Medical Image Analysis 16 (2012) 976–990

Contents lists available at SciVerse ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

Structural analysis of fMRI data: A surface-based framework for multi-subject studies

Grégory Operto [a,b], Denis Rivière [b], Bernard Fertil [a], Rémy Bulot [a], Jean-François Mangin [b], Olivier Coulon [a,*]

[a] LSIS Laboratory, CNRS, Marseille, France
[b] Neurospin Center, CEA, Gif-Sur-Yvette, France

ARTICLE INFO

Article history:
Received 19 May 2011
Received in revised form 14 February 2012
Accepted 17 February 2012
Available online 28 February 2012

Keywords:
Functional magnetic resonance imaging (fMRI)
Cortical analyses
Structural methods
Surface-based methods
Group analysis

ABSTRACT

We present a method for fMRI data group analysis that makes the link between two distinct frameworks: surface-based techniques, which process data in the domain defined by the surface of the cortex, and structural techniques, which use object-based representations of the data as opposed to voxel-based ones. This work is a natural surface-based extension of the volume-based structural approach presented in a previous paper. A multi-scale surface-based representation of individual activation maps is first computed for each subject. Then the inter-subject matching and the activation detection decision are performed jointly by optimization of a Markovian model. Finally, a significance measure is computed in a non-parametric way for the results, in order to assess their relevance and control the risk of type I error. The method is applied on simulated and real data and the results are compared to those produced by standard analyses. The surface-based structural analysis is shown to be particularly robust to inter-subject spatial variability and to produce relevant results with good specificity and sensitivity. We also demonstrate the advantages of the surface-based approach by comparing with the results of a 3D structural analysis.

1. Introduction

Multi-subject studies in neuroimaging open the way to exploring fundamental questions related to the human brain function. In particular, group analysis of functional magnetic resonance imaging (fMRI) data aims at building descriptions of unicity and diversity across a pool of subjects for a same cognitive experiment. When the pool is sufficiently large, results can then be generalized to a population. Still, matching these subjects together strongly suffers from the inter-subject variability that exists at different levels, anatomical, physiological or functional. To circumvent this issue, the historical and still usual procedure consists in normalizing every scanned subject to a common anatomical space (Gholipour et al., 2007). Hence, any location (i.e. voxel) in that space is supposed to correspond to the same location in the brain of each subject and multi-subject comparisons can consequently be carried out on the basis of this spatial matching. As the usual anatomical normalization does not entirely solve the problem of cross-subject spatial matching, an additional spatial smoothing is commonly performed on the functional data in order to increase the overall inter-subject matching between homologous brain structures. Then, due to the variability across responses from one subject to another, a second-level model is defined to make inference on voxels that showed a group-scale effect (Beckmann et al., 2003). Decisions are classically taken using Student *t*-tests, upon the usual hypothesis that the signals across subjects are normally distributed. A specific threshold is chosen on the statistical values and this results in a group activation map showing voxels which were activated across subjects. The literature abounds with activation detection techniques (Roche et al., 2004; Mériaux et al., 2006; Woolrich et al., 2004; Penny et al., 2005; Flandin and Penny, 2007; Keller et al., 2009; Smith and Nichols, 2009). Most of those methods encounter three limitations:

- They are voxel based. However, voxels are only the acquisition space and do not have any anatomical meaning, other than the approximate localization provided by spatial normalization. At this level, variability is difficult to encode, whereas at the object level it is clearer and variability can be encoded, e.g. in terms of size or in terms of variance of spatial location. In this context, normalization tends to neglect the inter-subject variability. Notably, a single triplet of coordinates in this common space is likely to point to different – especially cortical – structures from one subject to another (Uylings et al., 2005; Crivello et al., 2002). Regions of interest (ROI)-based analyses are generally a way to

* Corresponding author.
E-mail address: olivier.coulon@univmed.fr (O. Coulon).

1361-8415/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.media.2012.02.007

establish correspondence between subjects that overcomes spatial variability, as long as the ROIs are defined individually. They can be based on anatomical information but it implies a strong prior on anatomy and is rarely automatic. They can also be defined based on functional information (Fedorenko et al., 2010). This requires to add a short but very robust fMRI experiment, such as a localizer, at the beginning of the functional experiment, that will be used to preselect functional ROIs at the individual level. Similarly, Function-based inter-subject registration aims at addressing this issue (Sabuncu et al., 2010) but is difficult to generalize to other functional studies and does not improve registration over the whole cortex.

- The large amounts of information contained in raw 3D volumes require heavy computation, and have to cope with multiple testing, even if the statistical cost is generally reduced by taking into account the image autocorrelation level (i.e. a number of resolution elements (*RESELS*)) (Worsley et al., 1996) or by defining ROIs within which the signal is considered homogeneous.
- The detection results are presented in the common space, as a 3D group activation map, depicting for every voxel the likelihood of being activated across subjects. However, this map provides results at the group level and does not store any individual information: anatomo-functional correlation is only visible in the common space, ignoring inter-subject anatomical variability. The use of this single map, as well as the choice of a single threshold, in order to describe results from a whole cohort of different subjects, are questionable (Coulon et al., 2000; Thirion et al., 2007).

In comparison to this general iconic approach, the framework of structural techniques allows one to deal with representations of data closer to the object under study rather than voxels. Analysis is hence driven on objects of interest and their relationships, which were both formerly extracted from the original images, using anatomical (Rivière et al., 2002; Cachia et al., 2003; Roca et al., 2009), functional (Coulon et al., 2000; Thirion et al., 2007; Lashkari and Golland, 2009), or anatomo-functional (Thirion et al., 2006) criteria. These objects are built as representations of real neuroscientific objects, e.g. sulci (Rivière et al., 2002), white matter fibers (Guevara et al., 2011), gyri (Cachia et al., 2003), functional areas (Coulon et al., 2000; Thirion et al., 2007; Lashkari and Golland, 2009), functional connectivity patterns (Langs et al., 2011) taking into account a number of properties, in terms of geometry, topology, and relationships between each other. For the particular case of functional analysis, Fig. 1 presents an abstract description of these structural approaches. Let one consider ten individual activation maps, each composed of two clusters of interest plus background noise. The upper cluster (circled in red) varies in shape from one map to another but is quite stable spatially. In comparison, the lower cluster (circled in blue) has a much more variable position. Any kind of analysis aims there at finding the most significant, reproducible objects in the group. A classical voxel-by-voxel analysis would there produce a group activation map similar to the one illustrated at the left bottom: the upper cluster can still be spotted but the lower one has clearly suffered from the spatial inter-subject variability, while false positives still remain in the background. In contrast, the results of an ideal structural analysis would actually provide one map per subject with individual representations of the clusters detected at the group level, hence allowing their localization on each subject's anatomy. This preserves the possibility to study anatomo-functional correlation while taking into account the inter-subject variability. The analysis is here focused on objects rather than voxels, a level at which variability is better addressed. Potentially this approach is also more sensitive and robust to spatial normalization limitations (Thirion et al., 2007). To date, this approach has been developed through few particular works (Coulon et al., 2000; Thirion et al., 2007; Lashkari et al., 2010) which address three major problems: first, the choice of a structural representation for functional data, using, e.g. scale-space primal sketches, or watershed parcels; second, the matching method between objects from one subject to another; finally, the inference on putative group activations.

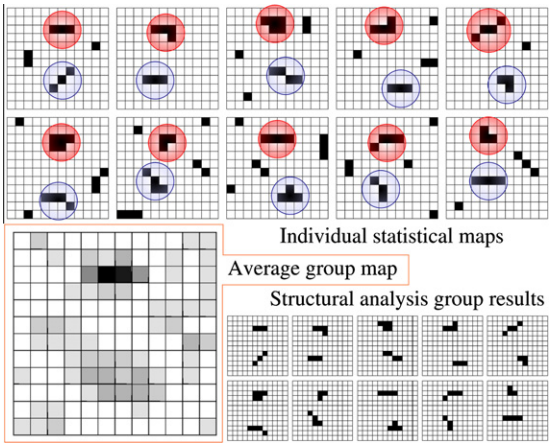

Fig. 1. Schematic comparison between standard and structural analysis – the upper grids represent ten individual maps, each containing two objects of interest (red and blue circles) plus noise. On the bottom left corner is a group map obtained by averaging the individual ones. On the bottom right corner is the optimal group result of a structural analysis, *i.e.* ten individual maps showing two group activations each. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Another aspect of functional data analysis is the volume-based approach that considers the voxels of the whole brain, including those in the white matter, whereas the main sources of the functional signal are located in the cortical ribbon (Mountcastle, 1997; Lauwers et al., 2008; Rakic, 1995). With respect to this, surface-based analysis schemes especially gain interest as they focus on the main location of the cerebral activity. At standard observation scale, the human cortex can be considered as a two-dimensional structure, with a highly convoluted geometry and a high inter-subject variability. Hence, surface-based studies differ from volume-based ones by relying for instance on geodesic distances, computed on this particular manifold, rather than Euclidean distances. This approach is particularly advocated in the scope of cortex-based studies, since the generally employed surface models do not include deep subcortical structures or cerebellum. Considering surfaces instead of volumes also helps to address the question of inter-subject matching in a way more suited to the study of the human cortex, by taking into account its geometry, and it has been shown that considering the cortex as the reference for inter-subject studies is of great interest (Fischl et al., 1999a; Van Essen et al., 1998; Clouchoux et al., 2010). Surface-based functional analysis has however not been much studied in the literature. Most of existing works either perform standard voxel-based univariate (Goebel and Singer, 1999) or multivariate (Formisano et al., 2004) analyses restricted in the cortical ribbon or a parcellation of it (Flandin et al., 2002). Some methods define a general linear model (GLM) using matrices containing surface nodes values instead of voxels intensities (Andrade et al., 2001; Saad et al., 2004), or embedding anatomical information using constrained spatial basis functions (Kiebel and Friston, 2002). Some others realize Bayesian inference of brain activations along the cortical mesh (Vincent et al., 2010). Still, very few of these methods tackle surface-based group studies (Hagler et al., 2006). Regarding this matter, a number of inter-subject surface-based matching

techniques allow to warp a spherical coordinate system onto individual cortical surfaces (Fischl et al., 1999a; Fischl et al., 2004; Lyttelton et al., 2006; Joshi et al., 2010) so as to establish correspondences across the surfaces without any actual 3D deformation and hence opening the way to surface-based group analyses.

In this paper, we propose a fusion of the two frameworks, structural and surface-based, in order to develop a method for group analyses gathering advantages from both sides. The method uses individual surface-based functional maps built from fMRI data. It proposes a multi-scale object-based representation of these maps and performs a surface-based group analysis from these representations. A schematic overview of the process is shown in Fig. 2. The inter-subject matching and the activation detection are performed in a single step by optimization of a Markovian model. Finally, an inference is made on the results by estimating their significance and controlling the risk of type I error.

The contributions of the paper are manifold:

- It proposes a surface-based extension of the volume-based structural method presented in (Coulon et al., 2000). The structural representation of previously computed surface-based individual statistical maps, together with the preprocessing steps and the graph representation of the group of subjects, are presented in Section 2. This section is mostly a reminder of methods previously presented in (Coulon et al., 2000).
- It improves our previous structural method with proper data-driven parameters estimated from real data (Section 3.2), the definition of an inter-subject spatial similarity between objects (Section 3.4), and a regionalized optimization scheme (Section 3.5).
- It adds a non-parametric control of the risk of type I error (Section 3.6).

