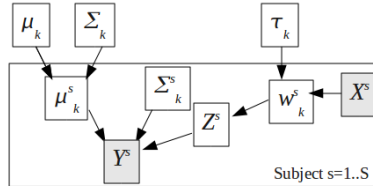


Fig. 1. Generative model of the data used in this work: in each subject $s \in 1..S$, the observed data Y_s , results from a spatial model, shared across subjects, that provides the probability w that each surface point belongs to a parcel, and a hierarchical model of the functional parameters, with both subject-specific (μ_k^s, Σ_k^s) and group-level (μ_k, Σ_k) mean and covariance parameters. The observed variables are shaded.

2.3 Estimation of the model

We use an alternate optimization scheme, in which Θ and \mathcal{T} are optimized in turn in order to maximize the log-likelihood of the data in Eq. (4). $\max_{\Theta} p(Y|\Theta, \mathcal{T})$ is obtained through the standard EM algorithm while $\max_{\mathcal{T}} p(Y|\Theta, \mathcal{T})$ is obtained through gradient descent performed simultaneously.

- E-step: let $Z^s = (z_i^s)_{i=1..I}$ be the allocation variables of the mixture model.

$$p(z_i^s = k) = \frac{w_{ik}^s \mathcal{N}(y_i^s; \theta_k)}{\sum_{l=1}^K w_{il}^s \mathcal{N}(y_i^s; \theta_l)} \quad (6)$$

- M-step: $(\mu_k, \Sigma_k, \Sigma_k^s) = \operatorname{argmax}_{\Theta} \mathbb{E}_z \log p(Y, Z|\theta, \mathcal{T})$, which yields an internal EM algorithm, where μ_k^s are the hidden variables, while the maximization is carried out over the other variables $(\mu_k, \Sigma_k, \Sigma_k^s)$:

$$p(\mu_k^s | \mu_k, \Sigma_k, \Sigma_k^s, Y^s, Z^s) = \mathcal{N}(\mu_k^s; \Lambda_k^s [(\Sigma_k)^{-1} \mu_k + m_k^s (\Sigma_k^s)^{-1} n_k^s], \Lambda_k^s) \quad (7)$$

where $m_k^s = \frac{\sum_{i=1}^{I_s} p(z_i^s = k) y_i}{\sum_{i=1}^{I_s} p(z_i^s = k)}$, $n_k^s = \sum_{i=1}^{I_s} p(z_i^s = k)$ and $\Lambda_k^s = [(\Sigma_k)^{-1} + n_k^s (\Sigma_k^s)^{-1}]^{-1}$.

Then the internal M-step is performed:

$$\mu_k = \frac{1}{S} \sum_{s=1}^S \mu_k^s, \quad (8)$$

$$\Sigma_k = \frac{1}{S} \sum_{s=1}^S (\mu_k^s - \mu_k)^T (\mu_k^s - \mu_k), \quad (9)$$

$$\Sigma_k^s = \frac{\sum_{i=1}^{I_s} p(z_i^s = k) (y_i^s - \mu_k^s)^T (y_i^s - \mu_k^s)}{\sum_{i=1}^{I_s} p(z_i^s = k)} \quad (10)$$

Eqs. (7) and (8-10) are iterated until convergence. Furthermore, we use in this algorithm a regularization procedure (Fraleay and Raftery (2007)) in order to ensure that the different terms do not converge toward a degenerated solution (e.g. null variance).

