Group-level impacts of within- and between-subject hemodynamic variability in fMRI

Solveig Badillo, Thomas Vincent, Philippe Ciuciu

Group-level impacts of within- and between-subject hemodynamic variability in fMRI

Solveig Badillo[a,b], Thomas Vincent[a,c], and Philippe Ciuciu[a,b,*]

[a]*CEA/DSV/I²BM/NeuroSpin, CEA Saclay, Bât. 145, Point Courrier 156, 91191 Gif-sur-Yvette cedex, France*
[b]*Parietal team, INRIA Saclay Ile-de-France, NeuroSpin, Bât 145, 91191 Gif-sur-Yvette cedex, France*
[c]*Mistis team, Inria Grenoble Rhône-Alpes, 655 avenue de l'Europe, Montbonnot, 38334 Saint Ismier Cedex, France*

Abstract

Inter-subject fMRI analyses have specific issues regarding the reliability of the results concerning both the detection of brain activation patterns and the estimation of the underlying dynamics. Among these issues lies the variability of the hemodynamic response function (HRF), that is usually accounted for using functional basis sets in the general linear model context. Here, we use the Joint Detection-Estimation approach (JDE) [34, 49] which combines regional nonparametric HRF inference with spatially adaptive regularization of activation clusters to avoid global smoothing of fMRI images. We show that the JDE-based inference brings a significant improvement in statistical sensitivity for detecting evoked activity in parietal regions. In contrast, the canonical HRF associated with spatially adaptive regularization is more sensitive in other regions, such as motor cortex. This different regional behavior is shown to reflect a larger discrepancy of HRF with the canonical model. By varying parallel imaging acceleration factor, SNR-specific region-based hemodynamic parameters (activation delay and duration) were extracted from the JDE inference. Complementary analyses highlighted their significant departure from the canonical parameters and the strongest between-subject variability that occurs in the parietal region, irrespective of the SNR value. Finally, statistical evidence that the fluctuation of the HRF shape is responsible for the significant change in activation detection performance is demonstrated using paired t-tests between hemodynamic parameters inferred by GLM and JDE.

Keywords: Hemodynamics, FIR modeling, Bayesian inference, JDE, group-level analysis, parallel imaging

1. Introduction

Since its early discovery twenty years ago [40], Blood Oxygenation Level Dependent (BOLD) effect has been largely studied and fMRI has become a popular neu-

---

*Corresponding author (philippe.ciuciu@cea.fr).

roimaging modality. However, the relationship between the BOLD signal and the neural activity is imperfectly known [23, 32, 31]. Drawing conclusions about unobserved neural activity from the BOLD signal thus remains challenging and requires model-based approaches, where the main concerns are first the detection of evoked brain activity and second the characterization of the human BOLD response to a brief and unique stimulus, named the Hemodynamic Response Function (HRF) [24]. The cognitive interest for HRF inference lies in its potential to recover the magnitude, latency and duration of neural activity through a nontrivial mapping of such parameters onto the hemodynamic ones (eg, peak value, time-to-peak and full width at half maximum, respectively), which can be extracted from the HRF time course. This allows neuroscientists to interpret fMRI results in terms of brain function localization. However, such results can be complicated to interpret. The specification of an accurate HRF model remains indeed difficult as the fluctuations of the BOLD response may originate from different neural activity sources and their link is not necessarily linear [29]. Moreover, the estimation of hemodynamic parameters from the HRF shape may be noisy and biased by several factors (prior specification of a constrained model, unbalanced experimental paradigm, etc.).

In the present paper, both accurate activation detection and robust HRF estimation are addressed. Detecting evoked brain activity is usually performed in the General Linear Model (GLM) context and postulates an invariant or canonical form for the HRF throughout the brain. However, several contributions have exhibited empirically derived HRFs that differ significantly from "the canonical model" [2, 36, 38, 22], and demonstrated that using such HRFs may improve the statistical sensitivity. Thus, GLM extensions have been developed to account for spatial fluctuations of the HRF shape by considering temporal and dispersion derivatives of the canonical model [17, 24]. In some cases, this approach lacks flexibility to take large deviations from canonical features into account. To overcome this limitation, nonparametric Finite Impulse Response (FIR) models have been proposed in the GLM framework. These models do not assume any functional form for the HRF and amounts to estimating a large number of parameters to identify its characteristics properties [19, 25]. This FIR-based modeling thus makes it possible to perform activation detection and HRF estimation. However, the achieved gain has a direct counterpart, namely a cost in terms of statistical sensitivity since it induces a loss of degrees of freedom [34]. Moreover, FIR-based inference may lack robustness and often yields disrupted shapes that are difficult to interpret, due to over-fitting problems. Hence, to cope with this issue, temporal [21, 11, 35, 5] but also spatial regularization [20, 52, 10, 33, 41, 16, 27, 13, 6, 51] have been introduced in models to improve estimation accuracy.

Such estimation methods can be categorized according to three main criteria: *i.)* the HRF model they rely on (parametric [13, 29] or FIR [11, 10, 1, 5, 34, 51]), *ii.)* the type of inference that they perform (univariate [24, 13, 1] vs. multivariate [51, 49]) and *iii.)* the adaptability degree of regularization they introduce, namely the supervised [29] vs. unsupervised [13, 51, 49] setting of hyper-parameters.

Besides, robust HRF estimation only makes sense in voxels or regions eliciting evoked activity in response to a given experimental condition. To derive relevant interpretation of the estimated HRF time courses, the best strategy is thus to focus on such voxels or regions. This calls for methods that perform detection and estimation

jointly: the main challenge is to optimally combine the above mentioned features at the subject-level, ie. estimate the HRF using a multivariate and sparse generative model with unsupervised spatio-temporal regularization. Part of the answer can be found in [29], where an optimal strategy for HRF inference is achieved either by a parametric modeling through Inverse Logit basis functions or by smooth FIR modeling, the former remaining more robust to model misspecification. Nonetheless, this study only considered univariate inference. Here, we focus instead on multivariate inference where detection and estimation are combined in a Joint Detection-Estimation (JDE) Bayesian framework [49, 8, 9]. JDE relies on a regional bilinear model of the BOLD response involving FIR modeling for the HRF and constraining the parameters by physiological temporal and spatial priors. Spatial correlation is modeled through a hidden Markov model, the stiffness of which is spatially adaptive and automatically tuned.

This paper presents two main contributions. First, we compare group-level activation detection analyses relying on three different intra-subject methods: the intra-subject JDE analysis, a standard GLM-based fitting procedure on the one hand or an intermediate Spatially Adaptive GLM (SAGLM) method. The SAGLM approach is a simplification of the JDE approach where the HRF is set to its canonical shape. This comparison is performed first on artificial data and second on a population of 15 healthy subjects using different fMRI data sets corresponding to varying SNRs so as to assess the reproducibility of our conclusions on the same population. This comparison aims at identifying brain regions where HRF estimation has a positive impact on brain activity detection. Second, we extract hemodynamic parameters, in particular the Time-To-Peak (TTP) and Full Width at Half Maximum (FWHM), from the JDE and GLM-based HRF estimates and investigate for both which factor dominates the others between regions, subjects and SNR changes through repeated measures analyses of variance. A comparable group-level hemodynamic study has been performed in [45], but their hemodynamic model was based upon a function basis that offers less flexibility than FIR modeling. In this respect, our study extends previous works [29, 45]. We compare JDE- and GLM-based hemodynamic parameter estimates: amongst the regions we tested, the one eliciting the most significant discrepancy to the canonical HRF and the strongest variability across individuals is located in the parietal cortex.

The rest of this paper is organized as follows. In Section 2, we summarize the intra-subject GLM, SAGLM and JDE inference schemes. Simulation results are discussed in Section 3. The fMRI data sets and our group-level analysis pipeline are presented in Section 4. The results of our comparative study on real data sets are given in Section 5. Finally, the pros and cons of the JDE framework are discussed in Section 6 in the light of our results and future extensions are envisaged. Conclusions are drawn in Section 7.