In Section 4, the method is applied to synthetic and real data. Finally, the last section presents and discusses the results, as well as a comparison with the results of a volume-based structural analysis and a random effect (RFX) analysis. In the conclusion we highlight the possible developments of such structural methods on the cortical surface domain.

2. Structural representation of surface-based fMRI data and graph-based group representation

We present here the data preprocessing and recall the structural representation of our data, already presented in (Coulon et al., 2000) in a volume-based context. We also recall the graph-based representation of the multi-subject data from which group analysis is performed. For more details about structural representations of statistical maps, the reader can refer to (Coulon et al., 2000).

2.1. Preprocessing steps

Once data is acquired, raw T1 volumes are processed using the anatomical pipeline from the BrainVisa package (http://brainvisa.info (Mangin et al., 2004)). This includes, for every subject, image intensity bias correction, brain hemisphere separation and tissue segmentation, and extraction of gray/white matter interface as spherical meshes (Mangin et al., 2004). Usual preprocessing steps are applied to functional volumes, including slice-timing, spatial realignment, registration with the anatomy, such as performed by the SPM package (Friston et al., 1994). Then, the following step consists in converting the 3D functional volumes into 2D cortex-based representations. The method proposed in (Operto et al., 2008) allows the creation of such representations, by projecting the functional data onto the corresponding anatomical mesh, taking consideration of the local cortical geometry. The result is a matrix depicting the signal time course associated with every mesh node. From every of these surface-based functional maps, we compute a surface-based statistical *t*-map by defining a GLM over the surface nodes (in particular, we employed the method described in (Roche et al., 2004), provided with the BrainVisa fMRI toolbox (Favre et al., 2009)). This results in individual surface-based activation maps, for every subject of the group, for a single particular contrast of interest. An example of such a maps is shown in Fig. 3.

2.2. Surface-based structural representation: the scale-space primal sketch

Since the works of Poline and Mazoyer (Poline and Mazoyer, 1994) and Worsley (Worsley et al., 1996), it is established that a

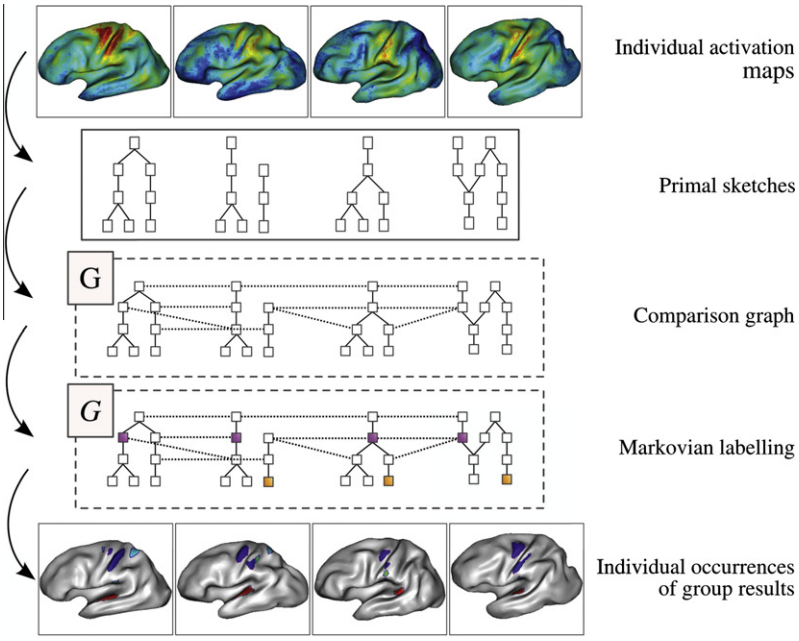

Fig. 2. Overview on the different steps of the structural analysis pipeline.

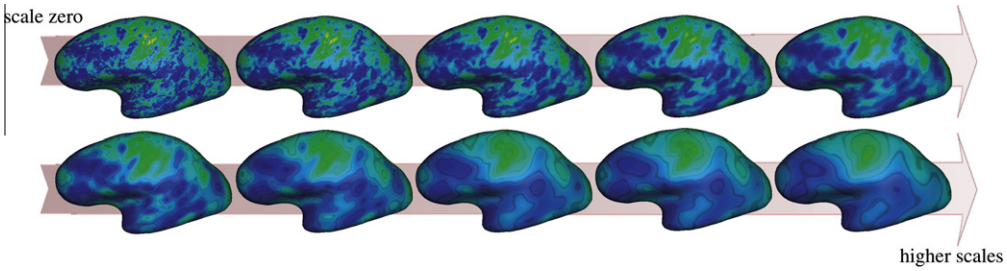

Fig. 3. Scale-space computed from a surface-based statistical student *t* map.

multi-scale approach is beneficial to functional image analysis. In (Coulon et al., 2000) we introduced the use of such representation, the scale-space primal sketch, in the context of group analysis of functional data.

The scale-space primal sketch is a hierarchical object-based representation of data derived from a linear scale-space computed from the original data (Koenderink, 1984; Lindeberg et al., 1999). In our case, a linear scale-space is computed from each individual by solving the heat equation on the surface (Koenderink, 1984):

$$\frac{\partial L(n,\sigma)}{\partial \sigma} = \Delta L(n,\sigma), \tag{1}$$

with the individual map as initial condition. The Laplacian operator can be estimated on a triangular mesh using various methods (López Pérez, 2006; Zosso and Thiran, 2009; Chung and Taylor, 2004; Chung et al., 2008; Chung et al., 2005). Here, we use the finite elements method described in (Chung and Taylor, 2004), which was efficiently used in related works, e.g. (Cachia et al., 2003), and leads to a Laplacian operator computed as a weighted sum of values at neighboring nodes. The linear scale-space results in a series of surface-based maps derived from the original map and showing progressive smoothing of fine details, as the scale parameter increases, in favor of the most robust features, as illustrated in Fig. 3.

From this linear scale-space, the scale-space primal-sketch of the individual map can be computed. The primal sketch is an object-based representation that describes the structure of the map. It has been shown to provide good representation of functional statistical maps (Coulon et al., 2000; Lindeberg et al., 1999). The primal sketch of an individual map is a graph-like hierarchical structure composed of multi-scale objects, scale-space blobs (SSB), linked by relationships called bifurcations, that described their relative behavior through increasing scales. Every local maximum in the original map is associated to a SSB. Each SSB embeds two descriptive features: a measure of their activation level and their spatial support on the cortical surface mesh. The activation level *T* associated to a scale-space blob *ssb* is defined as:

$$T(ssb) = \sum_{\sigma_a \leq \sigma_i < \sigma_d} \frac{\mathcal{T}(\sigma_{i+1}) + \mathcal{T}(\sigma_i)}{2} (\ln \sigma_{i+1} - \ln \sigma_i) \tag{2}$$

with $[\sigma_a, \sigma_d]$ the lifetime interval of the SSB, and $\mathcal{T}$ a function defined as:

$$\mathcal{T}(\sigma_i) = \max_{n \in S_{\sigma_i}} t(n, \sigma_i) \tag{3}$$

with $t(n,\sigma_i)$ as the Student *t*-test value at node *n* on the blob's support S_{σ_i}. The integration of this measure across scales generally favors the objects showing good activation on the original *t*-map and with sufficient long lifetime across scales (Lindeberg, 1993) (see Fig. 4), as illustrated in Fig. 5. For more in-depth details about the primal sketch, the reader should refer to the work of Lindeberg (Lindeberg, 1993; Lindeberg, 1994), and for a specific use in the context of neuroimaging data to (Coulon et al., 2000)

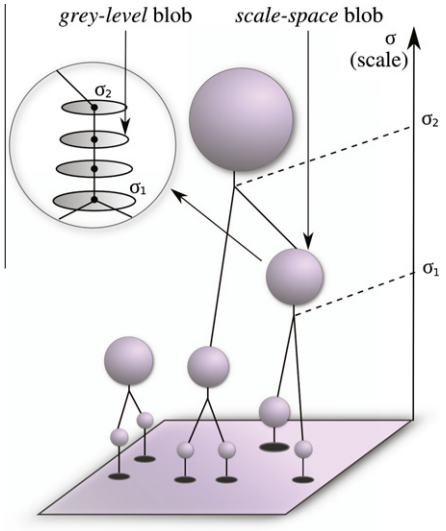

Fig. 4. Symbolic representation of the primal sketch.

2.3. global comparison graph

In order to represent the group of subjects, the set of primal sketches is embedded in a graph $\mathcal{G}$ as defined in (Coulon et al., 2000): nodes are the SSB of primal sketches of all the subjects individual maps, and an edge is created between two nodes if they belong to two different subjects and show some inter-subject similarity (this similarity is detailed further in Section 3.4). We stress that this notion of edge is in no way a process of decision: it does not imply that the inter-subject spatial proximity corresponds to the repetition of a functional activation across two subjects. Instead it shows that two blobs are simply spatially close enough in the common group space. Decision is taken in a further step. Nodes in the graph are valued with the *T* value of the corresponding SSB, and edges are valued with the similarity measure between two objects defined (Section 3.4). The topology of the graph and the values associated to nodes and edges are the key information provided to the group-level detection.

3. Surface-based group analysis

In this section, we define our model and show how labels are assigned to nodes in $\mathcal{G}$. The group analysis process entails assigning a label to every node: a non-null label if considered as an activation, or null label if of non-interest. Nodes (SSB) carrying the same non-null label are representing the same group activation across different subjects. The label field on the graph is represented by *X*, the data represented by *Y*, and maximizing the posterior probability $P(X|Y)$ leads to the optimal labeling that finds objects with high activation measures in individual maps and a

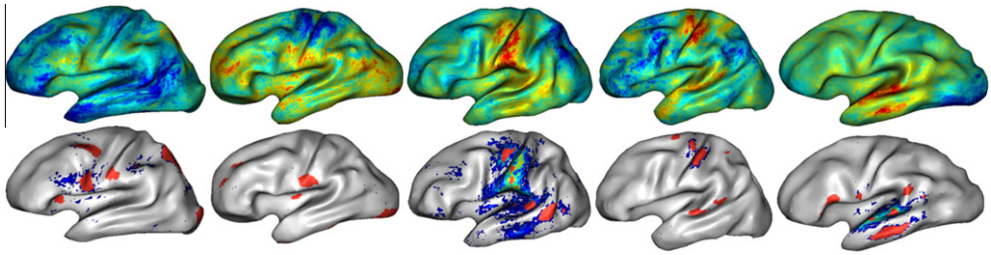

Fig. 5. Blobs presenting the highest $T(.)$ measures for five different t-maps of five different subjects. (*Top*) original statistical maps – (*bottom*) blobs showing the highest measures (*in red*) are superimposed on the thresholded statistical maps ($t > 3.9$). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

good spatial stability across subjects. In (Coulon et al., 2000), we showed that by modeling X with a Markov random field it is possible to maximize the posterior probability by minimizing the energy $U(X|Y)$:

$$U(X|Y) = \sum_{s=1}^{N} V_d(y_s|x_s) + \sum_{c \in C} V_c(X), \tag{4}$$

with $V_c(X)$ the potential functions associated to cliques in $\mathcal{G}$ and $V_d(y_s|x_s)$ the data-driven potential functions associated to sites in $\mathcal{G}$.