- C-step:

$$\frac{\nabla \mathcal{L}}{\nabla \tau_k} = \frac{1}{\gamma^2} \sum_{s=1}^S \sum_{i=1}^{I_s} (x_i^s - \tau_k) w_{ik}^s \left(\frac{\mathcal{N}(y_i^s; \theta_k)}{\sum_{l=1}^K w_{il}^s \mathcal{N}(y_i^s; \theta_l)} - 1 \right) \quad (11)$$

We perform parameters updates that are reminiscent of the mean-shift procedure (Comaniciu and Meer, 2002): $\tau_k \rightarrow \tau_k + \delta \tau_k$, where

$$\delta \tau_k = \sum_{s=1}^S \frac{1}{I_s} \sum_{i=1}^{I_s} (x_i^s - \tau_k) \omega_i^s, \text{ and } \omega_i^s = w_{ik}^s \left(\frac{\mathcal{N}(y_i^s; \theta_k)}{\sum_{l=1}^K w_{il}^s \mathcal{N}(y_i^s; \theta_l)} - 1 \right) \quad (12)$$

Alternating these three steps (6,7-10,12) is very effective in practice: the log likelihood often converges in 5 to 10 iterations.

2.4 Optimizing the parameters of the model

The two free parameters of the model are the number of parcels K and the spatial shrinkage parameter γ in each gyrus g . We optimize conjointly these parameters during a first analysis session using the cross-validated likelihood as criterion and a grid search approach. To compute the log-likelihood on a new dataset (X^σ, Y^σ) , we need to estimate the covariance matrices Σ_k^σ within the new subject; we take the maximum likelihood estimator:

$$\mathcal{L}(Y^\sigma | X^\sigma, \Theta, \mathcal{T}) = \max_{(\Sigma_k^\sigma)_{k \in \{1..K\}}} \sum_{i=1}^{I_\sigma} \log \left(\sum_{k=1}^K w_{ik}^s(x_i^\sigma, \mathcal{T}) \mathcal{N}(y_i^\sigma; \mu_k, \Sigma_k + \Sigma_k^\sigma) \right) \quad (13)$$

The optimal values $(K^*(g), \gamma^*(g))$ are then retained for the RFX procedure.

2.5 Random-effects (RFX) inference procedure

In order to control the specificity of the parcel-based statistical procedure, we need to know the distribution of the statistic (5) under the null hypothesis, i.e. when no activation is present. This cannot be done analytically, because the value of the statistic depends on the whole parcellation procedure. We tabulate the distribution of the null hypothesis by randomly swapping the sign of the data related to the tested contrast across subjects, and then recomputing the parcels and the associated statistic \tilde{t} . Next, this procedure has to be carried out on the whole volume. After $R = 10^3$ randomizations, the maximal parcel-level statistic across gyri is tabulated:

$$\bar{t}_r = \max_{g \in 1..G} \max_{k \in \{1.., K^*(g)\}} \tilde{t}_k, \forall r \in \{1, \dots, R\} \quad (14)$$

and the threshold t_α for a specificity α ($\alpha = 0.05$ typically), corrected for multiple comparisons across parcels and gyri, is chosen as the $(1 - \alpha)$ quantile of the distribution of (\bar{t}_r) . The probability of a parcel-based t-value being greater than t_α in any parcel of any gyrus by chance is thus lower than α .

3 Results on a real dataset

Dataset. A localizer protocol was acquired on a 1.5T GE MRI scanner. The Freesurfer package was used to segment different anatomical compartments from the anatomical image of the brain of each subject, providing white and grey matter mesh, and segmenting the sulci (Fischl et al. 1999). This sequence of processing was applied systematically to all available brains and the quality of resulting segmentation was visually checked. A surface-based coordinate system that represents sulco-gyral anatomy is finally obtained, and the cortical surface is subdivided into gyri. All cortical meshes are then resampled

so that the nodes of the mesh are in one-to-one correspondence across subjects. For all subjects a standard preprocessing of fMRI data was performed using the SPM5 software. Functional images were then projected onto the grey/white interface using the method described in (Operto et al. (2008)). Subsequently, on each functional dataset a Linear Model-based analysis was carried out to obtain task-related activity maps Y for different contrasts of experimental conditions. We have tested 4 different contrasts related to motor, auditory, computation and reading functions respectively, but in separate analyses, so that we always have $n_f = 1$: taking $n_f > 1$ is possible, but would complicate the permutation framework described in Section 2.5.