2. Intra-subject analysis techniques in fMRI

2.1. Standard GLM-based analysis

The GLM involved in this study is a massively univariate approach where the effects are assumed independent in space. Instead of modeling spatial correlation between neighboring voxels, the fMRI time-series are smoothed with a fixed Gaussian kernel that introduces a spatially invariant amount of correlation. In any voxel j the

3

GLM allows for some hemodynamic fluctuations by modeling the canonical HRF $\boldsymbol{h}_\mathrm{c}$ and its first order derivative $\boldsymbol{h}'_c$ as proposed in [17] (see Fig. 1):

$$\forall j \in [\![1; J]\!], \quad \boldsymbol{y}_j = \sum_{m=1}^{M} \boldsymbol{X}^m \big(\beta_j^m \boldsymbol{h}_\mathrm{c} + \widetilde{\beta}_j^m \boldsymbol{h}'_\mathrm{c}\big) + \boldsymbol{P}\boldsymbol{\ell}_j + \boldsymbol{b}_j, \tag{1}$$

where β_j^m and $\widetilde{\beta}_j^m$ are the unknown effects associated with the mth stimulus-induced regressors constructed with $\boldsymbol{h}_c$ and $\boldsymbol{h}'_c$, respectively. J corresponds to the total number of voxels considered in the functional mask of the brain and M is the number of experimental conditions involved in the experimental paradigm. Let us also denote $\boldsymbol{x}_{t_n}^m = \big(x_{t_n - d\Delta t}^m\big)_{0 \leqslant d \leqslant D}^\mathrm{t}$ such that matrix $\boldsymbol{X}^m = \big[\boldsymbol{x}_{t_1}^m, \ldots, \boldsymbol{x}_{t_N}^m\big]^\mathrm{t}$ is the $N \times (D+1)$ occurrence matrix consisting of the lagged stimulus covariates. Matrix $\boldsymbol{P} = \big[\boldsymbol{p}_1, \ldots, \boldsymbol{p}_G\big]$ of size $N \times G$ comprises the values at times t_n of an orthonormal basis (*i.e.*, $\boldsymbol{P}^\mathrm{t} \boldsymbol{P} = \boldsymbol{I}_G$) consisting of G functions $\boldsymbol{p}_g = (p_{g,t_n})_{t_1 \leqslant t_n \leqslant t_N}^\mathrm{t}$ that take a potential drift and any other nuisance effect (eg. motion parameters) into account. Vector $\boldsymbol{\ell}_j = (l_{g,j})_{1 \leqslant g \leqslant G}^\mathrm{t}$ contains the corresponding unknown regression coefficients in voxel j. Vector $\boldsymbol{b}_j = (b_{j,t_n})^\mathrm{t}$ defines the noise or error term in j which is assumed to be an $AR(1)$ Gaussian process.
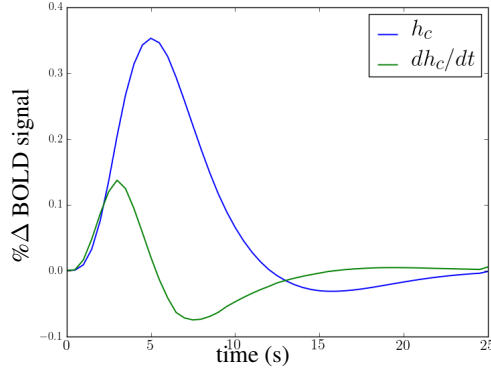

Figure 1: The canonical HRF time course and its first temporal derivative. Canonical parameters are $\mathrm{TTP}(h_c) = 5$s and $\mathrm{FWHM}(h_c) = 5.39$s.

The formulation in Eq. (1) highlights the construction of the stimulus-induced regressor as well as the contribution of other components making the introduction in the next section of the JDE modeling easier to link to GLM formulation. In practice, BOLD effects and noise statistics are respectively estimated in the maximum and restricted maximum likelihood sense [18]. Here, they have been computed using the SPM5 software [1].

2.2. *Joint Detection-Estimation framework*

Spatially varying HRF models have been used in the JDE framework [33, 34, 49] in order to keep a *single* but *spatially adaptive* regressor per experimental condition,

---

[1]Wellcome Department of Imaging Neuroscience, UK (www.fil.ion.ucl.ac.uk)

and thus enable direct statistical comparison through one-dimensional contrasts, making cognitive interpretations more straightforward. In this setting, a local estimation of the HRF is performed at a regional scale, whereas the detection task remains voxel-specific. A critical issue then consists of exhibiting a functionally homogeneous clustering or parcellation of the fMRI data sets over the whole brain to which the JDE inference can be applied.

2.2.1. Within-subject parcellation

The functional mask of each subject's brain is a priori divided in Γ functionally homogeneous *parcels* using a spatially constrained hierarchical clustering of functional features extracted via a classical GLM analysis [46]. Technically speaking, several parcellation or clustering procedures can be used (connected k-means, Ward algorithm) as long as functional homogeneity measured over the fMRI times series through information criteria is guaranteed within each parcel [46]. Hence, the assumption of a shape-invariant HRF for each parcel is tenable.

The larger the number of parcels, the stronger the level of within-parcel homogeneity, but potentially with a lower signal-to-noise ratio. This means that if the number of voxels is too small in a given parcel, the corresponding HRF estimation may become inaccurate, specifically in regions where no voxel elicits a specific response to any experimental condition. To objectively choose an adequate number of parcels, theoretical information criteria have been investigated in [47]: converging evidence for $\Gamma \approx 500$ at a spatial resolution of $3 \times 3 \times 3 \text{ mm}^3$ has been shown for a whole brain analysis leading to typical parcel sizes around a few hundreds voxels ($\approx 2.7 \text{cm}^3$). Fig 2 shows the parcellations in the same slice for three subjects. The brain was divided in 500 parcels, and the subject-specific histograms of parcel size are shown in Fig. 2 to illustrate that the distributions are quite stable whereas the underlying geometry may vary substantially across subjects in order to adapt to local subject-specific fluctuations.

2.2.2. Parcel-based modeling of the BOLD signal

Here, the parcel-based model of the BOLD signal introduced in [33, 34] is adopted. As illustrated in Fig. 3[Left], the input BOLD data set is split into Γ parcels $(\mathcal{P}_\gamma)_{\gamma=1:\Gamma}$ and JDE analysis is then performed independently on each $\mathcal{P}_\gamma = \{j\}_{j \in [\![1;J_\gamma]\!]}$ (J_γ corresponds to the total number of voxels contained in the parcel). We stress the fact that the corresponding fMRI signals are not spatially smoothed prior to the fitting procedure. A time-invariant model, illustrated in Fig. 3[Right], characterizes each parcel $\mathcal{P}_\gamma$ by a single HRF shape and accounts for voxel-dependent and stimulus-related fluctuations of the BOLD signal magnitude. This bilinear parcel-based generative model of the BOLD signal is thus generally defined by:

$$\forall j \in \mathcal{P}_\gamma, \quad \boldsymbol{y}_j = \sum_{m=1}^{M} a_j^m \boldsymbol{X}^m \boldsymbol{h}_\gamma + \boldsymbol{P}\boldsymbol{\ell}_j + \boldsymbol{b}_j. \tag{2}$$

where $\boldsymbol{y}_j$, $\boldsymbol{P}$, $\boldsymbol{\ell}_j$ and $\boldsymbol{b}_j$ match the variables introduced in Subsection 2.1. Here a_j^m stands for the Neural Response Level (NRL) in voxel j for condition m, and the whole set of NRLs is denoted $\boldsymbol{A} = \{\boldsymbol{a}^m, m = 1 \ldots M\}$ where $\boldsymbol{a}^m = \{a_j^m, j \in \mathcal{P}_\gamma\}$.
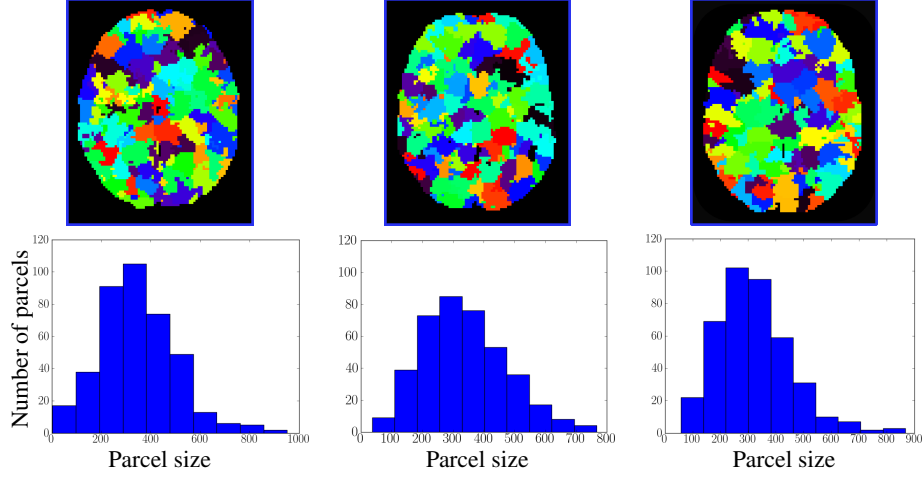
Figure 2: Parcellations ($\Gamma = 500$) shown in the same slice for three different subjects. For each subject, the histogram corresponding to the distribution of parcel sizes is also reported. The *group-level* minimum, mean and maximum parcel sizes are (33, 340, 951) voxels, respectively.