Any configuration of labels gets therefore associated to an energy, whose different terms correspond to specific potential functions, which can be either data-driven or contextual. In practice, finding the optimal realization of the label field amounts to pointing out a trade-off between these terms, which minimizes the global energy. We present below the different functions that define our model.

3.1. Potential functions and energy minimization

The potential functions actually define the configurations to be encouraged (associated to negative potentials) or to be penalized (associated to positive potentials).

1. **Data-driven potential**: the data-driven potential $V(y_s|x_s)$ is estimated for every node s of the graph. We recall that y_s is the data associated to node s, i.e. $T(s)$ (see Eq. (2)). According to our model, this potential has different expressions, depending on whether the label carried by s is null or non-null.
   Blobs of non-interest have no prior on their measurements, therefore we define $V_d(y_s|x_s = 0) = 0$.
   In either case, blobs have more chances to represent an activation as their associated activation measure is high.
   Hence we define two values y_1 and y_2, and:

$$\begin{aligned} &\text{If } y_s < y_1, \quad V_d(y_s|x_s \neq 0) = K_d \\ &\text{If } y_s \geqslant y_1, \quad V_d(y_s|x_s \neq 0) = \frac{K_d \cdot \left(\frac{y_2}{2}\right)^2}{\left(\frac{y_2}{2}\right)^2 + (y_s - y_1)^2} \end{aligned} \tag{5}$$

   with $K_d = N_{ps} \cdot k_d$, k_d being the weight given to this data-driven potential in the global energy $U(X|Y)$ and N_{ps} the number of subjects (i.e. primal sketches) in the analysis (Fig. 6).
   Note that the data-driven potential is positive everywhere, which implies that, whatever its measure, a blob requires that other potentials (in particular, the spatial similarity potential) be negative in order to be considered as a putative activation.
2. **Spatial similarity potential**: this potential is related to second-order cliques, i.e. every pair of blobs which are linked in the graph, and encourages these pairs in getting the same non-null label provided they present a good spatial similarity. Hence we define:

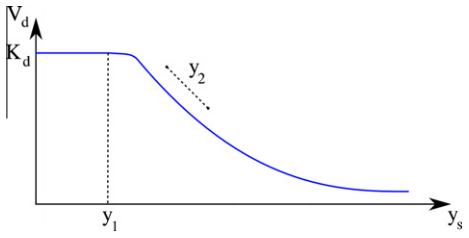

Fig. 6. Datadriven potential function.

$$\begin{aligned} &\text{If } x_{s_1} \neq 0, x_{s_2} \neq 0 \text{ and } x_{s_1} = x_{s_2}, \quad V_{sim} = -K_{sim} \cdot sim(s_1, s_2) \\ &\text{Otherwise}, \quad V_{sim} = 0 \end{aligned} \tag{6}$$

   with x_{s_1} and x_{s_2} being the labels carried by nodes s_1 and s_2 and K_{sim} the weight of this potential in the global energy (Fig. 7). *sim* is the function measuring this similarity between s_1 and s_2: the estimation of the inter-subject similarity on cortical meshes is done using a surface-based coordinate system and will be defined further in Section 3.4.
3. **Lower-scale potential**: the lower-scale potential is estimated for every blob in the graph, i.e. for first-order cliques. This function aims at compensating the trend of high scaled objects to overlap by chance on the cortical surface, compared to those at finer scales (Worsley et al., 1996). Therefore, when the system finds ambiguous cases, where two blobs are in competition to represent the same activation in one same subject, this potential will favor the one at lower scale.
   Blobs carrying null label can be found at any scale level, so that we define $V_e(x_s = 0) = 0$. Otherwise

$$V_e(x_s \neq 0) = K_e . \frac{\sigma_s}{\sigma_{max}} \tag{7}$$

   where $K_e = k_e \cdot N_{ps}$, with k_e being the weight associated to this potential in the global energy and N_{ps} the number of primal sketches in the analysis (Fig. 8). σ_s is called *representation scale* and equals the logarithmic mean scale between the scales at which the blob s appears and disappears, and the

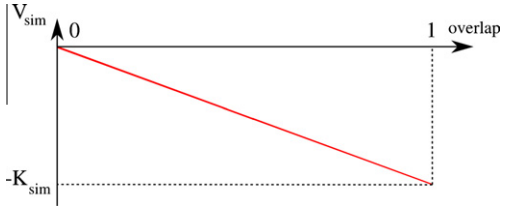

Fig. 7. Similarity potential function.

potential is normalized with respect to σ_{max}, the maximum scale in the PSs.

4. **Intra primal sketch potential**: according to our model, the same non-null label should not be associated with more than one SSB per subject, because one same activation can logically not be identified several times on a single cortex. This intra-primal sketch potential hence controls the number of each non-null label in the graph and penalizes any label which appears multiple times in a same primal sketch. The graph topology then gets more complex since this potential concerns all the SSB of a same primal sketch, which all become neighbors in a single maximal clique. Hence, if every non-null label appears at most once in every primal sketch, we define $V_{ps} = 0$. Otherwise

$$V_{ps} = K_{ps}.\sum_{l\neq 0} n_l \tag{8}$$

where n_l is the number of occurrences of label l in the primal sketch and $K_{ps} = k_{ps}.N_{ps}$ with k_{ps} the weight associated to that potential in the global energy and N_{ps} the number of primal sketches in the analysis.

Among all the previously defined potentials, only the spatial similarity one is negative. This underlines the importance of group-level information in the detection process, and suggests that, in this model, activations are first of all objects that can be found at a similar location in different subjects.

3.2. Parameters of the model

Taking into account the nature of the data (functional blobs) and their domain of representation (cortical surface meshes) is essential. V_d, and in particular parameters y_1 and y_2, depend on $T(.)$ (Eq. (5)). Using the cumulated distribution of $T(.)$, (computed from 50 individual t-maps, 10 subjects and five contrast maps per subject, shown in Fig. 9), we define y_1 as the limit corresponding to the top one percent blobs. A blob with lower measure than y_1 is supposed to have a low chance to make part of a group activation. Besides, the y_2 parameter, which is linked to the data-driven potential decay, is defined such as the 0.5% best ranked blobs get a potential value below 1×10^{-3}. We noted that among the top one percent blobs we could generally find all the blobs representing visually identifiable activated areas in these 50 individual maps (Fig. 5).

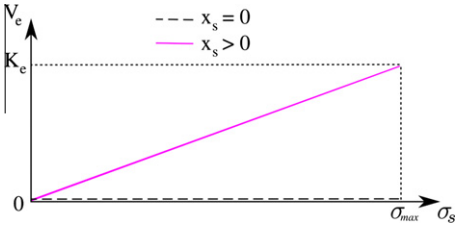

Fig. 8. Lower-scale potential function.

On the other hand, parameters such as potential weights are chosen empirically so as to observe system stability and robustness to small variations:

- $K_{sim} = 1.0$. All other parameters are tuned relatively to this one.
- $K_d = 0.8$. Increasing this parameter would favor the data-driven term and therefore the $T(.)$ value of individual SSB. This means that SSBs with high activation value but not reproducible across subjects would be more likely to be labeled as activations. K_d and K_{sim} are the two main parameters that balance the competition between individual activation level and inter-subject reproducibility.
- $K_{ps} = 4.0$. This parameter has little influence on the results provided it is high enough to prevent any label from appearing multiple times within the same subject.
- $K_e = 0.2$. This parameter has a low value in order to only generate an unbalance in favor of lower scale SSBs in the energy when there is an ambiguity between two SSBs of the same subject that might represent the same activation (as described in Section 3.1).

3.3. Energy minimization

The global energy $U(X|Y)$ has resultingly the following expression:

$$U(X|Y) = \sum_{s\in\mathcal{G}} V_d(y_s|x_s) + \sum_{s\in\mathcal{G}} V_e(x_s) + \sum_{c\in C_{sim}} V_{sim}(c) + \sum_{p\subset\mathcal{G}} V_{ps}(p) \tag{9}$$

composed of the different potentials, previously described, over the sites in the graph $\mathcal{G}$. This energy is clearly non-convex and therefore its minimization is not trivial. To perform it, we use a Gibbs sampler with simulated annealing (Geman and Geman, 1984), coupled with a region-based specialization of the labels over the cortical surface – detailed in Section 3.5 – which facilitates optimization. After convergence, the resulting label field is showed to be optimally satisfying the detection model defined in Section 3. It includes a set of positive labels, each one representing a group activation. We therefore know the occurrence, or the absence of occurrence, of each activation for any subject. Individual representations can then be mapped on each subject's anatomy for localization considerations (see illustration on Fig. 2).

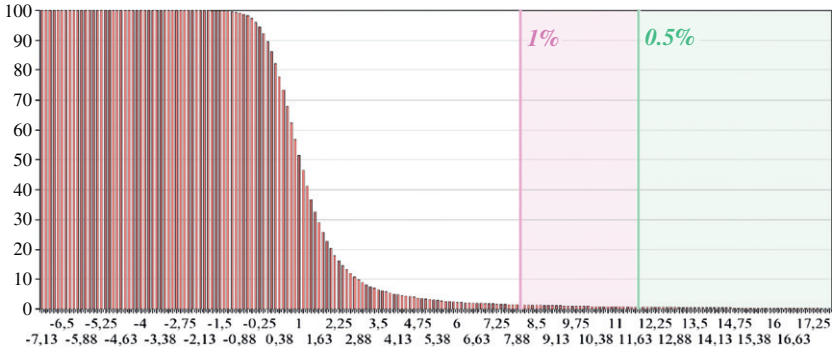

Fig. 9. Cumulated distribution of the $T(.)$ measure associated to SSBs (computed from 50 individual t-maps, 10 subjects and five contrast maps per subject).

3.4. Surface-based inter-subject similarity

The inter-subject matching is performed directly on the cortical surface using an anatomically-constrained surface-based referential defined over the gray/white matter interface of every subject (Clouchoux et al., 2010) that provides an implicit inter-subject matching. As opposed to the classical approach, the method is hence exempt from any iconic normalization step, and can use any other 2D groupwise spherical coordinate-system over the cortex (e.g. (Van Essen, 2005; Fischl et al., 1999b)), although it is dependent on the quality of this system.

Every SSB is represented by a patch on the cortical mesh. A 2D coordinate bounding box is computed for this patch. It is then possible to compute a spatial overlap between bounding boxes from two SSB from distinct subjects. In Eq. (6), we defined the similarity potential as depending on a function $sim(.,.)$ for assessing the spatial proximity of two SSB from two different subjects. This similarity function between two blobs s_1 and s_2 is defined as the overlap ratio between their bounding boxes b_1 and b_2:

$$sim(s_1, s_2) = 2 \times \frac{\mathcal{A}_{\cap}(b_1, b_2)}{\mathcal{A}_{\cup}(b_1, b_2)} \tag{10}$$

This similarity between s_1 and s_2 is therefore based on the ratio between the areas of the intersection $\mathcal{A}_{\cap}(b_1, b_2)$ and the union $\mathcal{A}_{\cup}(b_1, b_2)$ of their bounding boxes b_1 and b_2.