Assessment of the PRFX statistic. Average left and right brain hemisphere meshes have been derived and are used for display. These average meshes are then parcelled using the maximum a posteriori label of each node given its position and the \mathcal{T} parameters learnt by the algorithm. Our parcel-based inference (PRFX) is compared to more classical statistical procedures (Hayasaka et al. 2003, Rocheet al. 2007) used on the cortical surface: *i*) a vertex- or node-level control procedure yields the threshold t_α so that the probability of the random effects statistic at a single node being greater than t_α is less than α (0.05 typically); it is obtained by tabulating the maximal t -value of any node under the null hypothesis by a permutation procedure (VRFX); *ii*) a cluster-level procedure (CRFX) that considers all the clusters (connected components) of nodes with a t values higher than a certain threshold (we take the threshold corresponding to $p < 0.01$, uncorrected) and tabulates the distribution of the largest cluster size under the null hypothesis using the same permutation approach, so that the risk of detecting one cluster larger than the size threshold is less than α .

Outcome of the procedure. We concentrate on the contrast *sentence reading minus checkerboard viewing* that yields regions specifically involved in the reading task. Altogether, the parcellation outlines about 500 regions in each hemisphere. Activation detection images are given in Fig. 2. In the left hemisphere, activation specific to the reading task is found along the Superior Temporal Sulcus (superior and middle temporal gyrus), in Broca’s area, in the middle part of the pre-central gyrus, and in the Superior Frontal Gyrus (Supplementary Motor Areas, SMA) by both PRFX and CRFX. Moreover, PRFX also detects significant activity in the temporal pole and the inferior temporal gyrus. VRFX detects tiny spots, barely visible in Fig. 2 in all these regions, except the superior Frontal region. In the right hemisphere, all three methods detect some activity in the Superior Temporal Sulcus (middle temporal gyrus), but only PRFX detects activity in the right SMA. Overall, the PRFX procedure is more sensitive than the other techniques. We also tested other contrasts and found that the PRFX procedure is at least as sensitive as the others. The VRFX procedure detects very few active nodes, but with a stronger control, in the sense that the null hypothesis is indeed rejected in

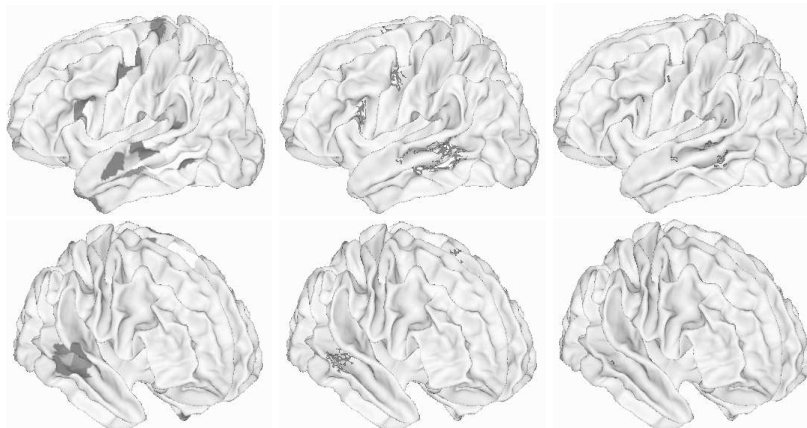


Fig. 2. Outcome of the parcel-based (left), cluster-based (middle) and node-based (right) random effects analyses in the left (top) and right (bottom) hemisphere. All the maps are corrected at the $p < 0.05$ parcel-, cluster- and voxel-level, respectively.

each detected node. Nevertheless, it detects at least one active node in almost all the regions found with the other approaches.

The CRFX procedure detects more extended regions than VRFX, but rejects only the global null hypothesis in these clusters, i.e. it does not reject the null hypothesis in any particular node. The same applies for the Parcel-based random effect procedure: it allows the rejection of the null hypothesis in a certain portion of a pre-defined gyrus, not on all the nodes of the finally outlined region.

