

# Accurate Definition of Brain Regions Position through the Functional Landmark Approach

Bertrand Thirion<sup>1,2</sup>, Gaël Varoquaux<sup>1,2</sup>, and Jean-Baptiste Poline<sup>1,2</sup>

<sup>1</sup> Parietal team, INRIA Saclay-Île-de-France, Saclay, France

`bertrand.thirion@inria.fr`

<http://parietal.saclay.inria.fr/>

<sup>2</sup> CEA, DSV, I<sup>2</sup>BM, Neurospin, Saclay, France

**Abstract.** In many application of functional Magnetic Resonance Imaging (fMRI), including clinical or pharmacological studies, the definition of the location of the functional activity between subjects is crucial. While current acquisition and normalization procedures improve the accuracy of the functional signal localization, it is also important to ensure that functional foci detection yields accurate results, and reflects between-subject variability. Here we introduce a fast functional landmark detection procedure, that explicitly models the spatial variability of activation foci in the observed population. We compare this detection approach to standard statistical maps peak extraction procedures: we show that it yields more accurate results on simulations, and more reproducible results on a large cohort of subjects. These results demonstrate that explicit functional landmark modeling approaches are more effective than standard statistical mapping for brain functional focus detection.

## 1 Introduction

With the improvement of acquisition hardware, such as parallel coils and the advent of high-field MRI scanners, functional Magnetic Resonance Imaging (fMRI) provides increasingly precise information on the spatial structure of brain activity, and is expected to ultimately yield individual mappings with a resolution of 2mm. Functional activity is currently believed to be the most accessible marker to define cortical regions *in vivo* [1]. However, as far as cross-subject activity detection is concerned, the accuracy of activation position is largely determined by anatomical normalization procedures, which are also constantly improving (see [2] for a review). The precise localization of brain foci represents an important information, in particular for pharmacological MRI, where the impact of drugs in several subjects for particular brain targets is assessed [3], or in case studies where one needs to extrapolate the position of identified foci of activity to a new individual to assess lesion impact or for surgery.

However, fMRI has not been considered as a reliable marker of brain activity so far [4,3]. This is related to two problems: one is the lack of accuracy in the location of brain regions, the other one is the lack of sensitivity to detect peaks of interest, that can thus be missed when comparing functional foci across subjects.

In this paper, we investigate the reliability and the reproducibility of several fMRI activation detection procedures; we do not focus on the reproducibility of supra-threshold regions, but on the positions of activity peaks, for which we introduce specific metrics. In particular, we propose a fast functional landmark detection procedure, that directly aims at modeling the spatial distribution of activity peaks. This is in contrast with more standard approaches that first compute statistical maps voxel-by-voxel, then extract the peaks of these maps. This procedure, detailed in Section 2, is conceptually simpler and faster than existing procedures that perform inter-subject activation position modeling [5,6,7]. We present two validation procedures in Section 3: one based on simulations, where the true foci positions are known: we show that, when there is some jitter in the position of functional foci across subjects, the proposed functional landmark approach clearly outperforms all peak detection procedures derived from standard group statistical maps. This gain in detection reliability is characterized by a procedure designed specifically, analogous to the Receiver Operating Characteristic (ROC) curve, that shows the (sensitivity, specificity) trade-off of peak detection techniques. We proceed with an experiment based on a very large dataset, where we show that our functional landmark detection procedure yields more reproducible positions than its alternatives by using a jackknife subsampling procedure. More details and results can be found at the following address: <http://parietal.saclay.inria.fr/research/supplementary-material-miccai-2010>.

## 2 A Spatial Model of Brain Functional Foci

In this paper, we consider that a first-level analysis has been performed in a dataset acquired in  $S$  subjects after normalization in the MNI space, so that individual contrast, variance and statistical maps are available for certain functional contrasts (combination of condition-related effects).