The estimation of this overlap is adimensional, and independent from any metric. As a result, while creating the global comparison graph (see Section 2), an edge is built between two nodes/SSBs only if their spatial overlap is non-null, as well as their scale overlap.

3.5. Region-based label specialization

Gibbs sampling with simulated annealing converges (in an infinite time) to the global minimum, under theoretical conditions which are difficult to comply to in practice (Geman and Geman, 1984). Moreover, the intra-primal sketch potential (Eq. (8)), which controls the numbers of occurrences of non-null labels over each primal-sketch embedded in the global graph, complexifies the topology of $\mathcal{G}$ by introducing high-order cliques and makes the optimization process troublesome. Nevertheless, once a non-null label appears in the graph it is bound to be spatially limited to an area around the specific location it has appeared at, and assigning that same label to any other SSB elsewhere on the cortex becomes irrelevant. We therefore decide to specialize every non-null label to a particular area on the cortex. Each label is associated to a region it can only appear in, and focuses on detecting activations in that region only. Labels and their associated local fields of search hence form a cover of the whole cortical surface. Size of the regions is chosen high enough to overcome the spatial variability of activations across subjects (Fig. 10). For that same purpose, the regions should overlap with their neighbors. Additionally, several different labels are, in practice, associated with every single field of search: in the case of multiple activations located in the same region, several labels would be available in order to identify them distinctly.

This region-based approach leads to rigorously identical results, in terms of detected activations, compared to the original global approach. In contrast, calculation costs are drastically reduced: with a total of 108 different labels, the number of possible states for each site of the graph then varies between 3 and 12 labels, depending on the site localization. Moreover, in a future work, this approach could allow us to introduce extra anatomical information in the process: for instance, by defining regions according to a gyral parcellation (Clouchoux et al., 2006; Cachia et al., 2003; Fischl et al., 2004), the detection results become associated with specific cortical regions and could therefore get anatomically identified.

3.6. Assessing structural results significance

The structural analysis generates a number of labels on the graph, resulting from the prior energy minimization step. Each of these labels corresponds to a group activation detected in a certain number of subjects (equal to the number of primal sketches in which a SSB carries this label).

Standard analyses classically provide an estimation of the relevance of their results. Decisions on voxels are taken under the hypothesis of normality of the data distribution (or for certain methods, only symmetry (Roche et al., 2004; Mériaux et al., 2006)). For these methods relying on a parametric model of the data distribution, results are generally provided with a single threshold on a significance level, which controls the risk of type-I error (false positive). In practice, results with low p-values can be interpreted as statistically significant and are less likely to wrongly reject the null hypothesis.

As opposed to this, the decision process of the structural surface-based analysis maximizes a posterior probability and offers results under the form of connected sets of SSBs in $\mathcal{G}$ with the same non-null label. Some labels can appear in every subject of the group, as those representing stable primary activations, while

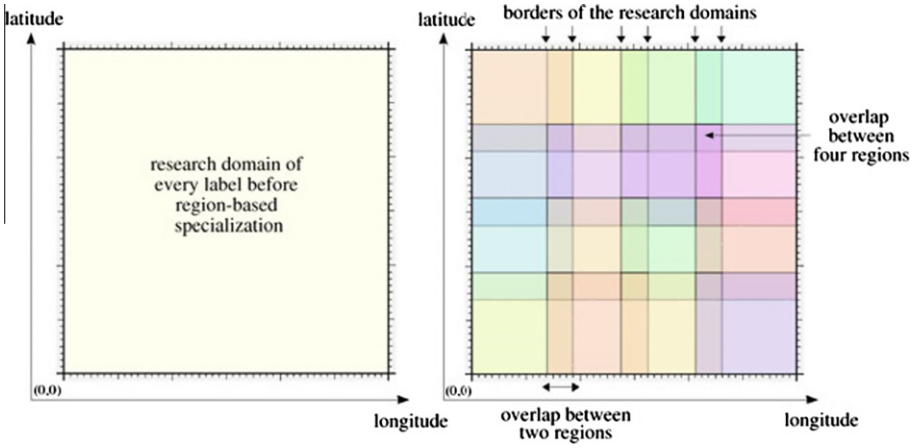

Fig. 10. Region-based specialization of the labels: (*left*) all the labels are in competition over the cortical surface domain – (*right*) the domain is divided into partially-overlapping regions.

others may appear only in subgroups. This representation of results aims at being more extensive, compared to the simple voxel-based approach. Still, completing this representation by assessing the significance of these labels is an essential point in order to infer *post hoc* which of the matched blobs are reliable in terms of false positive control.

Results of such nature satisfy no trivial distribution model such as normality. What confidence level could then be associated to one label appearing in each subject of the group and another in a subgroup of subjects? This question underlines here the particular nature of structural results. We present here a non-parametric approach in order to evaluate the risk of accepting a false positive among these results and to associate a significance index to every detected group activation. An overview of the method is illustrated in Fig. 11.

3.6.1. Surface-based structural results: general nature

Surface-based structural analysis generates a set of labeled blobs. More exactly, every non-null label is associated to one *connected subgraph* (CS) in the global comparison graph. Each of these CSs can be characterized by an average t-value, as well as an average measure of inter-subject spatial overlap. Both of these are standard criteria to assess functional activations significance (Poline et al., 1997). Still, the size of these CSs can vary as it equals the number of subjects in which the activation was identified.

The estimation of the relevance of these resulting labels and their associated CSs first entails assessing the distribution of the population to which these CSs belong.

3.6.2. Population sampling

Let l be a non-null label attributed to n_l SSBs and forming the CS $\mathcal{G}_l \subset \mathcal{G}$. The subgraph $\mathcal{G}_l$ is associated with energy E_l, has a measure of average t-value t_l and an average inter-subject spatial overlap s_l. $\mathcal{G}_l$ was detected on basis of its relevant profile (low energy due to high activation and inter-subject spatial stability) in comparison to the numerous other CSs of same size n_l existing in the global graph. Consequently, with regard to $\mathcal{G}_l$, estimating the distribution of average t-value and spatial overlap over the population of n_l-order CSs is a way to assess its atypical nature. This recalls the typical null-hypothesis tests performed in functional imaging: if we assume that there is no group activation in our graph, what is the probability of observing a value at least as high as the one of our object?

The distribution of n_l-order CSs in the graph is sampled as follows: first, a random blob is drawn, then a second one is chosen among its neighbors in the graph. Until size n_l is reached, additionnal blobs are randomly drawn from the set containing the neighbors of every previously added blob. In case the first blob is chosen in too small a connected component, it is drawn randomly again. For each sample, average t-value and average spatial overlap are computed. The whole process iterates until N samples are drawn (in practice, N = 10,000). Note that samples are drawn with replacement. This sampling algorithm is illustrated by Fig. 12.

3.6.3. Assessing results significance by rank analysis

Once the population of n_l-order CSs is sampled, the distributions of t_l and s_l can be estimated, as illustrated in Fig. 11 or shown for real data in Fig. 17. t_l and s_l are relevant information since they are classical criteria for group detection significance assessment (Poline et al., 1997) and independent of our detection model, unlike the energy. We want to find out if the sampling process is able to draw random components as atypical as those returned by the former analysis step. Since no assumption is made about the distributions, we use a non-parametric rank analysis approach to compare the putative activated blobs sets to the population they

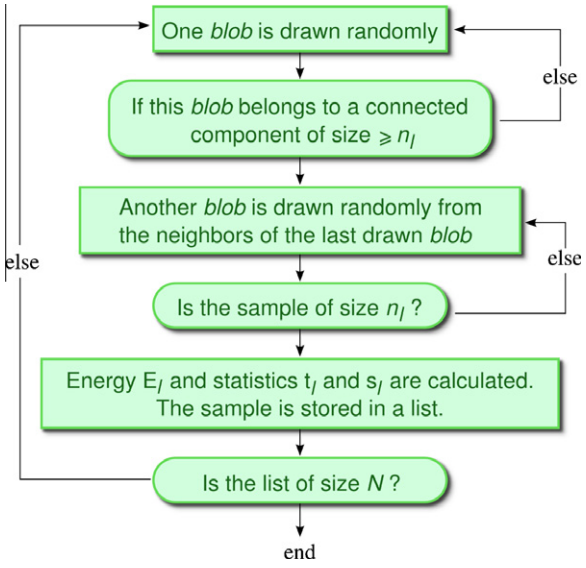

Fig. 12. Flowchart of the sampling algorithm used in the estimation process of results significance.

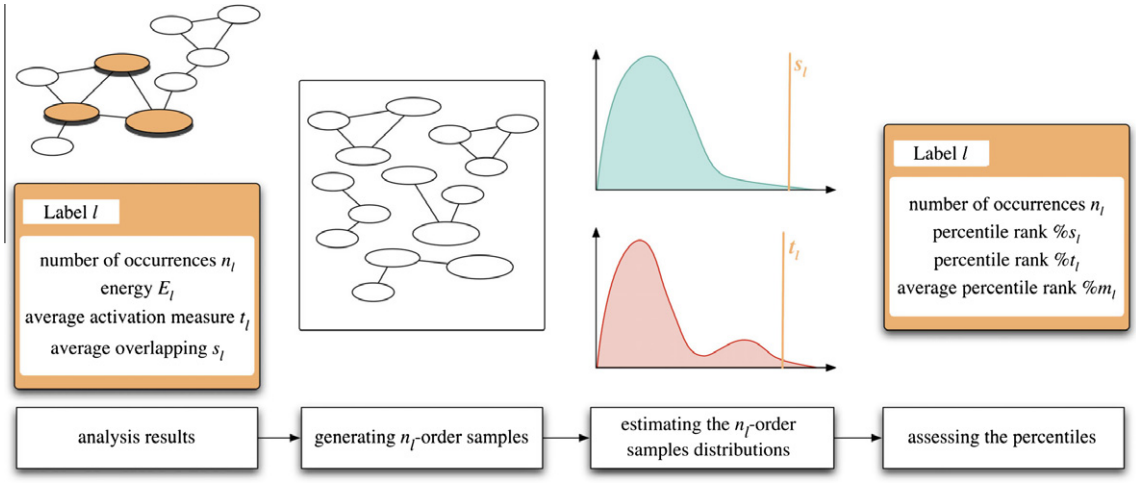

Fig. 11. Assessing structural results significance by rank-percentile analysis: overview on the method.

belong to. Therefore, from the distributions of t_l or s_l we compute the rank-percentiles $\%t_l$ and $\%s_l$ of the CS $\mathcal{G}_l$. This informs about the (hopefully small) proportion of similar samples obtained randomly and therefore about the significance of the putative activation represented by label l.

To summarize, for each group activation l with n_l SSBs, the distributions of t_l and s_l are computed by sampling the population of n_l-order CSs, and two rank-percentiles $\%t_l$ and $\%s_l$ are computed, as illustrated in Fig. 11. We observed that the arithmetic mean $\%\bar{m}_l$ between these two rank-percentiles is a relevant indicator of the significance (see result section), which describes the average risk of type-I error associated to each generated CS.

4. Experiments

fMRI data analysis systematically faces the question of results validation. The lack of a reliable "ground truth" to which the results of existing methods could be compared, is an obstacle. This lack is partly due to the existing inter-subject variability, which is still an open problem. Regarding this, in the following we present here experiments on both simulated and real data in order to illustrate the behavior of the surface-based structural analysis and to assess its performance.