Hence, the activation response to the mth stimulus type in voxel j is given by $\boldsymbol{h}_\gamma \times a_j^m$. In contrast to Eq. (1), the fixed HRF components $\boldsymbol{h}_c$ and $\boldsymbol{h}_c'$ are replaced by an *unknown* parcel-based HRF $\boldsymbol{h}_\gamma$. In the same respect, each unknown NRL a_j^m embodies a single magnitude parameter per regressor whereas the GLM formulation implies that the magnitude is distributed between both β_j^m and $\widetilde{\beta}_j^m$. To summarize, the HRF shape and the BOLD response magnitude are coupled in the GLM formulation whereas they are decoupled in the JDE formulation.
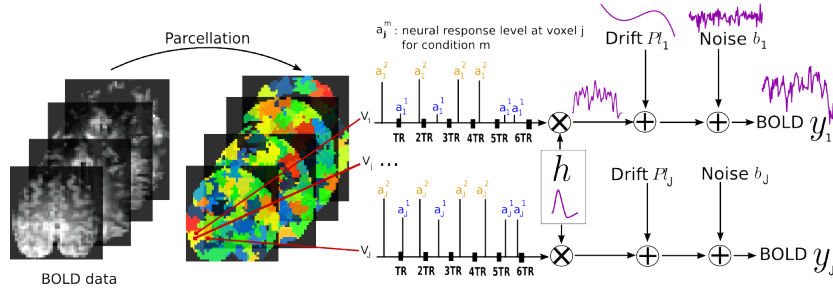

Figure 3: General overview of the subject-level processing in the JDE framework: the BOLD time series are first parcellated (left) and then for every parcel $\mathcal{P}_\gamma$ the JDE forward model described in Eq. (2) is inferred.

2.2.3. *Summary of the involved priors*

The JDE approach developed in [34] and augmented in [49, 8, 9] introduces proper priors on every unknown parameter $(\boldsymbol{h}_\gamma, \boldsymbol{A})$ in order to robustly estimate the brain activity (localization and HRF time course).

6

The HRF. The prior density for the HRF reads $h_\gamma \sim \mathcal{N}(0, v_h R)$ where $R = (D_2^t D_2)^{-1}$ embodies a smoothness constraint on the second order derivative of h_γ to regularize its estimation. The extreme HRF time points are also set to zero because the response is assumed to be causal and to return to its baseline after about 25s.

The Neural Response Levels (NRLs). It is assumed that different stimulus types induce statistically independent NRLs [34, 49], i.e. $p(A \mid \theta_a) = \prod_m p(a^m \mid \theta^m)$ with $\theta_a = (\theta^m)_{m=1:M}$ and θ^m the set of unknown hyper-parameters related to the mth stimulus type. Detection is handled through the introduction of activation class assignments $Q = \{q^m, m = 1 \ldots M\}$ where $q^m = \{q_j^m, j \in \mathcal{P}_\gamma\}$. q_j^m represents the *activation class* at voxel j for condition m, with $q_j^m = 0$ if voxel j is non-activating for condition m and $q_j^m = 1$ otherwise. The NRLs will therefore be expressed conditionally to these hidden variables and are assumed independent in space conditionally to q^m: $p(a^m \mid q^m; \theta^m) = \prod_{j \in \mathcal{P}_\gamma} p(a_j^m \mid q_j^m; \theta^m)$. If $q_j^m = i$ then $p(a_j^m \mid q_j^m = i; \theta^m) \sim \mathcal{N}(\mu_i^m, v_i^m)$. The statistical parameters $\theta^m = \{\mu_i^m, v_i^m, i = 0, 1\}$ are unknown, but for non-activated voxels we set $\mu_0^m = 0, \forall m$.

Activation classes. Prior independence is assumed between the M experimental conditions regarding the activation class assignments. A hidden Markov (Ising) model is chosen as spatial prior on q^m so as to introduce spatial regularization. The regularization parameter is estimated for each stimulus. The reader can refer to the methodological publications for more details [34, 49].

The hyper-parameters. Conjugate prior distributions have been chosen for all hyper-parameters related to the unknown variables of interest $\{h_\gamma, A, Q\}$ as well as confounds L. A non-informative Jeffreys prior has been used for noise variance and a uniform prior density over $[-1, 1]$ has been considered for the auto-regressive parameters; see [49] for details. The notation Θ refers to all hyper-parameters of the models.

2.3. The intermediate SAGLM approach

In the present paper, we also investigate the sole impact of spatial regularization involved in the JDE formalism, by setting the HRF shape h_γ to its canonical value in every parcel $\mathcal{P}_\gamma$. In this intermediate approach called *Spatially Adaptive* GLM or SAGLM, the generative BOLD model remains linear and is given by Eq. (1) but the inference becomes multivariate in contrast to the classical GLM approach. Indeed, inference is still carried out in each parcel, enabling a parcel-dependent spatial regularization.

The interest for testing the SAGLM approach lies in the ability to separate *i.)* the influence of spatially adaptive regularization vs. invariant smoothing while keeping the canonical HRF (SAGLM vs GLM comparison) and *ii.)* the impact of estimating the HRF in a spatially adaptive regularization scheme (JDE vs SAGLM comparison).

2.4. Bayesian inference

In [49], JDE inference is performed using a hybrid Metropolis-Gibbs sampler to draw realizations from the full posterior distribution $p(h_\gamma, A, L, Q, \Theta \mid Y)$ in every parcel $\mathcal{P}_\gamma$. Numerical details are provided in [49]. The same principle applies also here for SAGLM inference except that it involves $p(A, L, Q, \Theta \mid Y)$ since h_γ is fixed.

3. Simulation results

Simulations were conducted at the group-level to compare the statistical activation detection performance of the different subject-level inferences. They also make it possible to separate the impact of discrepancy to canonical HRF from that of spatial smoothing vs. unsupervised regularization. To this end, we considered two artificial data sets on a single 2D grid (20×20 pixels; cf. Fig. 4(a)) that simulate 8 subjects with realistic SNRs. This number of subjects is smaller than classical group size (e.g., 15 individuals) and was chosen so as to counterbalance the fact that analyzing artificial data under the true generative model (either GLM or JDE-based) is easier than extracting relevant information from real fMRI time series. Indeed, the latter may embody nonlinear effects like saturation, habituation or non-stationarities due, for instance, to learning: all these aspects as well as subject varying confounds (noise term, drift basis) require a sufficient cohort size (varying from 12 to 20 individuals according to the experimental paradigm) to derive significant effects at the population level from real fMRI data. Thus, the improved quality of synthetic data and our trust in the involved model at the inference step are compensated by a smaller group-level effect and thus a smaller cohort size.

In both simulation scenarios, the BOLD time series were simulated according to Eq. (2) where $\boldsymbol{h}_\gamma$ was set to the canonical HRF shape for all subjects in the first set of simulations. In contrast, we introduced subject-varying HRF shapes in the second simulated scenario using three different shapes that strongly deviate from the canonical function (see Fig. 5 for details). All the remaining parameters that entered the simulation (NRLs, drift parameters and noise realizations) were kept constant for the two scenarios so as to draw fair conclusions regarding the impact of HRF shape fluctuations. As regards the BOLD signal, we simulated a fast event-related paradigm comprising a single condition ($M = 1$, ISI=3.5 s) encoded in $\boldsymbol{X}^1$. Matrix $\boldsymbol{P}$ was set to a fixed polynomial basis of order 4 for all subjects and subject-specific drift coefficients were drawn according to $\boldsymbol{\ell}_j^s \sim \mathcal{N}(0, 0.3), \forall s = 1 : 8$. Finally, white noise was superimposed as follows: $\boldsymbol{b}_j^s \sim \mathcal{N}(0, 1.1)$. To limit the number of factors impacting the group-level results, the same three activation patterns were chosen for all subjects (see Fig. 4(b)). This amounts to neglecting inter-individual anatomo-functional variability. Nonetheless, the contrast-to-noise ratio (CNR)[2] between activated and non-activated pixels varied between subjects. This allowed us to mimic what is usually encountered on real fMRI data, namely a BOLD effect size strongly varying across subjects. More precisely, for each subject s, the NRLs $a_j^{1,s}$ in non-activated pixels ($q_j^{1,s} = 0$) were drawn according to $\mathcal{N}(0, v_0^1)$ with $v_0^1 = 0.3$. Also, for 6 out of 8 subjects, the NRLs $a_j^{1,s}$ in activated pixels ($q_j^{1,s} = 1$) were sampled from $\mathcal{N}(\mu_1^1, v_1^1)$ with $(\mu_1^1, v_1^1) = (2, 0.5)$. For the remaining two subjects, a degraded CNR situation (smaller BOLD effects) was simulated using $\mu_1^1 = 1.2$ instead. Owing to these fluctuations, the simulated SNR[3] was subject-dependent and varied within the [2.1, 5.3] dB range across subjects. The sim-

---

[2]$\mathrm{CNR}_1 = 2(\mu_1^1)^2/(v_0^1 + v_1^1)$.