### 2.1 Extracting Peak Positions from Standard Group Analyzes

Most activation detection procedures rely on a univariate modeling procedure that provides group-level statistical maps; these maps are then used to detect supra-threshold regions or activity peaks. The corresponding statistics can be *i*) the classical t-test, that we call random-effects statistic (*rfx*); we also consider the use of data smoothing (12mm Full Width at Half Maximum, FWHM) *srfx*; *iii*) the conjunction statistic, that assesses that the null hypothesis can be rejected in  $k$  subjects over  $S$  [8]; here we consider 2 cases, namely  $k = S/2$  (half conjunction), that we call *cjh*, and *iv*) the case  $k = S$  (full conjunction), called *cjf*. To make inference results comparable across statistics, the significance has been evaluated using the same sign swap procedure [9] in all cases: this procedure performs inference on the positivity of the voxel mean effect in the population, with a built-in correction for multiple comparisons.

In most fMRI studies, a few positions in the MNI space are reported as activation foci by extracting the maps local maxima. These can be ordered by decreasing importance by considering their associated statistical value.

## 2.2 The Functional Landmark (FL) Model

Here, we describe our new procedure to detect functional landmarks. It is related to the approach in [6], with several differences that make it simpler and faster. It consists in 3 steps: a peak extraction procedure in individual data, the specification of a spatial model at the group level, that includes a modeling of false positives; the final specification of the group-level foci of activity, with their position and their statistical significance.

*High-level descriptions of individual maps.* The FL detection procedure takes as input each subject  $t$  statistical map  $(\phi^s)_{s=1..S}$  obtained in the cohort. For a given map  $\phi$  we call a *terminal blob* a set  $a_j$  of connected voxels that are above a saddle point of  $\phi$ , and that contain a unique local maximum of  $\phi$ . We retain only the blobs that contain at least  $s_{min}$  voxels. In the sequel, we denote the extracted terminal blobs  $(a_j^s)_{j=1..J(s), s=1..S}$  where  $s$  is the individual dataset under study and  $j \in [1, J(s)]$  is a region index. The peak position within  $a_j^s$  is denoted  $t_j^s$ , while the average signal within the blob is denoted  $\phi_j^s$ . Next,  $\phi_j^s$  is converted to a probability that the blob is indeed active  $p(H_1(a_j^s)|\phi_s)$ , by learning the distribution of activations through a mixture model of the corresponding map  $\phi_s$  [10]. In this work, we consider only blobs above a threshold that corresponds to a p-value  $p < 0.01$  uncorrected, and use  $s_{min} = 5$ .

*A Dirichlet Process Model for spatial data.* The next step consists in selecting the blobs at similar positions across subjects. Importantly, this is carried out in a probabilistically consistent framework i.e. by refining the p-values  $p(H_i(a_j^s))$ ,  $i \in \{0, 1\}$  based on spatial information:

$$p(H_i(a_j^s)|t_j^s, \phi^s) = \frac{p(t_j^s|H_i(a_j^s), \phi^s)p(H_i(a_j^s)|\phi^s)}{p(t_j^s|\phi^s)}, \tag{1}$$

where  $p(t_j^s|\phi^s) = \sum_{i=0}^1 p(t_j^s|H_i(a_j^s), \phi^s)p(H_i(a_j^s)|\phi^s)$ . We further assume that  $p(t_j^s|H_i(a_j^s), \phi^s) = p(t_j^s|H_i(a_j^s))$ . Next we specify the spatial densities under each alternative  $p(t_j^s|H_i(a_j^s))$ : under  $H_0$ , the selected blobs are distributed uniformly across the brain, so that  $p(t_j^s|H_0(a_j^s)) = 1/|\Omega|$ , where  $\Omega$  is the brain volume; under  $H_1$ , the distribution  $p(t_j^s|H_1(a_j^s))$  is unknown, but is expected to be clustered in some regions of the brain. To model this density, we use a Gaussian Mixture Model with an unspecified number of components implemented through a Dirichlet Process Mixture Model (DPMM) [11]:

$$\begin{aligned} \gamma &\sim DP(\theta, \mathcal{G}), \\ (\mu_j^s, \Lambda_j^s) &\sim \gamma, \forall j[1, J(s)], \forall s \in [1, S] \\ t_j^s &\sim \mathcal{N}(t_j^s; \mu_j^s, \Lambda_j^s), \forall j[1, J(s)], \forall s \in [1, S], \end{aligned} \tag{2}$$