4.1. Experiments on simulated data

4.1.1. Creation of surface-based simulated activation maps

We built synthetic statistical maps on spherical meshes. The use of a common domain such as a sphere allows to proceed to inter-subject comparisons while simulating anatomical variability but without taking into account any bias or error that a normalization step might introduce. For all individual synthetic maps we use the same spherical mesh (22184 nodes). Five nodes were chosen (same nodes for all subjects) as the location of five different activation foci. For each of those nodes, inter-subject variability was then added by selecting a random node around the initial position at a distance $d \approx \delta$, with δ the parameter that controls the spatial variability. Around this new position, all nodes in a 7 mm radius neighborhood were given an activation value of $t_A = 6.0$. Finally, the resulting clusters were smoothed with a Gaussian smoothing ($FWHM \approx 20$ mm).

Noise was added to those individual activation maps, under the form of noise maps built by using at each node the realization of a normal random variable of standard deviation σ_a and followed by a Gaussian smoothing ($FWHM \approx 20$ mm). Using $\sigma_a = 1.27$ ($FWHM \approx 3.0$) one gets simulated activation maps that reproduce the intensity distribution of real ones, as illustrated on Fig. 13. A value of $\sigma_a = 0.05$ will also be used to simulate low noise activation maps, in order to focus on the effects of inter-subject spatial variability (controlled by the parameter δ).

4.1.2. Experiments

Experiments were run to study the performance of the method with respect to variations of the background noise level and inter-subject spatial variability of activation foci. Four datasets were generated, each composed of 20 simulated maps (simulating 20 subjects) and characterized by particular noise level and spatial variability, namely:

- dataset 1: low background noise ($\sigma_a = 0.05$) and no spatial variability ($\delta = 0$ mm),
- dataset 2: standard background noise level ($\sigma_a = 1.27$) and no spatial variability ($\delta = 0$ mm),
- dataset 3: low background noise ($\sigma_a = 0.05$) and high spatial variability ($\delta = 19$ mm),
- dataset 4: standard background noise level ($\sigma_a = 1.27$) and high spatial variability ($\delta = 19$ mm).

The analysis process is then performed on these maps: creation of the primal sketches, structural group analysis, and significance assessment on the resulting labels. Performance of structural analysis is then measured according to its ability to correctly detect and match all five activation foci across these maps. Separately, for each dataset, a Random Effect (RFX) map is computed (Penny and Holmes, 2004). These RFX maps are then thresholded at a standard level ($t > 3.9, p < 5 \times 10^{-2}$) in order to compare the remaining clusters to the structural analysis results. Thresholded RFX maps are shown in (Fig. 14) for various levels of inter-subject anatomical variability, showing a deterioration of results as inter-subject spatial variability increases.

4.2. Experiments on real data

The analysis pipeline was applied to real data, from the NMR database (Poupon et al., 2006), composed of 12 subjects, including T1 anatomical volumes and functional series acquired with a *Localizer* protocol (Pinel et al., 2007). For each subject, a gray/white matter interface mesh was extracted from its T1 volume (Mangin et al., 2004), and functional volumes, corrected by slice timing, were registered with anatomy and projected onto the subject's cortical mesh using (Operto et al., 2008). Every subject has a surface-based coordinate system on their cortical mesh, obtained using (Clouchoux et al., 2010). Finally, from the surface-based functional time series, we create contrast maps (Roche et al., 2004) and individual statistical t maps. From this preprocessed data, we chose to retain a subset of 10 subjects, for which these prior steps (i.e. mesh extraction, parameterization) were conducted satisfactorily.

Left-hemisphere surface-based individual t maps were processed following the structural method presented in this paper: creation of individual primal sketches/definition of the global comparison graph/ labeling process via energy minimization/significance estimation. Finally, the labeling process results in a set of different labels, whose significance can then be estimated.

Analyses were performed on two different contrasts, in order to assess performance on a variety of localizations and activation levels, namely: (*Right–Left Hand*) and (*Audio–Video*). Both are standard functional contrasts and data have been previously analyzed using

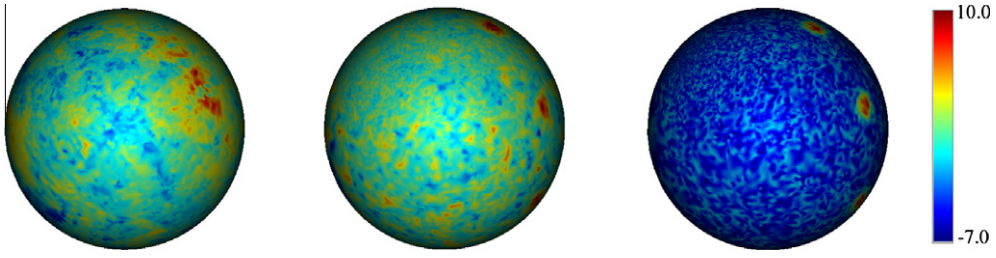

Fig. 13. Comparison between real and simulated maps: (*left*) real statistical t-map presented on spherical mesh – (*center*) simulated map presented on the same mesh ($\sigma_a = 1.27$) – (*right*) same map with low background noise ($\sigma_a = 0.05$).

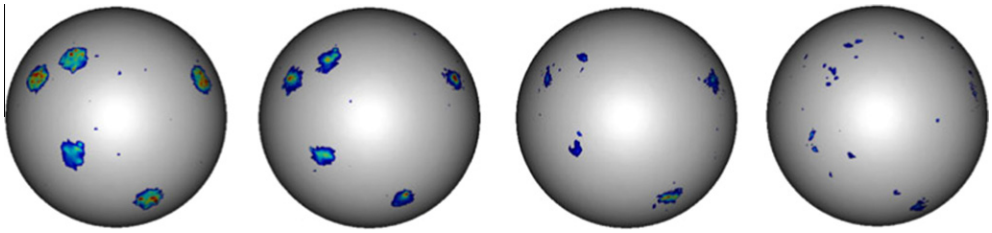

Fig. 14. Thresholded RFX maps ($t > 3.9$, $p < 5 \times 10^{-2}$) obtained for four series of simulated maps with various levels of inter-subject spatial variability and low noise ($\sigma_a = 0.05$): (*from left to right*) $\delta = 0$ mm, $\delta = 5$ mm, $\delta = 10$ mm, $\delta = 19$ mm.

a standard univariate RFX test (Pinel et al., 2007) that we use as a reference to evaluate our results.

Finally, volume-based structural analysis was also performed on the functional data in order to assess the benefit of the surface-based approach. Functional volumes were normalized to the ICBM152 space using the SPM software (Ashburner and Friston, 1999), then individual t-maps were built using the same model than for surface-based data. From each t-map a 3D primal sketch was built (algorithm described in (Coulon et al., 2000)). A group analysis of 3D primal-sketches was then performed using the same potential functions and same weights than for the surface-based model. Inter-subject SSB similarities were computed in the normalized space using Eq. (10) with 3D bounding boxes in the normalized space.

5. Results and discussion

5.1. Simulated data

Results are summarized in Table 1. The positive labels generated by the method depict activations that were detected.

Overall, the generated labels matched the simulated activations:

- datasets 1 and 2 generated 5 positive labels each appearing once per map with no false positives (FP) and no false negatives (FN). Every label's significance was over 99.5%,
- dataset 3 generated 6 positive labels. Three of them correctly appeared 20 times (once per map). One appeared 19 out of 20 expected. Two labels appeared respectively 7 and 12 times: this difference illustrates the disturbance effect related to variability. While actually matching the same single activation, two distinct labels (2 and 6) were created on the basis of their spatial distribution. This series hence counts 2 FN at the subject level, but none at the group level,
- dataset 4 generated 6 positive labels, instead of five expected. Three of them appeared correctly 20 times, i.e. once per map. One of them appeared 19 out of 20 expected. Two labels were generated respectively 18 and 3 times. As for series 3, two labels (2 and 6) were detected and represent a single activation. Among these 21 occurrences, 20 match a true simulated focus and one FP appeared on one individual map (therefore at the subject level but none at the group level). The six labels have percentiles $\%\bar{m}_l$ ranged between 99.46% and 99.99%.

Activations are generally well localized, despite the introduction of background noise or spatial variability. Among the four analyses, we counted a total of 2 FN and 1 FP, when a label wrongly did/did not appear on an *individual map*. Nevertheless, considering the group level, the five simulated activation foci were systematically assigned a positive label and there was no FP; additionally, every positive label was associated to a significance index $\%\bar{m}$ higher than 99.0%. This shows excellent sensitivity and specificity.

Table 1
Results on simulated data with activation label l, mean t-value t_l, mean inter-subject spatial overlap and s_l, corresponding rank-percentiles $\%t_l$ and $\%s_l$, mean rank-percentile $\%\bar{m}_l$ and number of individual occurrences n_l.

l	n_l	t_l	s_l	$\%t_l$	$\%s_l$	$\%\bar{m}_l$
No background noise + no spatial variability						
1	20	6.73	0.8	99.99	99.99	99.99
2	20	6.67	0.93	99.98	99.99	99.99
3	20	6.81	0.91	99.99	99.99	99.99
4	20	6.61	0.84	99.96	99.99	99.98
5	20	6.32	0.82	99.04	99.99	99.52
Standard background noise + no spatial variability						
1	20	6.9	0.78	99.99	99.99	99.99
2	20	7.27	0.90	99.99	99.99	99.99
3	20	7.14	0.83	99.99	99.99	99.99
4	20	7.66	0.68	99.99	99.99	99.99
5	20	6.92	0.7	99.99	99.99	99.99
No background noise + spatial variability ($\delta = 19$ mm)						
1	20	6.86	0.42	99.99	99.99	99.99
2	12	6.89	0.53	99.99	99.99	99.99
3	19	6.48	0.34	99.99	99.99	99.99
4	20	6.47	0.47	99.98	99.99	99.99
5	20	6.60	0.43	99.99	99.99	99.99
6	7	6.72	0.87	99.70	99.99	99.85
Standard background noise level + spatial variability ($\delta = 19$ mm)						
1	20	7.05	0.35	99.99	99.99	99.99
2	18	7.41	0.38	99.99	99.98	99.99
3	20	6.90	0.34	99.99	99.99	99.99
4	20	7.21	0.36	99.99	99.95	99.99
5	19	6.96	0.27	99.99	99.89	99.94
6	3	7.26	0.66	99.81	99.12	99.46

False positive at the individual level are correlated with the introduction of a strong spatial cross-map variability between foci: this creates configurations, where some blobs get spatially separated from the others representing the same activation on the other maps. The system therefore creates no edge between these isolated blobs and the others. The formation of subgroups (several labels to describe a single activation) happens in comparable circumstances, the optimization process seeking for blobs groups with the best trade-off between activation values and spatial stability.