4 Discussion

This work presents a new procedure to segment brain regions at a spatial scale that is intermediate between the mesh vertices and the anatomical gyri, which are too coarse (34 in each hemisphere with Freesurfer in (Fischl et al. (2004))) for an accurate functional description of the cortical surface. The main novelty of the presented work is to introduce a probabilistic model with random effects, which introduces the distinction between two sources of variance: *i*) the variance related to the spatial spread of the parcels, and thus simply represents the resolution which is chosen to analyse the data and *ii*) the between-subject variance, that represents the intrinsic functional variability between individuals, as well as potential spatial misfits. Besides, the introduction of the different variance components allows group-level inference, i.e. the computation of statistics that represent the magnitude of the average effect in the population, when compared to between-subjects fMRI signal variability.

As can clearly be seen in the results section, the method compares very favorably in terms of sensitivity with random effects analyses performed on the coregistered and resampled meshes, both at the cluster- and vertex- level. Our interpretation of this gain is that the proposed approach better adapts to the individual configurations/features. Importantly, the method outlines extended regions or parcels, which potentially provides a less biased representation than a few mesh vertices: indeed, parcels represent the position of a *region* in the standard space. Finally, the results are easily interpreted, given that each region belongs to a pre-defined anatomical gyrus.

The proposed model still requires the calibration of two parameters γ and K , which can be made automatically using standard model selection procedures (BIC, cross-validation). When these parameters are fixed, the proposed model is not expensive computationally, so that permutation-based tests remain affordable. Moreover, the computation can be performed in parallel for the different gyri. Using a python implementation, we could run the whole framework in less than 24 hours.

A relatively straightforward extension of the present framework includes the adaptation to more complex populations, where behavioural or clinical score are available to characterize the between subject variability of subgroup structure in the observed population. This might be particularly useful to derive interpretable, i.e. few discriminative features to separate the populations.

References

- ANDRADE, A., KHERIF, F. et al. (2001): Detection of fMRI activation using cortical surface mapping. *Hum. Brain Mapp.* 12, 79–93
- COMANICIU, D. and MEER, P. (2002): Mean shift: a robust approach toward feature space analysis. *IEEE PAMI* 24(5), 603–619
- FISCHL, B., SERENO, M.I., TOOTELL, R.B. and DALE, A.M. (1999): High-resolution intersubject averaging and a coordinate system for the cortical surface. *Hum. Brain Mapp.* 8(4), 272–284
- FISCHL, B., VAN DER KOUWE, A., et al. (2004): Automatically parcellating the human cerebral cortex. *Cereb. Cortex* 14(1), 11–22
- FRALEY, C. and RAFTERY, A. (2007): Bayesian regularization for normal mixture estimation and model-based clustering. *Journal of Classif.* 24, 155–181
- HAYASAKA, S. and NICHOLS, T.E. (2003): Validating Cluster Size Inference: Random Field and Permutation Methods *NeuroImage* 20(4):2343–2356
- OPERTO, G., BULOT, R., ANTON, J.L. and COULON, O. (2008): Projection of fMRI data onto the cortical surface using anatomically-informed convolution kernels. *NeuroImage* 39(1), 127–135
- ROCHE, A., MÉRIAUX, S., KELLER, M. and THIRION, B. (2007): Mixed-effects statistics for group analysis in fMRI: A nonparametric maximum likelihood approach. *NeuroImage*, 38:501–510
- TUCHOLKA, A., THIRION, B., PERROT, M., PINEL, P., MANGIN, J.F. and POLINE, J.B. (2008): Probabilistic anatomo-functional parcellation of the cortex: how many regions? In: Metaxas, D.; Axel, L.; Fichtinger, G.; Szekely, G. (Eds.) *11th Proc. MICCAI LNCS* Vol. 5242, Springer Verlag, Berlin.