As a base measure  $\mathcal{G}$  we choose a uniform density on the compact brain volume for the mean parameters, and an Inverse Wishart distribution for the covariance

$$\mathcal{G} = p(\mu, \Lambda) = \frac{1}{|\Omega|} W^{-1}(\Lambda; \nu \sigma^2 \mathbb{I}_3, \nu), \tag{3}$$

where  $\sigma$  and  $\nu$  are the hyper-parameters of the model. The fundamental property of Dirichlet processes is the following: if  $X_J = \{(x_j), j = 1..J\}$  is sampled from a DPMM, The predictive density of a new sample is:  $x|X_J, \theta, \mathcal{G} \sim \theta\mathcal{G} + \sum_{j=1}^J \delta_{x_j}$ . Model estimation can thus be performed fairly easily using Gibbs sampling; the algorithm iteratively samples the membership  $z_j^s$  of  $a_j^s$  from

$$\begin{aligned} p(z_j^s = k|t_j^s, z^{-s}, \theta) &\propto p(t_j^s|t^{-s}, z^{-s}, z_j^s = k)p(z_j^s = k|z^{-s}, \theta), \text{ where} \\ p(t_j^s|t^{-s}, z^{-s}, z_j^s = k) &= \begin{cases} \mathcal{N}(t_j^s; \mu_k, \Lambda_k) & \text{if } n_k^{-s} > 0, \\ \frac{1}{|\Omega|} & \text{otherwise,} \end{cases} \quad \text{and} \\ p(z_j^s = k|z^{-s}, \theta) &= \begin{cases} \frac{n_k^{-s}}{\theta + N^{-s}} & \text{if } n_k^{-s} > 0, \\ \frac{\theta}{\theta + N^{-s}} & \text{otherwise,} \end{cases} \end{aligned} \quad (4)$$

where  $z^{-s}$  and  $t^{-s}$  represents the membership and position variables for the blobs in subjects other than  $s$ ;  $n_k^{-s}$  is the number of instances of  $z = k$  for all regions in subjects other than  $s$  and  $N^{-s} = \sum_k n_k^{-s}$ ;  $n_k^{-s} > 0$  amounts to considering that  $k$  is a previously seen component;  $k$  is unvisited otherwise. In our setting, we include an alternative case, in which the blob  $a_j^s$  is a false positive. The sampling scheme is thus:

$$p(H_0(a_j^s)|t_j^s, \phi^s) = \frac{1}{Z} \frac{1}{|\Omega|} p(H_0(a_j^s)|\phi^s), \quad (5)$$

$$p(z_j^s = k|t_j^s, z^{-s}, \theta, \phi^s) = \frac{1}{Z} \begin{cases} \frac{\theta}{\theta + N^{-s}} \frac{1}{|\Omega|} p(H_1(a_j^s)|\phi^s) & \text{if } n_k^{-s} = 0, \\ \frac{n_k^{-s}}{\theta + N^{-s}} \mathcal{N}(t_j^s; \mu_k, \Lambda_k) p(H_1(a_j^s)|\phi^s) & \text{otherwise,} \end{cases} \quad (6)$$

where the normalizing constant  $Z$  is simply obtained by summation. The parameters  $(\mu_k)_{k=1..K}$  and  $(\Lambda_k)_{k=1..K}$  and the number  $K$  of classes are updated at each iteration based on the current values of  $(z_j^s)$ . In this work,  $Q = 1000$  iterations are used, as we found that this was sufficient to yield stable estimates. We choose  $\theta = 0.5$  as usually done in the literature,  $\sigma = 5\text{mm}$  and  $\nu = 10$ . These could eventually be further optimized by cross-validation.