In order to measure the effect of introducing spatial variability on the system, these results can be discussed against those produced by standard RFX analysis on the same data (Fig. 14). On map 1 ($\delta = 0$ mm), the thresholded RFX map presents five regions, corresponding to the five simulated activations. On map 2 ($\delta = 5$ mm), these regions are still visible but smaller. On map 3 ($\delta = 10$ mm), only small clusters survive above the threshold. Finally, on map 4 ($\delta = 19$ mm), the regions corresponding to true activations can hardly be distinguished from false positives. Fig. 15 illustrates this fact, showing four of the 20 activation maps with low noise ($\sigma_a = 0.05$) and high spatial variability ($\delta = 19$ mm). One can see that all activations are detected on all subjects with structural analysis, whereas the RFX map shows poor results. This

strikingly recalls the schematic qualitative difference shown in Fig. 1. All these observations depict the RFX test's poor performances when facing high variability. In comparison, structural analysis shows ability to robustly match spatially variable activated regions.

5.2. Real data

For the (Right–Left Hand) contrast, the analysis generated three positive labels. Labels **1** and **2** appeared 10 times, once per subject. Label **3** appeared in two subjects. Labels **1** and **2** are located in primary sensorimotor area (*M1* and *S1*) with significance indices $\%\bar{m}_1 = 99.95\%$ and $\%\bar{m}_2 = 99.7\%$. Label **3** is located in the frontal lobe and is poorly represented across the group. It has a significance index $\%\bar{m}_3 = 91.99\%$ and therefore should be rejected. Results are presented in Table 2 and in Fig. 16. Fig. 17 illustrates the significance assessment step for these three activations, showing the estimated distributions of t_l and s_l for 10-order and 2-order connected subgraphs.

For the (Audio–Video) contrast, the analysis returned two positive labels, both in the temporal superior area, with significance 97.53% and 96.69%, as presented in Table 2 and Fig. 18.

Figs. 16 and 18 also show that the detected activations are very consistent with the activity observed on individual surface-based maps. For both contrasts, results confirm those produced by a standard RFX analysis (Pinel et al., 2007) on the left hemisphere. Nevertheless the nature of the results carries additional information that a group map cannot provide. In particular, the fact that group results are provided with their individual occurrences allows the link between anatomy and function to be studied more precisely. For instance, it is interesting to notice on Fig. 16 that the M1 activation has been almost always detected at the level of the hand 'knob' (Yousri et al., 1997), while the S1 activation has been systematically detected at a lower position along the central sulcus, although the location of this hand 'knob' varies in the surface-based referential. If measuring the latitude coordinate in the surface-based coordinate system, which locally indicates the position along the central sulcus (Clouchoux et al., 2010), the difference between S1 and M1 is significantly positive ($p = 0.00027$). This level of description cannot be achieved by a group map that provides an average activation localized in an atlas (in fact, in (Pinel et al., 2007) a single sensorimotor activation is detected, as often with volume-based methods, with a maximum in the post-central gyrus which could be interpreted as a purely sensorial activation but is a result of the inter-subject averaging).

By looking at Figs. 16 and 18, one can see that the usual $p < 0.05$ threshold applied at the individual level would not provide very good results. It is particularly true for the (Audio–Video) contrast for which the superior temporal gyrus' activation does not show above the threshold for several subjects (5, 7, 8, and 10), although it is in the results of the structural analysis for those subjects. On the other hand, lowering the threshold would increase the probability to get false positive results. This shows several things: when one wants to look at individual results, the single threshold approach does not stand inter-subject variability; even when looking at individual results, the knowledge added by the whole group can help discriminate false and true positive; and it is particularly interesting for structural analysis **not** to separate the inter-subject matching and the detection steps as both cooperate to provide the final decision.

The same structural analysis performed in a volume-based fashion returned the results presented in Fig. 19. For the (Right–Left Hand) contrast, the primary sensorimotor activation was detected as a single structure. For the (Audio–Video) contrast, only the activation in the Heschl gyrus was detected. This could be the effect of a non-optimal choice of parameters for the 3D analysis. Nevertheless, in both cases, Fig. 19 shows that the delimitation of the detected activations in relation with the cortical surface is poor compared to the surface-based analysis. Clusters often cover both sides of a gyrus, or opposite banks of a sulcus. In particular, the (Right–Left Hand) contrasts confirm the ability of the surface-based approach to separate the sensory and motor components of the primary activity for every single subject, and to properly match each of these two components across subjects. This is gen-

Table 2
Labels returned by the analysis and their associated significance indices.

l	n_l	t_l	s_l	$\%t_l$	$\%s_l$	$\%\bar{m}_l$	E_l
Right motor–left motor contrast:							
1	10	5.44	0.49	99.91	99.98	**99.95**	−17.77
2	10	5.11	0.51	99.41	99.99	**99.7**	−17.44
3	2	4.04	0.7	86.26	97.71	91.99	−0.1

l	Localization
1	Postcentral (Hand)
2	Precentral (Hand)
3	Frontal

l	n_l	t_l	s_l	$\%t_l$	$\%s_l$	$\%\bar{m}_l$	E_l
Audio–video:							
1	10	7.67	0.54	97.28	97.77	**97.53**	−5.47
2	10	7.56	0.44	98.70	94.67	**96.69**	−4.32

l	Localization
1	Heschl
2	Superior temporal

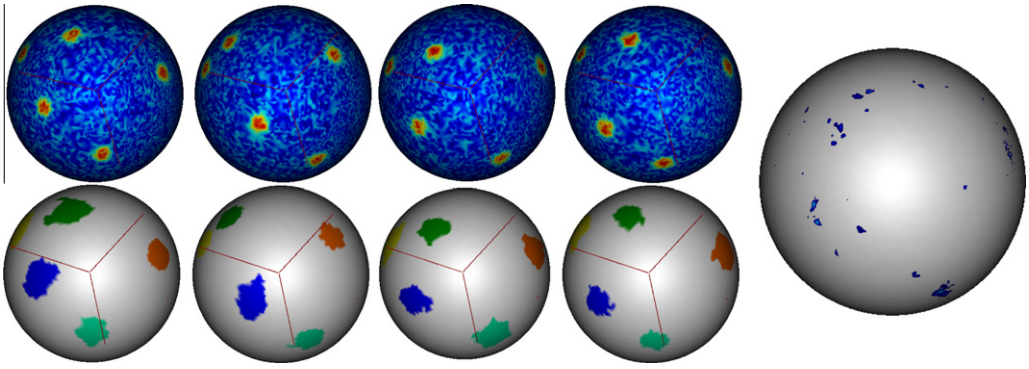

Fig. 15. *(Left top)* four of the 20 simulated maps with low noise ($\delta_a = 0.05$) and high spatial variability ($\delta = 19$ mm) – (*left bottom*) results of structural analysis – (*right*) Corresponding thresholded RFX map ($t > 3.9$, $p < 5 \times 10^{-2}$).

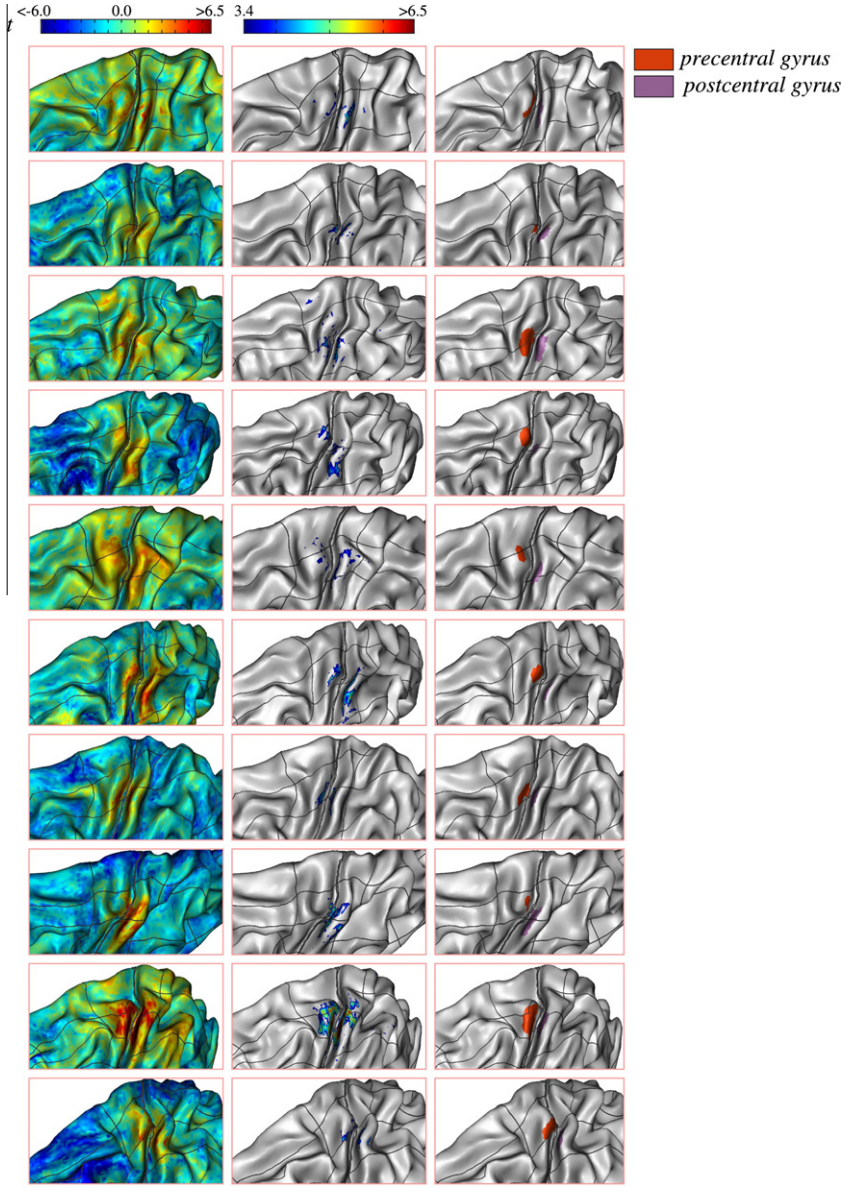

Fig. 16. Results for the *Right–Left Hand* contrast, 10 subjects: (*left*) individual *t* maps – (*center*) thresholded *t* maps ($t > 3.4$) – (*right*) positive labels generated by the structural analysis. Each color corresponds to a different label. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

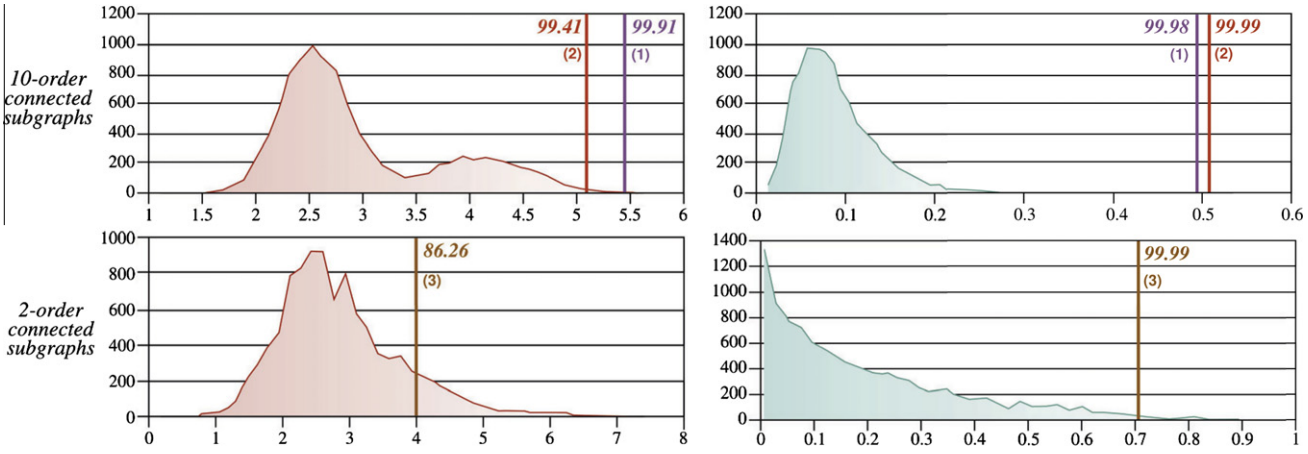

Fig. 17. Rank analysis and significance indices of labels **1**, **2** and 3 for the contrast (*Right–Left Hand*): rank-percentiles (colored bars with corresponding label in brackets) are superimposed on the corresponding distributions – (*left*) t_l distributions – (*right*) s_l distributions. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