[3]$\mathrm{SNR} = 10 \log \sum\limits_{j=1}^{J_\gamma} \|a_j^1 \boldsymbol{X}^1 \boldsymbol{h}_\gamma\|^2 / \sum\limits_{j=1}^{J_\gamma} \|\boldsymbol{b}_j\|^2$.

ulated data that entered in the SAGLM and JDE analyses remained unsmoothed. For the GLM analysis, data were smoothed with an isotropic Gaussian kernel, ensuring a level of smoothness of 1.5 pixels, as performed on the real data (see Subsection 4.3).
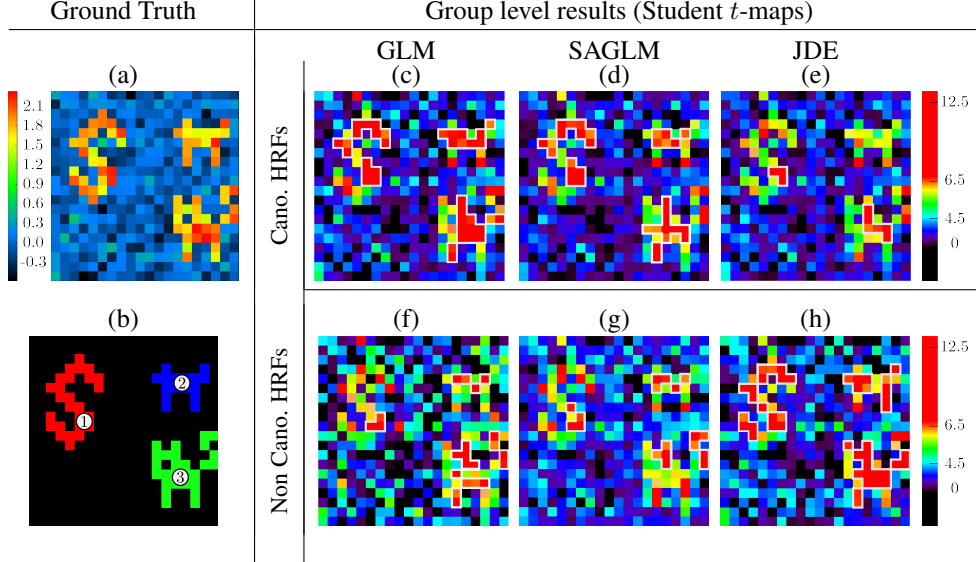

Figure 4: **(a):** Simulated averaged activity over subjects (true mean NRLs). **(b):** Spatial configuration of true activation clusters: clust. #1: 26 activated voxels; clust. #2: 16 activated voxels and clust. #3: 28 activated voxels. **(c)-(e):** Group-level Student t-maps related to the first scenario (simulated canonical HRFs) and obtained for the three intra-subject inference methods (GLM, SAGLM and JDE from left to right). **(f)-(h):** Group-level Student t-maps related to the second scenario (subject-varying non-canonical HRFs) and obtained for the same methods as above.

Figure 5: JDE-based HRF estimates for each subject along with the ground truth (denoted GT). **(a):**In the first simulation scenario, all subjects elicit BOLD responses according to the canonical HRF shape $\boldsymbol{h}_c$ shown in **black line**. **(b):**In the second simulation scenario, subject-varying HRFs that deviate from $\boldsymbol{h}_c$ have been used as illustrated by the color coding: 4/2/2 subjects were associated with the ground truth HRFs in blue, green and red shapes, respectively.

As all simulated data have low SNR, the group-level results depicted in Fig. 4 show

Table 1: Group-level comparison of activation detection performance of the JDE, SAGLM and GLM subject-level inferences in two simulation scenarios (canonical vs. subject-varying HRF). Cluster-level Family-wise Error Rate (FWER) correction at $p_{\mathrm{val}}^{\mathrm{clust}} < 0.05$. Only clusters comprising three voxels or more are reported as significant. Significant values are reported in **bold font**. The column %TP gives the percentage of true positives retrieved within each cluster. The rows are sorted first according to cluster index and second within each cluster according to Student t-scores at the pixel level.

		Group-level statistical results				
		Cluster-level				Pixel-level (peak)
	Subject-level inference	Cluster	corr. $p_{\mathrm{val}}^{\mathrm{clust}}$	Size	%TP	t-score
Canonical HRFs	GLM	1	0.05	5	0.19	11.7
		1	**0.03**	**10**	**0.38**	**11.24**
		2	**0.04**	**9**	**0.56**	**8.85**
		3	**0.01**	**14**	**0.5**	**11.1**
	SAGLM	1	**0.01**	**5**	**0.19**	**12.14**
		1	**0.04**	**10**	**0.38**	**9.85**
		2	**0.02**	**13**	**0.46**	**8.65**
		3	**0.04**	**6**	**0.375**	**8.66**
	JDE	1	0.07	3	0.12	10.01
		3	0.1	6	0.21	6.64
Non canonical HRFs	GLM	1	**0.04**	**3**	**0.11**	**8.5**
		2	0.05	4	0.25	6.1
		3	**0.02**	**11**	**0.39**	**7.5**
		3	0.07	4	0.14	7.1
	SAGLM	1	**0.04**	**4**	**0.15**	**8.3**
		2	0.1	5	0.31	5.8
		3	**0.04**	**5**	**0.18**	**8.9**
		3	0.07	4	0.14	6.85
	JDE	1	**0.007**	**14**	**0.54**	**12.26**
		1	**0.01**	**4**	**0.25**	**11.2**
		2	**0.01**	**5**	**0.31**	**10.1**
		2	**0.01**	**5**	**0.31**	**10**
		3	**0.009**	**17**	**0.6**	**11.4**

several false negatives but no false positive, for all intra-subject inference schemes. However, the group-level differences in terms of activation detection performance between the intra-subject models appear clearly in Tab. 1. In the first simulation scenario, the GLM and SAGLM inference schemes retrieved the most significant activated clusters ($p_{\mathrm{val}}^{\mathrm{clust}} < 0.05$) with positions that closely match the ground truth. As shown in Tab. 1[Top], these two approaches exactly found the same number of activated pixels in cluster 1 separated in two subclusters. The highest peak values were given by SAGLM inference for the smaller subcluster and by GLM inference for the larger one. Regarding the two remaining activation clusters, the two approaches yielded similar results although they found different compromises in terms of peak height/spatial extent, as expected given the difference between spatial smoothing and regularization. In contrast, JDE-based inference was much less sensitive since it only retrieved few activated pixels from clusters 1 and 3, and none from cluster 2. Even at the pixel level, the peak values were lower using JDE inference. Hence, these results confirm at the group-level that under canonical HRF assumptions, hemodynamics estimation in JDE is performed at the expense of the pixel and cluster-level sensitivity of detection.

In the second scenario, JDE-based inference yielded more sensitive results compared to those deriving from canonical assumptions. In each cluster, the peak t-value and the cluster-level p-values ($p_{\mathrm{val}}^{\mathrm{clust}}$) were more significant using JDE inference and the cluster sizes were larger. Hence, our group-level conclusions are unambiguous and in favour of the proposed JDE intra-subject analysis when departing from the canonical HRF shape (see Tab. 1[Bottom] for details). When pulling results of both scenarios together, it is also worth noticing that the most significant clusters representing the largest percentage of true positives were exhibited by JDE inference in cluster 1.

With respect to HRF estimation, despite the low SNR JDE inference was able to recover robust subject-dependent time courses, close to the canonical and non-canonical profiles in the first and second scenarios respectively, see Fig. 5(a)-(b). The second scenario highlights different hemodynamics regimes in which JDE outperforms its competing alternatives in terms of activation detection performance. These results provide a benchmark for the same group-level comparisons on real data in the next sections.

4. Data acquisition and group-level analysis pipeline

4.1. Imaging parameters

Fifteen ($S = 15$) healthy volunteers (average: 23.2 years, std: 2 years) gave written informed consent before participating to an imaging study on a 3T MRI wholebody scanner (Tim Trio, Siemens). The study received ethics committee approval by authorities responsible for our institution. Anatomical imaging used a T1-weighted magnetization prepared rapid acquisition gradient echo sequence (176 slices, repetition time TR = 2300 ms, echo time TE = 2.98 ms, sagittal orientation; field of view (FOV) 256 mm, voxel size $1 \times 1 \times 1$ mm^3). Functional imaging used a T2*-weighted gradient-echo, Echo-Planar-Imaging (EPI) sequence (40 slices, TR = 2400 ms, TE = 30 ms, FOV 192, voxel size $2 \times 2 \times 3$ mm^3). Each fMRI run consisted of $N = 128$ EPI volumes that were collected using a 32 channel head coil to enable parallel imaging during

the EPI acquisitions. Parallel SENSE imaging[4] was used to keep a reasonable TR value in the context of relative high spatial resolution for whole brain imaging. Two acceleration factors $R = 2$ and $R = 4$ were tested to measure the impact of the SNR loss on brain activity detection and hemodynamics estimation: the larger R, the lower the SNR [43][5]. Acquisition order with $R = 2$ and $R = 4$ was randomly balanced across the fifteen participants.