*Drawing explicit clusters from the data.* Final components are estimated by clustering all the blobs that have been assigned to the same model components in at least half of the iterations, hence can be reliably thought to belong to the same group-level component. Our algorithm provides as output the average positions  $(\bar{t}_k)_{k=1..K}$  of the blobs within each cluster. To assess the evidence that the resulting clusters are indeed true landmarks, we estimate a *representativity* statistic, which is simply the expectation that an active region corresponding to that cluster can be found in any subject in the group:  $\eta(k) = \sum_{s=1}^S \left(1 - \prod_{j: z_j^s = k} p(H_0(a_j^s)|\phi_s)\right)$ . It is similar to the *population prevalence* used in [7], but here the probability of each region being active is taken into account.  $\eta(k)$  takes values between 0 and  $S$ . Finally, the computational cost of the whole procedure is proportional to the total number of blobs, not to the image size, which makes it much faster than all the alternatives: : by a factor of 6 with respect to [6], by a factor of (about) 400 with respect to [7].

### 2.3 Validation Procedures

The quality of the model is established by kernel-based statistics that measure the discrepancy between different sets of positions. Given two sets of positions  $\tau = (\tau_1, \dots, \tau_F)$  and  $t = (t_1, \dots, t_D)$ , we can define the following asymmetric statistic, that measures how close the values in  $t$  approximate those in  $\tau$ :

$$\psi(t; \tau) = \sum_{f=1}^F \max_{d=1..D} \exp\left(-\frac{\|\tau_f - t_d\|^2}{2\delta^2}\right) \quad (7)$$

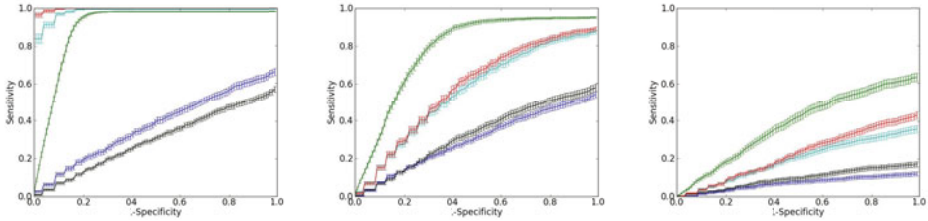
Now, assuming that  $\tau$  represents a ground truth,  $\psi(t; \tau)$  can be taken as a *sensitivity* measure, while  $\psi(\tau; t)$  measures the *specificity* of the detection: these quantities are continuous approximations of the number of true and false detections. As the detected peaks are a monotonous function of some threshold parameter in all the above described procedures, both quantities do monotonously increase with respect to the corresponding threshold, so that it is possible to define (1-specificity, sensitivity) plots that are analogous to ROC curves. In our experiments, we assess the sensitivity, normalized by a factor  $\frac{1}{F}$ , for values of the specificity below 1, so that the (specificity, sensitivity) plots remain with the  $[0, 1] \times [0, 1]$  interval, and area under the curve (AUC) values can be used.

When dealing with real data, no ground truth is available, but we obtain sets of positions  $(t^g) = (t_1^g, \dots, t_{D(g)}^g)$  form different subgroups  $g = 1..G$ ; then we derive a concordance index  $\kappa(t) = \frac{1}{G(G-1)} \sum_{g=1}^G \sum_{h \neq g} \frac{1}{D(g)} \psi(t^g; t^h)$   $\kappa$  is comprised between 0 and 1, a value close to 1 indicating a perfect match between subgroups, while a value close to 0 indicates a poor correspondence. We use  $\delta = 10\text{mm}$  in all our experiments.

## 3 Experiments and Results

### 3.1 Simulated Data

*Data simulation.* We have simulated a multi-subject dataset as a set of  $F = 10$  distant active regions with cone-shaped activation patterns. These regions have been chosen at arbitrary positions in the mask of the standard MNI brain template. Data from 10 subjects are simulated as activation peaks corrupted with spatially correlated noise (FWHM=7mm in each direction) to mimic the spatial structure of real fMRI datasets; the spatial resolution is 3mm. The simulated activation strength is 3 times the noise standard deviation, which corresponds to realistic values. An isotropic spatial jitter of 0, 1.5, 3mm or 6mm standard deviation in each direction is added to the individual position of the foci. 100 simulations are performed for each jitter value, and the average (specificity, sensitivity) characteristics are computed across all simulations. We compare the simulated peak positions in the dataset with those obtained from standard activation voxel-level detection statistics: random effects, with or without smoothing *srfx/rfx*, half *cjh* and full *cjf* conjunctions, and functional landmarks (*fls*).