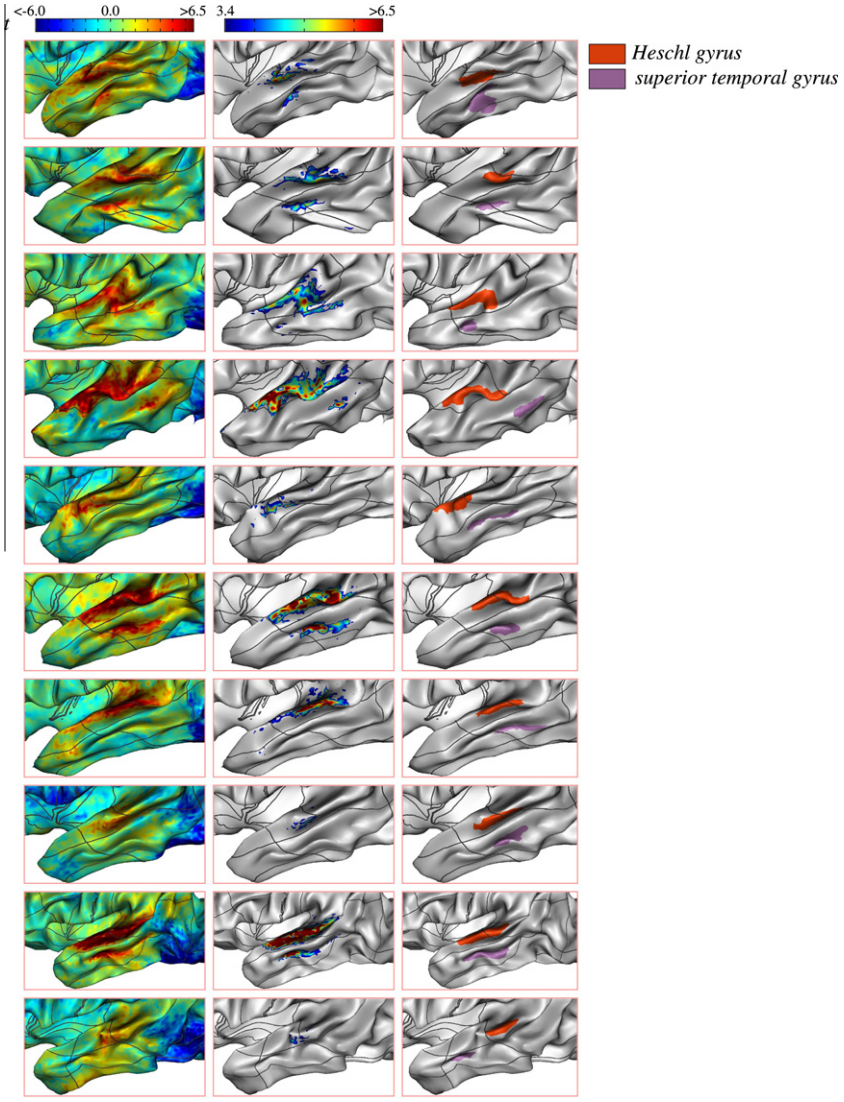

Fig. 18. Results for the *Audio–Video* contrast, 10 subjects: (*left*) individual t maps – (*center*) thresholded t maps ($t > 3.4$) – (*right*) positive labels generated by the structural analysis. Each color corresponds to a different label. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

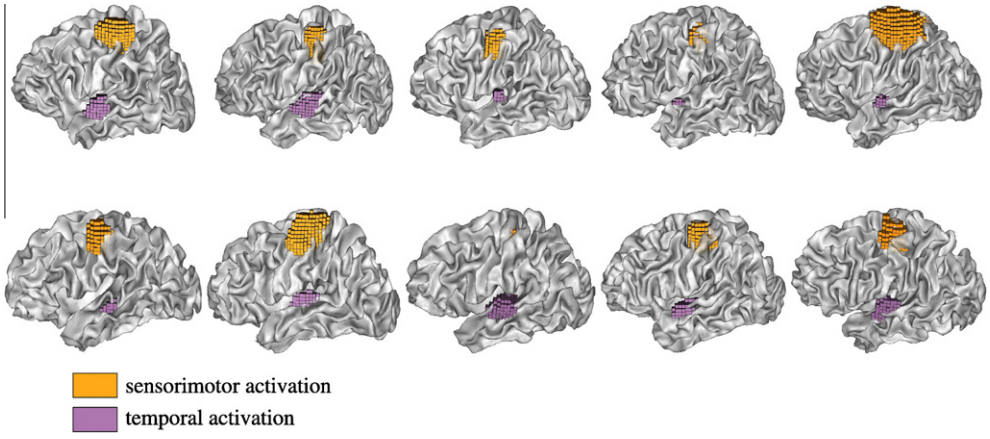

Fig. 19. Volume-based results for both contrasts *Right–Left Hand* (yellow) and *Audio–Video* (purple), 10 subjects. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

erally an essential motivation for using surface-based techniques and is due to the fact that smoothing and structure extraction are performed on the cortical surface, taking into account its geometry, whereas volume-based analyses consider opposite banks of a

sulcus as neighbors. About this specific point, another less controlled factor is the prior projection step, which also divides the cluster by driving the BOLD signal on both faces of the central sulcus. Comparative analyses address this particular question (Tucholka et al., 2009; Anticevic et al., 2008; Desai et al., 2005), which to date still remains open.

6. Conclusion

We presented here a multi-scale surface-based structural method for analyzing fMRI data. Our contributions were the surface-based extension of the volume-based framework presented in (Coulon et al., 2000), a proper data-driven parameter estimation, the definition of an inter-subject spatial similarity between objects, a regionalized optimization scheme, and a non parametric control of the risk of type I error. Results on synthetic data show good sensitivity and specifity and demonstrate an excellent robustness to intersubject spatial variability. Results on real data are good and the advantage of the surface-based approach over the classical volume-based approach was shown via a comparison with the 3D structural analysis presented in (Coulon et al., 2000). Beside the extension of structural analysis to the cortical surface domain, our method provides a significance assessment and posterior control of the risk of type I error.

In this paper, the surface-based detection model was defined in the context of fMRI data analysis. Future research will address the application of the method to other types of data, e.g. connectivity (Roca et al., 2010) or cortical curvature maps (Operto et al., 2010). Dedicated detection models could be implemented for these specific data, in application to matching and identifying group-wise resembling objects such as white matter fiber bundles or sulcal subcomponents. Such analyses might result in pointing out cortical landmarks that could be reversely introduced in the cortical localization system used in the fMRI activation detection process. In conclusion, these future developments take place in an actually wider perspective of establishing a global structural model of the cortex (Mangin, 2005), in which different modalities could possibly be jointly used.

Acknowledgement

O. Coulon is funded by the Agence Nationale de la Recherche (ANR-09-BLAN-0038-01, BrainMorph).

References

Andrade, A., Kherif, F., Mangin, J., Worsley, K., Paradis, A., Simon, O., Dehaene, S., Le Bihan, D., Poline, J.B., 2001. Detection of fMRI activation using cortical surface mapping. Human Brain Mapping 12, 79–93.

Anticevic, A., Dierker, D., Gillespie, S., Repovs, G., Csernansky, J., Essen, D.C.V., Barch, D., 2008. Comparing surface-based and volume-based analyses of functional neuroimaging data in patients with schizophrenia. Neuroimage 41, 835–848.

Ashburner, J., Friston, K., 1999. Nonlinear spatial normalization using basis functions. Human Brain Mapping 7, 254–266.

Beckmann, C., Jenkinson, M., Smith, S., 2003. General multi-level linear modelling for group analysis in fMRI. Neuroimage 20, 1052–1063.

Cachia, A., Mangin, J.F., Rivière, D., Papadopoulos-Orfanos, D., Kherif, F., Bloch, I., Régis, J., 2003. A generic framework for parcellation of the cortical surface into gyri using geodesic Voronoi diagrams. Medical Image Analysis 7, 403–416.

Chung, M., Taylor, J., 2004. Diffusion smoothing on brain surface via finite element method. In: IEEE International Symposium on Biomedical Imaging (ISBI).

Chung, M., Robbins, S., Dalton, K., Davidson, R., Alexander, A., Evans, A., 2005. Cortical thickness analysis in autism with heat kernel smoothing. Neuroimage 25, 1256–1265.

Chung, M., Hartley, R., Dalton, K., Davidson, R., 2008. Encoding cortical surface by spherical harmonics. Statistica Sinica, Special Issue on Statistical Challenges and Advances in Brain Science 18, 1269–1292.

Clouchoux, C., Coulon, O., Anton, J.L., Mangin, J.F., Régis, J., 2006. A new cortical surface parcellation model and its automatic implementation. In: Heidelberg, S.V.B. (Ed.), 9th International Conference on Medical Image Computing and Computer Assisted Intervention, pp. 193–200.

Clouchoux, C., Rivière, D., Mangin, J.F., Operto, G., Régis, J., Coulon, O., 2010. Model-driven parameterization of the cortical surface for localization and inter-subject matching. Neuroimage 50, 552–566.

Coulon, O., Mangin, J.F., Poline, J.B., Zilbovicius, M., Roumenov, D., Samson, Y., Frouin, V., Bloch, I., 2000. Structural group analysis of functional activation maps. Neuroimage, 11.

Crivello, F., Schormann, T., Tzourio-Mazoyer, N., Roland, P., Zilles, K., Mazoyer, B., 2002. Comparison of spatial normalization procedures and their impact on functional maps. Human Brain Mapping 16, 228–250.

Desai, R., Liebenthal, E., Possing, E., Waldron, E., Binder, J., 2005. Volumetric vs. surface-based alignment for localization of auditory cortex activation. Neuroimage 26, 1019–1029.

Favre, L., Fouque, A.L., Vincent, T., Tucholka, A., Keller, M., Operto, G., Thyreau, B., Clouchoux, C., Risser, L., Moreno, A., Geffroy, D., Cointepas, Y., Coulon, O., Ciuciu, P., Thirion, B., Roche, A., 2009. A comprehensive fMRI processing toolbox for BrainVISA. In: 15th Human Brain Mapping Conference.

Fedorenko, E., Hsieh, P.J., Nieto-Castanon, A., Whitfield-Gabrieli, S., Kanwisher, N., 2010. New method for fMRI investigations of language: defining ROIs functionally in individual subjects. Journal of Neurophisiology 104, 1177–1194.