4.2. Experimental paradigm

We used a multi-functional cognitive localizer paradigm [42], consisting of a fast acquisition procedure based on a 5-minutes fMRI sequence. It was organized as a fast *event-related* design, comprising 100 short trials presented in a fixed sequence with a stochastic Stimulus Onset Asynchrony (SOA:2400 ms, 2700 ms, 3000 ms, 3300 ms or 3600 ms; mean SOA = 3 s). This protocol was designed to obtain a description of different levels of functional architecture, from the cerebral bases of auditory and visual perception, motor actions, to reading, language comprehension and mental calculation at an individual level. In this perspective, ten types of trials were mixed together and presented successively with the E-Prime software (Psychology Software Tool, Inc.): *1-2.)* passive viewing of flashing horizontal (10 trials) and vertical (10 trials) checkerboards, *3-4.)* pressing three times the left (resp., right) button with the left (resp., right) thumb button according to visual instructions (5 trials), *5-6.)* pressing three times the left (resp., right) button with the left (resp., right) thumb button according to auditory instruction (5 trials), *7.)* silently reading short visual sentences (10 trials), *8.)* listening to short sentences (10 trials), *9.)* silently solving visual subtraction problems (10 trials), *10.)* solving silently auditory subtraction problems (10 trials). 20 rest periods (black screen) were inserted in the sequence and served as null events for a better hemodynamic deconvolution.

4.3. fMRI data pre-processing

We used SPM5 for image pre-processing: realignment, coregistration, normalization to MNI stereotactic space and spatial smoothing with a 3 mm full-width at half-maximum isotropic Gaussian kernel for GLM-based analyses. Our smoothing level corresponds to 1.5 voxel resolution as FWHM parameter. This level may appear pretty low but the acquisition of spatially resolved fMRI data targets an improved localization of evoked activity which has not to be sacrificed by a large spatial filtering of images at the pre-processing stage. Indeed, a too high smoothing level would artificially dilates the activation clusters. JDE and SAGLM-based analyses were carried out on normalized but *unsmoothed* fMRI images.

---

[4]In the 2D k-space, one line out of R was sampled along the phase encoding direction but collected over 32 channels with complementary sensitivity profiles.

[5]We remind here that the SNR of the downsampled image behaves as $\text{SNR}^{\text{down}} = \text{SNR}^{\text{full}}/g\sqrt{R}$ where SNR^{full} denotes the SNR of the full k-space acquired image and g refers to as the geometric factor. The latter depends on the spatial correlation between the coils and to some extent on their number [43].

4.4. Random effect analyses for detecting evoked activity

In this part, we aim at measuring the incremental group-level impact of replacing the subject-level GLM fitting procedure by the SAGLM or JDE ones, in terms of activation detection performance.

Fig. 6 summarizes the different steps involved in these group-level comparisons. In particular, step 3 outlines that random effects analyses were conducted either from the voxel-based GLM parameter estimates $\widehat{\beta}_j^m$ or from the SAGLM/JDE neural response levels $\widehat{a}_j^m$. Group-level Student t-statistics were computed with a nonparametric calibration by permutations of the distribution under the null hypothesis. This approach provides an exact control of false positives under mild assumptions: for one-sided tests, the distribution under the null hypothesis is assumed symmetric which is more general than imposing normality as done in parametric inference [26, 39]. Moreover, nonparametric statistical inference provides an exact solution to the multiple comparison problem without any supplementary assumption in contrast to the random field theory (RFT) which requires multivariate normal hypothesis and thus smoothing the data with a large kernel so as to accurately estimate the Euler Characteristic Expectation [26, 39, 37, 44]. Thus, instead of the RFT correction, used the `Distance` toolbox for SPM5 [37, 44] to derive our group-level statistical results. This toolbox offers both cluster- and voxel-level corrections using the sign permutation framework. In the present study, we corrected for multiple comparisons at both the cluster and voxel-levels using $N = 2^{15}$ permutations and used $\alpha = 0.05$ as significance threshold. Moreover, for the cluster-level inference, we set the cluster forming threshold to *uncorrected* $p_{\text{value}} = 0.001$ at the voxel level.

A random effect (RFX) analysis was conducted for different contrasts of interest targeting brain activity in sensory and cognitive regions. In particular, we focused on:

- the global *Auditory vs. Visual* (**A - V**) contrast for which we expect evoked activity in temporal regions. The **A - V** contrast defines a compound comparison involving eight stimulus types (two motor actions, sentence listening/reading, mental calculation), which are presented either in the auditory or visual modality (see Appendix A of the Supplementary Materials). We also studied activation detection in the visual cortex via the *Visual vs. Auditory* (**V - A**) contrast that induces evoked activity in the occipital cortex. The results obtained for these contrasts are in agreement with previous observations in the literature.

- the *Left vs. Right click* (**Lc - Rc**) contrast for which we expect evoked activity in the right contralateral motor cortex. Indeed, the **Lc - Rc** contrast defines a compound comparison involving motor stimulus types either presented in the visual or auditory modality.

- the *Computation vs. Sentence* (**C - S**) contrast which is supposed to elicit activity in the frontal and parietal lobes since solving mental arithmetic task involves working memory and more specifically the intra-parietal sulcus [14]. The **C - S** contrast defines a compound comparison that relies on visual and auditory stimuli. In the visual case, the stimuli consist of word reading, one half being written calculations submitted to the subject, the other being made up of normal written sentences. In the auditory case, the stimuli consist of speech listening,
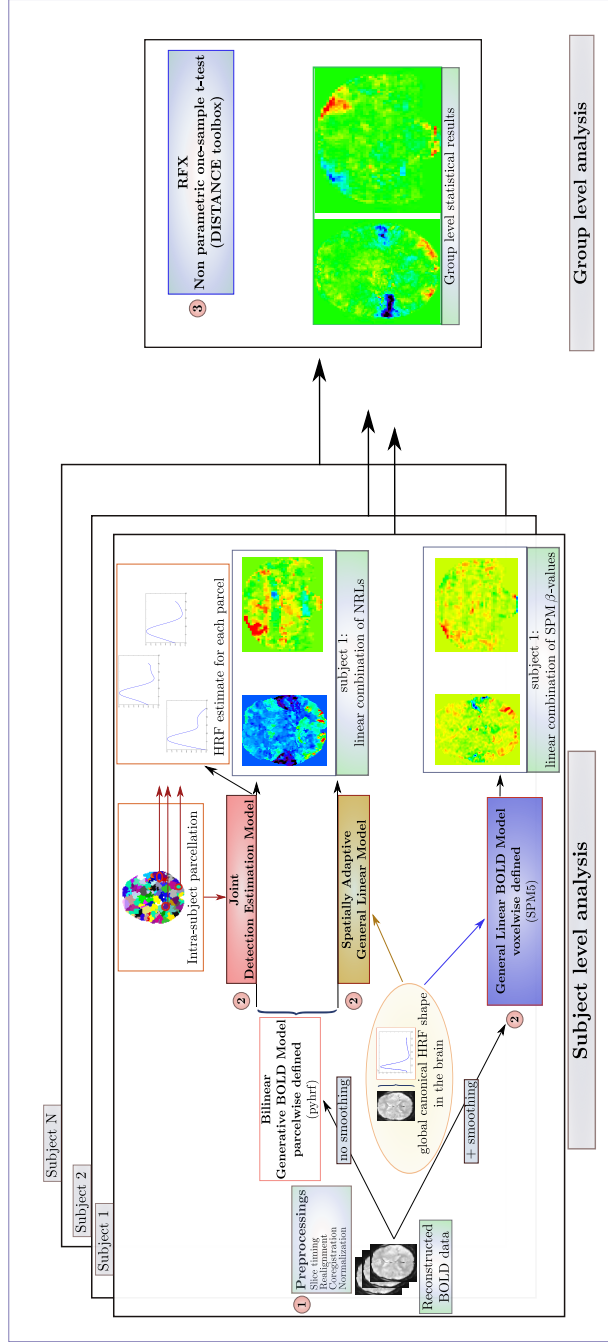
Figure 6: General sketch of the group-level comparison involving three kinds of subject-level inference (GLM, SAGLM and JDE) at step 2. Note that the group-level statistics have been computed using the nonparametric tests developed in the `Distance` toolbox for SPM5 [37, 44].

one half corresponding to calculations orally instructed to the subject, the other to normal spoken sentences.