**Fig. 1.** (Specificity, sensitivity) curves obtained by averaging over 100 draws of simulated data, with an isotropic jitter of 0, 3 or 6mm from left to right. The blue, black, cyan, red and green characteristics correspond to *rfx*, *srfx*, *cjh*, *cjh* and *fls* respectively.

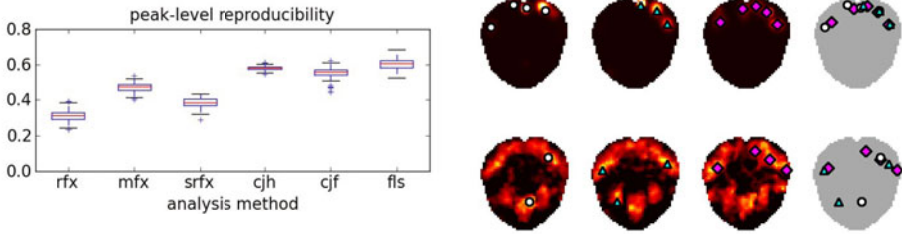
**Table 1.** Area under curves corresponding to the (specificity, sensitivity) curves shown in Fig. 1, based on 100 draws of simulated data

method	rfx	srfx	cjh	cjh	fls
AUC, jitter=0mm	0.372	0.296	0.983	<b>0.986</b>	0.898
AUC, jitter=1.5mm	0.317	0.330	0.761	0.816	<b>0.868</b>
AUC, jitter=3mm	0.297	0.323	0.560	0.577	<b>0.779</b>
AUC, jitter=6mm	0.007	0.09	0.194	0.221	<b>0.380</b>

*Results:* The characteristics are presented in Fig. 1 for three jitter values. The area under curve can be found in Table 1. For non-null jitter the FL approach achieves the best results, while *cjh*/*cjh* perform well only in the absence of jitter. Random effects yield poor performance, and smoothing helps only slightly in presence of jitter. A score of 0.38 (*fls*, jitter=6mm in each direction) means that 3.8 times more true positives than false positives are detected in average.

### 3.2 Experiments on Real Data

*Dataset and procedure.* We use here a dataset of 171 right-handed subjects, that has been described in detail in [12]. The experiment was based on an event-related fMRI paradigm that comprised ten experimental conditions. Among other tasks, the subjects had to perform left or right hand movement, which resulted in maps of the differential motor activity that we study here. FMRI data pre-processing and statistical analysis were performed using the SPM5 software ([www.fil.ucl.ac.uk](http://www.fil.ucl.ac.uk)). In particular, spatial normalization was performed using default parameters (non-rigid, low frequency deformation). We use the concordance measure to compare the reproducibility of peak positions: the different position sets are obtained through jackknife subsampling, by drawing  $G = 12$  random disjoint subgroups of 14 subjects from the population. Here, we use the same significance threshold, chosen as 0.05, corrected for multiple comparisons, in all procedures. Besides the statistics used in the simulation experiment, we introduced mixed-effects statistics *mfx*, as the individual variance maps were available.



**Fig. 2.** (left) Reproducibility index of peak position obtained in the jackknife subsampling procedure for several group analysis techniques. (Right) 2-dimensional example on an axial slice of the scatter of peaks as observed in 3 groups, using the FL approach (top), and the *rfx* statistic (bottom): FLs are clearly more stable across groups.

*Results on the real dataset.* The reproducibility plots are given in Fig. 2 for all the methods tested, based on 100 random population splits into disjoint groups. We observe that the FL positions are more reproducible in average than those of the other statistics, while *rfx* and *srfx* perform worst overall. Similar results were obtained for other functional contrasts (not shown).

## 4 Discussion

Here we have proposed a fast procedure for functional landmark detection and characterization. The approach is conceptually simpler and faster than the alternatives proposed in [5,6,7]; on a standard PC, the procedure takes a couple of minutes. This is crucial to allow cross-validation procedures, either to assess the merit of the method in various simulation experiments, such as those presented here, or merely to optimize the parameters used in the model. Importantly also, all the parameters are continuous and can be interpreted physically, so that their impact on the results can be assessed easily, and their choice can be guided by domain knowledge. Note that all the code used here is freely available in the nipy software suite (<http://neuroimaging.scipy.org/site/index.html>).