Fischl, B., Sereno, M., Tootell, R., Dale, A., 1999a. Cortical surface-based analysis, II: Inflation, flattening, and a surface-based coordinate system. Neuroimage 9, 195–207.

Fischl, B., Sereno, M., Tootell, R., Dale, A., 1999b. High-resolution intersubject averaging and a coordinate system for the cortical surface. Human Brain Mapping 8, 272–284.

Fischl, B., van der Kouwe, A., Destrieux, C., Halgren, E., Sonne, F., Salat, D., Busa, E., Seidman, L., Goldstein, J., Kennedy, D., Caviness, V., Makris, N., Rosen, B., Dale, A., 2004. Automatically parcellating the human cerebral cortex. Cerebral Cortex 14, 11–22.

Flandin, G., Penny, W., 2007. Bayesian fMRI data analysis with sparse spatial basis function priors. Neuroimage 34, 1108–1125.

Flandin, G., Kherif, F., Pennec, X., Rivière, D., Ayache, N., Poline, J., 2002. Parcellation of brain images with anatomical and functional constraints for fMRI data analysis. In: IEEE International Symposium on Biomedical Imaging, pp. 907–910.

Formisano, E., Esposito, F., Di Salle, F., Goebel, R., 2004. Cortex-based independent component analysis of fMRI time series. Magnetic Resonance Imaging 22, 1493–1504.

Friston, K.J., Holmes, A.P., Worsley, K.J., Poline, J.P., Frith, C.D., Frackowiak, R.S.J., 1994. Statistical parametric maps in functional imaging: a general linear approach. Human Brain Mapping 2, 189–210.

Geman, S., Geman, D., 1984. Stochastic relaxation, Gibbs distributions, and the bayesian restoration of images. IEEE Transactions on Pattern Analysis and Machine Intelligence 6, 721–741.

Gholipour, A., Kehtarnavaz, N., Briggs, R., Devous, M., Gopinath, K., 2007. Brain functional localization: a survey of image registration techniques. IEEE Transactions on Medical Imaging 26, 427–451.

Goebel, R., Singer, W., 1999. Cortical surface-based statistical analysis of functional magnetic resonance imaging data. Neuroimage.

Guevara, P., Poupon, C., Rivière, D., Cointepas, Y., Descoteaux, M., Thirion, B., Mangin, J.F., 2011. Robust clustering of massive tractography datasets. Neuroimage 54, 1975–1993.

Hagler, D.J., Saygin, A.P., Sereno, M.I., 2006. Smoothing and cluster thresholding for cortical surface-based group analysis of fMRI data. Neuroimage, 33.

Joshi, A., Pantazis, D., Li, Q., Damasio, H., Shattuck, D., Toga, A., Leahy, R., 2010. Sulcal set optimization for cortical surface registration. Neuroimage 50, 950–959.

Keller, M., Lavielle, M., Perrot, M., Roche, A., 2009. Anatomically informed bayesian model selection for fMRI group data analysis. In: 12th MICCAI, London, UK

Kiebel, S., Friston, K., 2002. Anatomically informed basis functions in multi-subject studies. Human Brain Mapping 16, 36–46.

Koenderink, J.J., 1984. The structure of images. Biological Cybernetics 50, 363–370.

Langs, G., Lashkari, D., Sweet, A., Tie, Y., Rigolo, L., Golby, A.J., Golland, P., 2011. Learning an atlas of a cognitive process in its functional geometry. In: IPMI 2011: Information Processing in Medical Imaging, LNCS 22, pp. 562–573.

Lashkari, D., Golland, P., 2009. Exploratory fmri analysis without spatial normalization. In: Proceedings of IPMI: Information Processing in Medical Imaging, pp. 398–410.

Lashkari, D., Vul, E., Kanwisher, N., Golland, P., 2010. Discovering structure in the space of fMRI selectivity profiles. Neuroimage 50, 1085–1098.

Lauwers, F., Cassot, F., Lauwers-Cances, V., Puwanarajah, P., Duvernoy, H., 2008. Morphometry of the human cerebral cortex microcirculation: general characteristics and space-related profiles. Neuroimage 39, 936–948.

Lindeberg, T., 1993. Detecting salient blob-like images structures and their scales with a scale-space primal sketch: a method for focus-of-attention. International Journal of Computer Vision 11, 283–318.

Lindeberg, T., 1994. Scale-space theory: a basic tool for analysing structures at different scales. Journal of Applied Statistics 21, 225–270.

Lindeberg, T., Lidberg, P., Roland, P.E., 1999. Analysis of brain activation patterns using a 3D scale-space primal sketch. Human Brain Mapping 7, 166–194.

López Pérez, L., 2006. Régularisation d'images sur des surfaces non-planes. Ph.D. Thesis. INRIA Sophia-Antipolis.

Lyttelton, O., Boucher, M., Robbins, S., Evans, A., 2006. An unbiased iterative group registration template for cortical surface analysis. Neuroimage, 10.

Mangin, J.F., 2005. Une vision structurelle de l'analyse des images cérébrales. Habilitation thesis. Université Paris–Sud. Orsay, France.

Mangin, J.F., Rivière, D., Cachia, A., Duchesnay, E., Cointepas, Y., Papadopoulos-Orfanos, D., Scifo, P., Ochiai, T., Brunelle, F., Régis, J., 2004. A framework to study the cortical folding patterns. Neuroimage 23, 129–138.

Mériaux, S., Roche, A., Dehaene-Lambertz, G., Thirion, B., Poline, J., 2006. Combined permutation test and mixed-effect model for group average analysis in fMRI. Human Brain Mapping 27, 402–410.

Mountcastle, V., 1997. The columnar organization of the neocortex. Brain 120, 701–722.

Operto, G., Bulot, R., Anton, J.L., Coulon, O., 2008. Projection of fMRI data onto the cortical surface using anatomically-informed convolution kernels. Neuroimage 39, 127–135.

Operto, G., Coulon, O., Cachia, A., Rivière, D., Régis, J., Mangin, J.F., 2010. Finding stable sulcal subunits: a group analysis of primal sketches of the cortex mean curvature. In: 16th Human Brain Mapping Conference.

Penny, W., Holmes, A., 2004. Human Brain Function, 2nd ed. Academic Press (Chapter Random Effect Analysis).

Penny, W., Trujillo-Barreto, N., Friston, K., 2005. Bayesian fMRI time series analysis with spatial priors. Neuroimage 24, 350–362.

Pinel, P., Thirion, B., Meriaux, S., Jobert, A., Serres, J., Le Bihan, D., Poline, J., Dehaene, S., 2007. Fast reproducible identification and large-scale databasing of individual functional cognitive networks. BMC Neuroscience 8, 91–108.

Poline, J.B., Mazoyer, B., 1994. Analysis of individual brain activation maps using hierarchical description and multiscale detection. IEEE Transactions on Medical Imaging 13, 702–710.

Poline, J.B., Worsley, K.J., Evans, A.C., Friston, K.J., 1997. Combining spatial extent and peak intensity to test for activations in functional imaging. Neuroimage 5, 83–96.

Poupon, C., Poupon, F., Allirol, L., Mangin, J.F., 2006. Nmr: a free database dedicated to the anatomo-functional study of the human brain connectivity. In: Proceedings of the 12th Annual Meeting of the Organization for Human Brain Mapping.

Rakic, P., 1995. Radial versus tangential migration of neuronal clones in the developing cerebral cortex. Proceedings of the National Academy of Science USA 92, 11323–11327.

Rivière, D., Mangin, J.F., Papadopoulos-Orfanos, D., Martinez, J.M., Frouin, V., Régis, J., 2002. Automatic recognition of cortical sulci of the human brain using a congregation of neural networks. Medical Image Analysis 6, 77–92.

Roca, P., Riviere, D., Poupon, C., Mangin, J.F., 2009. Dwi-based parcellation of the human cortex with a new dimension reduction of the connectivity matrix. In: 15th Human Brain Mapping Conference.

Roca, P., Tucholka, A., Rivière, D., Guevara, P., Poupon, C., Mangin, J., 2010. Inter-subject connectivity-based parcellation of a patch of cerebral cortex. In: MICCAI 2010. LNCS, Vol. 5762. Springer-Verlag.

Roche, A., Lahaye, P.J., Poline, J.B., 2004. Incremental activation detection in fMRI series using kalman filtering. In: Proceedings of 2st Proceedings of IEEE ISBI, Arlington, VA, pp. 376–379.

Saad, Z., Reynolds, R., Cox, R., Argall, B., Japee, S., 2004. SUMA: An interface for surface-based intra- and inter-subject analysis with AFNI. In: IEEE International Symposium on Biomedical Imaging, pp. 1510–1511.

Sabuncu, M., Singer, B., Conroy, B., Bryan, R., Ramadge, P., Haxby, J., 2010. Function-based intersubject alignment of human cortical anatomy. Cerebral Cortex 20, 130–140.

Smith, S., Nichols, T., 2009. Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. Neuroimage 44, 83–98.

Thirion, B., Flandin, G., Pinel, P., Roche, A., Ciuciu, P., Poline, J., 2006. Dealing with the shortcomings of spatial normalization: multi-subject parcellation of fMRI datasets. Human Brain Mapping 27, 678–693.

Thirion, B., Pinel, P., Tucholka, A., Roche, A., Ciuciu, P., Mangin, J.F., Poline, J., 2007. Structural analysis of fMRI data revisited: improving the sensitivity and reliability of fMRI group studies. IEEE Transactions on Medical Imaging 26, 1256–1269.

Tucholka, A., Keller, M., Poline, J.B., Roche, A., Thirion, B., 2009. Surface-based versus volume-based fMRI group analysis: a case study. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI'09, Workshop fMRI Statistical Analysis.

Uylings, H.B.M., Rajkowska, G., Sanz-Arigita, E., Amunts, K., Zilles, K., 2005. Consequences of large interindividual variability for human brain atlases: converging macroscopical imaging and microscopical neuroanatomy. Anat Embryol (Berl) 210, 423–431.

Van Essen, D.C., 2005. A population-average, landmark- and surface-based (PALS) atlas of human cerebral cortex. Neuroimage 28, 635–662.

Van Essen, D., Drury, H., Joshi, S., Miller, M., 1998. Functional and structural mapping of human cerebral cortex: solution are in the surfaces. Proceedings of the National Academy of Science USA 95, 788–795.

Vincent, T., Tucholka, A., Ciuciu, P., 2010. Surface-based joint detection–estimation of brain activity in functional MRI. In: 16th Human Brain Mapping Conference.

Woolrich, M., Jenkinson, M., Brady, J., Smith, S., 2004. Fully Bayesian spatio-temporal modeling of fMRI data. IEEE Transactions on Medical Imaging 23, 213–231.

Worsley, K.J., Marrett, S., Neelin, P., Evans, A.C., 1996. Searching scale space for activation in PET images. Human Brain Mapping 4, 74–90.

Yousri, T., Schmid, U., Alkadhi, H., Schmidt, D., Peraud, A., Buettner, A., Winkler, P., 1997. Localization of the motor hand area to a knob on the precentral gyrus. A new landmark. Brain 120, 141–157.

Zosso, D., Thiran, J.P., 2009. A scale-space of cortical feature maps. IEEE Signal Processing Letters 16, 873–876.