4.5. Group-level hemodynamic studies

We performed hemodynamics analyses in four regions of interest selected from significantly activated clusters that were exhibited at the group-level in the above mentioned contrasts of interest. More precisely, left parietal (P), left motor (M), temporal (T) and occipital (O) regions were extracted from contrasts **C - S**, **Lc - Rc**, **A - V** and **V - A**, respectively, as shown in Fig. 7. Hereafter, we denote r as region index such that $r \in \Upsilon$ and $\Upsilon = \{P, M, T, O\}$. Choosing these regions from GLM-based analyses could induce a potential bias in the HRF estimates towards the canonical HRF. However, we have proceeded in such manner since our goal was to compare subject-level inferences on real data where the activation localization was already known and documented [42]. As shown in Section 5.2.1, significant discrepancies from the canonical function could be retrieved even in this context.

For comparison purposes, we focused on two hemodynamics parameters, namely the time-to-peak (TTP) and full width at half-maximum (FWHM) and analyzed how they fluctuate between regions, subjects and with respect to the SNR (acceleration factor). We statistically assessed to which extent the HRF shape significantly departs from the canonical form in several brain regions by comparing GLM and JDE-based hemodynamics parameters. In what follows, the extraction of region-specific HRFs and the computation of TTP and FWHM parameters are detailed. Next, we summarize the computation of group-level JDE-based HRF time courses so as to study their shape variability across regions.
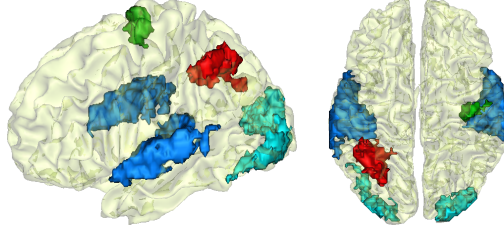

Figure 7: Definition of regions of interest to investigate hemodynamics variability from JDE-based group-level analysis ($R = 2$). Significant clusters shown using $p_{\mathrm{val}}^{\mathrm{clust}} = 0.05$, corrected for multiple comparisons. **Left**: Sagittal view; **Right**: axial/top view. Left parietal area (P) appears in red, left motor area in the pre-central cortex is shown in green, Bilateral temporal regions along auditory cortices and bilateral occipital regions in the visual cortices are shown in blue and cyan, respectively.

4.5.1. Regionwise HRF extraction

To study the *shape* of the regionwise HRFs eliciting evoked activity we consider normalized HRF profiles, ie $||\widehat{\boldsymbol{h}}||^2 = 1$. As GLM-based inference only provides voxel-wise estimates, subject-specific regionwise HRF estimates were computed at the local

activation peak within each region r as follows:

$$\widehat{\boldsymbol{h}}_{r,s}^{\text{GLM}} = \widehat{\beta}_{j^\star,s}^{m^\star}\boldsymbol{h}_c + \widehat{\widetilde{\beta}}_{j^\star,s}^{m^\star}\boldsymbol{h}_c', \quad (j^\star, m^\star) = \underset{j,m}{\arg\max}\big\{\beta_{j,s}^m \,|\, j \in r, m = 1\dots M\big\} \quad (3)$$

$$\widehat{\boldsymbol{h}}_{r,s}^{\text{JDE}} = \widehat{\boldsymbol{h}}_{\gamma^\star,s}^{\text{JDE}}, \quad (\gamma^\star, j^\star, m^\star) = \underset{\gamma,j,m}{\arg\max}\big\{\widehat{a}_{j,s}^m \,|\, j \in r \cap (\mathcal{P}_\gamma)_\gamma, m = 1\dots M\big\} \quad (4)$$

Hence $j^\star \in \mathcal{P}_{\gamma^\star}$. We are now interested in extracting the TTP and FWHM since the former has been found the most stable parameter and the latter is known to be more variable [38]. From individual regionwise HRF estimates for any method $\psi \in \big\{\text{GLM}, \text{JDE}\big\}$ we are given these parameters according to:

$$\text{TTP}_{r,s}^\psi = \Delta t \underset{d\in\{0,\dots,D\}}{\arg\max} (\widehat{h}_{d\Delta t}^\psi)_{r,s} \quad (5)$$

$$\text{FWHM}_{r,s}^\psi = \Delta t(d_2 - d_1) \,|\, (\widehat{h}_{d_1\Delta t}^\psi)_{r,s} = (\widehat{h}_{d_2\Delta t}^\psi)_{r,s} = \frac{\max(\widehat{\boldsymbol{h}}_{r,s}^\psi)}{2}, \ d_1 < d_2. \quad (6)$$

Each parameter in turn was entered in two-way repeated measures ANOVA so as to study the effect of the region $r \in \Upsilon$ and the acceleration factor (or SNR) with $R = 2$ and $R = 4$. Then, we statistically compared JDE vs GLM-based hemodynamic parameters using two-way repeated measures ANOVA and paired t-tests.

4.5.2. Group-level JDE-based HRF

To compute group-level HRFs and take both the activation level and hemodynamics shape variability into account, individual JDE-based HRF estimates were weighted by their corresponding maximum activation magnitude identified in the previous step (see Fig. 8 for details). Hence, the group-level JDE-inferred HRF profile thus reads:

$$\bar{\boldsymbol{h}}_r^{\text{JDE}} = \frac{\sum_{s=1}^S \widehat{a}_{j^\star,s}^{m^\star} \widehat{\boldsymbol{h}}_{r,s}^{\text{JDE}}}{\sum_{s=1}^S \widehat{a}_{j^\star,s}^{m^\star}} \quad (7)$$

5. Result on real fMRI data

5.1. Group-level comparisons in activation detection

For all contrasts of interest, we first compared the group-level performance in detection on the fMRI data acquired with $R = 2$ as SENSE acceleration factor (best SNR) before measuring the impact of degrading the SNR ($R = 4$).

5.1.1. *Lc - Rc* contrast

As shown in Fig. 9(a)-(c) and Tab. 2 for *R=2*, all RFX analyses found the same activation cluster in the right motor cortex, whatever the intra-subject modeling. Also, the localization of the activation peak in the MNI space was very stable across all analyses (GLM, SAGLM and JDE). In terms of cluster-level significance, GLM-based inference yield more sensitive results at the group level than its alternatives, as outlined

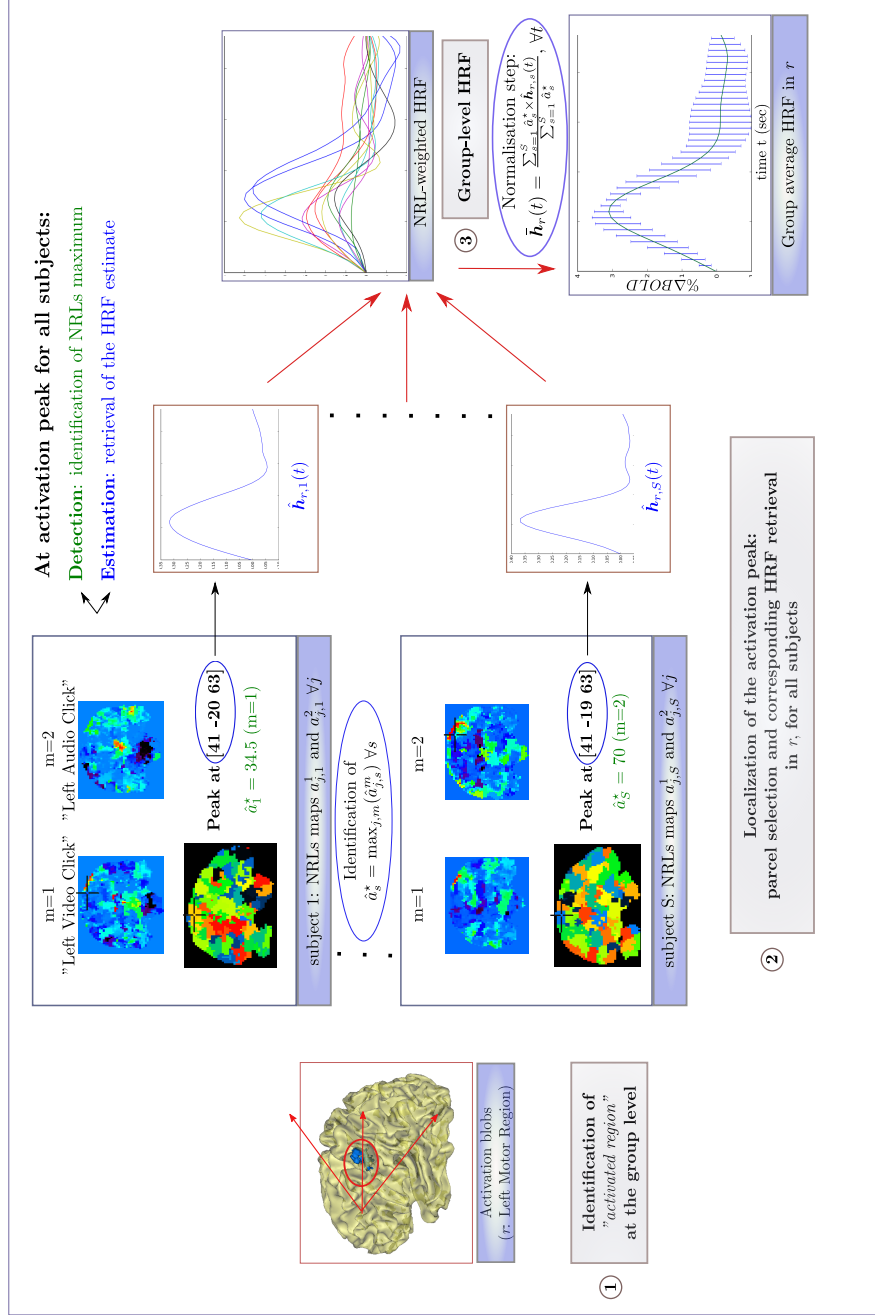Figure 8: General sketch summarizing the HRF computation at the subject and group-levels in activated regions $r \in \Upsilon$. **Left**: Position of the activation peak in r (here left motor cortex) given in mm in the Talairach space. **Center**: Individual weighted HRF time course extraction. **Right**: Computation of the Group average normalized HRF time course with corresponding error bars ($\pm \sigma$).