Going back to the problem of accurate activation position detection, the FL approach outperforms voxel-based statistics approach, except in simulations with no spatial jitter. This is expected, as voxel-based statistics are straightforwardly optimal in that case, but this is not a realistic situation. The FL method is less sensitive to jitter in the cross-subject activation position than the alternative approaches. This is confirmed by the experiments on real data, that show that the FL approach generally yields more stable results than its voxel-based alternatives. The relatively poor performance of *rfx* statistic can be explained by the fact that it down-weights regions where between-subject variability is large; as these regions coincide with activated regions, this reduces the power and the reproducibility of *rfx*. Smoothing does not help a lot in that respect. Conjunction statistics (especially *cjh*) perform much better, while mixed effects

are in-between. It is important to notice that all the approaches tested here have the same control of false detections:  $p < 0.05$ , corrected for multiple comparisons.

On a more technical note, we can notice that the reproducibility metric  $\kappa$  as well as the  $\psi$  measures on simulated data provide a meaningful and sensitive comparison of methods. One possible shortcoming of  $\kappa$  is that it tends to provide higher values when more detections are performed, simply by chance. However, this effect can easily be controlled by comparing the actual  $\kappa$  values to a value obtained by permutation, thus tabulating the value of  $\kappa$  under the null hypothesis.

*Conclusion.* We have introduced a new functional landmark procedure for inter-subject detection of functional regions, that explicitly models the statistical distribution of activity peaks, and provides more reliable foci positions than standard group statistical maps. Our detection procedure is fast and its parameters are easy to calibrate; it can thus be used easily in practical settings. As future work, we plan to extend this approach to cortical maps in order to cumulate the advantage of functional landmarks and accurate brain surface modeling.

## References

1. Brett, M., Johnsrude, I.S., Owen, A.M.: The problem of functional localization in the human brain. *Nat. Rev. Neurosci.* 3(3), 243–249 (2002)
2. Klein, A., Andersson, J., et al.: Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *Neuroimage* 46(3), 786–802 (2009)
3. Mitsis, G.D., Iannetti, G.D., Smart, T.S., Tracey, I., Wise, R.G.: Regions of interest analysis in pharmacological fMRI: how do the definition criteria influence the inferred result? *Neuroimage* 40(1), 121–132 (2008)
4. Duncan, K.J., Pattamadilok, C., Knierim, I., Devlin, J.T.: Consistency and variability in functional localisers. *Neuroimage* 46(4), 1018–1026 (2009)
5. Coulon, O., Mangin, J.F., et al.: Structural group analysis of functional activation maps. *Neuroimage* 11(6 pt 1), 767–782 (2000)
6. Thirion, B., Tucholka, A., et al.: High level group analysis of fMRI data based on dirichlet process mixture models. *Inf. Process Med. Imaging* 20, 482–494 (2007)
7. Xu, L., Johnson, T.D., Nichols, T.E., Nee, D.E.: Modeling inter-subject variability in fMRI activation location: a bayesian hierarchical spatial model. *Biometrics* 65(4), 1041–1051 (2009)
8. Heller, R., Golland, Y., et al.: Conjunction group analysis: an alternative to mixed/random effect analysis. *Neuroimage* 37(4), 1178–1185 (2007)
9. Mériaux, S., Roche, A., Dehaene-Lambertz, G., Thirion, B., Poline, J.B.: Combined permutation test and mixed-effect model for group average analysis in fMRI. *Hum. Brain Mapp.* 27(5), 402–410 (2006)
10. Everitt, B.S., Bullmore, E.T.: Mixture model mapping of the brain activation in functional magnetic resonance images. *Hum. Brain Mapp.* 7(1), 1–14 (1999)
11. Neal, R.M.: Markov chain sampling methods for dirichlet process mixture models. *Journal of Computational and Graphical Statistics* 9(2), 249–265 (2000)
12. Pinel, P., Thirion, B., et al.: Fast reproducible identification and large-scale databasing of individual functional cognitive networks. *BMC Neurosci.* 8, 91 (2007)