in Tab. 2[Top]. However, SAGLM-based inference retrieved the highest peak values. As before on simulation results (first scenario in Section 3), the spatial smoothing of fMRI data (GLM-based analysis) and the spatial regularization (SAGLM-based analysis) induce a different trade-off between cluster extent and peak value, the former emphasizing the cluster size whereas the latter emphasizes more the peak excursion. This clearly indicates the benefit of regionally estimating the spatial regularization level from the data itself instead of applying a spatially invariant smoothing.

By comparing the SAGLM and JDE-based inferences at the group-level (see Fig. 9(b)-(c)), it can be pointed out that estimating the HRF in motor regions has a negative impact: although the JDE-based analyses remain significant with $p_{\mathrm{val}} < 0.05$ at cluster and voxel levels, a loss in statistical sensitivity was observed, as reported in Tab. 2[Top]. This confirms what we already found on simulation results (first scenario).
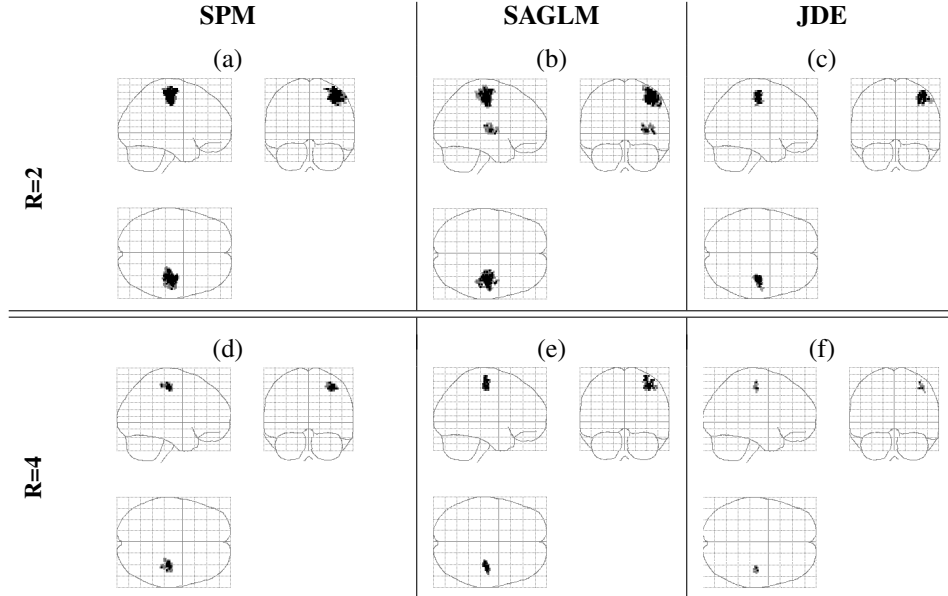

Figure 9: RFX statistical analysis for the **Lc - Rc** contrast and acceleration factors $R = 2$ **(a-c)** and $R = 4$ **(d-f)**, derived using the GLM **(a,d)**, SAGLM **(b,e)** and JDE **(c,f)** subject-level analyses. Statistical maps were thresholded at $p_{\mathrm{val}}^{\mathrm{vox}} < 0.001$ uncorrected, and corrected for multiple comparisons at $p_{\mathrm{val}}^{\mathrm{clust}} < 0.05$, using calibration by permutations.

On the more noisy fMRI data sets ($R = 4$), our investigation yields similar conclusions: first, all RFX analyses recovered activation clusters in the same position (see Tab. 2[Bottom]). Second, SAGLM-based inference was the most significant at the voxel-level. Third, GLM-based analysis provided us with the largest cluster extent at the expense of the peak value (see Tab. 2[Bottom]). As regards the JDE-based results, we observed a significant loss in statistical sensitivity: in contrast to $R = 2$, cluster level p_{val} is barely significant ($p_{\mathrm{val}}^{\mathrm{clust}} = 0.04$) whereas the voxel-level p_{val} is not significant ($p_{\mathrm{val}}^{\mathrm{vox}} > 0.1$). To summarize, modeling the HRF shape in the motor cortex induces

a loss of significance. As confirmed later in Section 5.2, we hypothesize that the underlying reason lies in the HRF shape: the canonical filter seems to match well the hemodynamic response in the motor cortex, as already demonstrated in [53]. Hence, it is suboptimal to assume the HRF unknown when an accurate model is available.

*5.1.2. **C - S** contrast*

For $R = 2$, all RFX analyses found a large activation cluster in the parietal cortex as shown in Fig. 10 (a)-(c) . In frontal regions, the GLM and SAGLM-based analyses retrieved smaller clusters in the middle and inferior left frontal cortices (Frontal Inf. Orb. L. and Frontal Mid. L. [AAL atlas]) while the JDE-based analysis found two frontal clusters with even smaller extent, localized only in the middle left frontal cortex.

In terms of cluster level significance, SAGLM-based inference yield more sensitive results than its alternatives, as outlined in Tab. 3[Top]. However, JDE-based inference retrieved the highest peak value at the voxel level compared to its alternatives. Importantly, the peak position in the intra-parietal sulcus remained stable whatever the intra-subject inference as reported in Tab. 3[Top]. Here, we hypothesize that the JDE approach demonstrates its group level positive impact since the HRF shape in the parietal cortex strongly departs from the canonical one. This assumption will be statistically assessed in Section 5.2, where the HRF shape estimates will be compared.

On the more noisy fMRI data set ($R = 4$), our investigation on the **C - S** contrast enforced the previous result. Only SAGLM and JDE-based inferences recovered a *significant* activation cluster in the left parietal cortex, close to the exhibited maximum on the RFX analyses performed at $R = 2$ (compare Fig. 10(b)-(e) with Fig. 10(c)-(f) and Tab. 3[Top] with Tab. 3[Bottom]). This confirms the interest of both adaptively regularizing the NRLs and modeling the HRF shape in regions where a large discrepancy to the canonical filter might occur, as shown on the second simulation scenario in Section 3. Tab. 3[Bottom] also showed that SAGLM-based inference exhibited the largest cluster while the JDE-based analysis retrieved close results, which were much significant than those deriving from GLM inference. At the voxel level, JDE inference

Table 2: Group-level comparison for the **Lc - Rc** contrast using the JDE, SAGLM and GLM subject-level inference ($R = 2$ and $R = 4$). Cluster-level FWER correction at $p_{\mathrm{val}}^{\mathrm{clust}} = 0.05$. Significant values are reported in **bold font** ($p_{\mathrm{val}} < 0.05$).

| | | Group-level statistical results | | | | |
| | | Cluster-level | | Voxel-level (peak) | | |
	Subj.-level inference	corr. $p_{\mathrm{val}}^{\mathrm{clust}}$	Size	corr. $p_{\mathrm{val}}^{\mathrm{vox}}$	t-score	Position (mm)
$R = 2$	GLM	**< 0.001**	**454**	**< 0.001**	**10.12**	[36 -22 54]
	SAGLM	**0.001**	**390**	**< 0.001**	**12.21**	[38 -22 57]
	JDE	**0.002**	**169**	**0.02**	**6.76**	[44 -16 48]
$R = 4$	GLM	**< 0.001**	**108**	**0.006**	**7.69**	[36 -20 51]
	SAGLM	**0.001**	**89**	**0.004**	**8.59**	[32 -20 63]
	JDE	**0.042**	**24**	0.72	5.1	[36 -20 51]

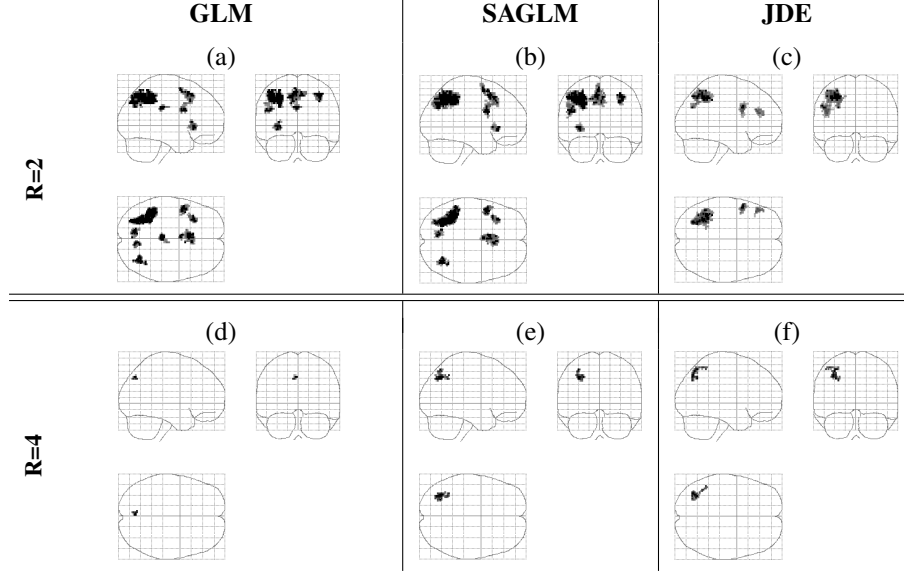|	GLM	SAGLM	JDE
R=2	(a)	(b)	(c)
R=4	(d)	(e)	(f)

Figure 10: RFX statistical analysis for the **C - S** contrast and acceleration factors $R = 2$ **(a-c)** and $R = 4$ **(d-f)**, derived using the GLM **(a,d)**, SAGLM **(b,e)** and JDE **(c,f)** subject-level analyses. Statistical maps were thresholded at $p_{\mathrm{val}}^{\mathrm{vox}} < 0.001$ uncorrected, and corrected for multiple comparisons at $p_{\mathrm{val}}^{\mathrm{clust}} < 0.05$, using calibration by permutations.

Table 3: Group-level comparison for the **C - S** contrast using the JDE, SAGLM and GLM subject-level inference ($R = 2$ and $R = 4$). Cluster-level FWER correction at $p_{\mathrm{value}}^{\mathrm{clust}} < 0.05$. Significant values are reported in **bold font** ($p_{\mathrm{val}} < 0.05$).

		Group-level statistical results				
		Cluster-level		Voxel-level (peak)		
	Subj.-level inference	corr. $p_{\mathrm{val}}^{\mathrm{clust}}$	Size	corr. $p_{\mathrm{val}}^{\mathrm{vox}}$	t-score	Position (mm)
$R = 2$	GLM	**< 0.001**	**617**	**0.002**	**8.7**	[-28 -66 48]
	SAGLM	**< 0.001**	**780**	**< 0.001**	**10.59**	[-30 -56 45]
	JDE	**< 0.001**	**574**	**< 0.001**	**10.77**	[-32 -54 45]
$R = 4$	GLM	**0.04**	**16**	0.98	4.7	[-4 -74 42]
	SAGLM	**0.004**	**81**	0.67	5.31	[-30 -56 45]
	JDE	**0.008**	**70**	0.1	6.7	[-32 -64 42]

yielded again the highest peak value while not significant given the low SNR.

For both values of R, our results for the **C - S** contrast still indicated that estimating the HRF shape brought more sensitivity around the peak. However, there was a statistical price to be paid at the cluster-level since we observed more focal activation patterns. Nevertheless, this price did not seem critical in parietal regions.

5.2. Hemodynamics study

Our aim is to bring a new insight into hemodynamics variability as well as providing meaningful HRF profiles linked to different cognitive states. First, between-subject and between-region hemodynamics variability are characterized from the JDE results only. Second, we bring statistical evidence that the previously exhibited differences in activation detection performance between GLM, SAGLM and GLM-based analyses are due to hemodynamic variability and thus to departure from the canonical HRF.

5.2.1. Group-level hemodynamics characterization based upon JDE analyses

The group-average HRF estimates for $R = 2$ and $R = 4$ in the four ROIs depicted in Fig. 7 are shown in Fig. 11. As explained before, these regions elicited evoked activity in response to auditory and visual stimuli, motor and mental computation task, respectively. We first observed a gradient of discrepancy to the canonical HRF shape between regions: the closest to canonical JDE-based HRF time courses are retrieved in the occipital and temporal regions (see Fig. 11(a)-(b)). The HRF profiles estimated in the motor region deviate a little bit more from the canonical filter, especially in terms of hemodynamic delay, as illustrated in Fig. 11(c). These results were in agreement with the literature [4, 28] and confirmed that the use of a canonical HRF was rather appropriate for detecting evoked activity in auditory and visual cortices.

Besides, as suggested by our activation detection comparisons and shown in Fig. 11(d), the largest discrepancy to the canonical HRF was found in the parietal region with respect to both the hemodynamic delay and activation duration. Interestingly, the SNR variation induced by switching from $R = 2$ to $R = 4$ had little impact on the group-average HRF shape for a given region. This means that the SNR decrease did not induce any bias or trend to another HRF profile. However, the standard deviation of each HRF time course increased as the SNR decreased in the occipital, temporal and motor regions. This especially occurred on the HRF tail close to the undershoot, mainly because of the BOLD effect decreased after the activation peak which induced more fluctuations. Noticeably, this phenomenon also arose at $R = 2$ owing to the fast-event related structure of the experimental paradigm. As the inter-stimuli interval was very short (around 3s), the beginning of an evoked response may actually overlap with the end of a previous one. This makes the estimation of the tail more difficult in the case where consecutive events were associated with experimental conditions that elicit activity in the same region.

Furthermore, when comparing the error bars across regions for $R = 2$ in Fig. 11, we observed that the parietal region was the most variable. This confirms the positive impact of modeling and estimating the impulse response in this area in an appropriate manner. Fig. 12 brings complementary visualizations to assess the incremental deviation from region to region between JDE-based HRF estimates and the canonical

impulse response. The same order is preserved between HRF time courses for $R = 2$ and $R = 4$, as shown in Fig. 12(a)-(b), respectively.
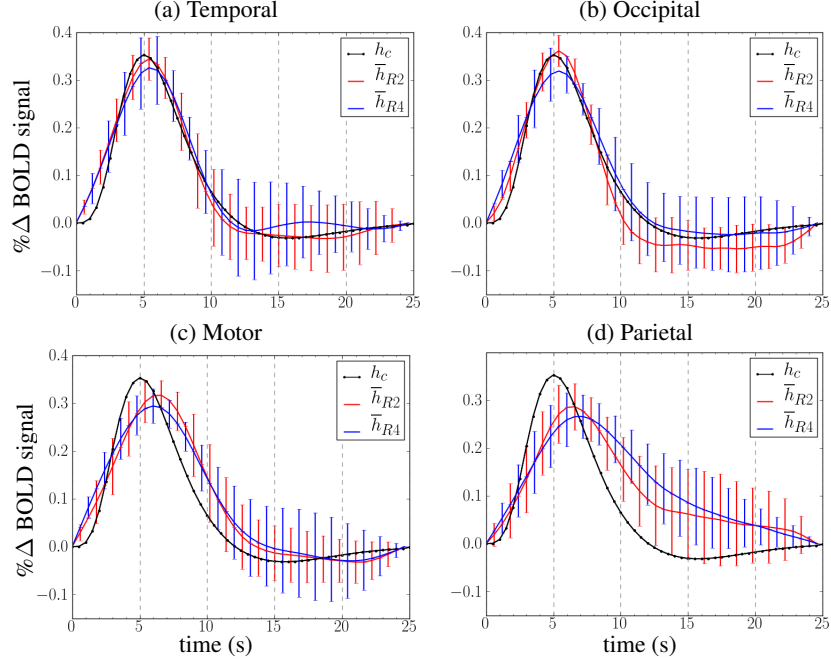

Figure 11: Group-average HRF estimates with superimposed error bars ($\pm\sigma$) for the temporal (a), occipital (b), motor (c) and parietal (d) regions. $\overline{\boldsymbol{h}}_{R2}$ and $\overline{\boldsymbol{h}}_{R4}$ stand for HRF means computed for $R=2$ and $R=4$, respectively. $\boldsymbol{h}_c$ stands for the canonical HRF.

The reader will find in the Supplementary Materials (cf. Tab. C.3) the values of group-average HRF estimates reported for each region and for both acceleration factors.

5.2.2. Summary statistics of hemodynamics parameters

The hemodynamic parameters ($\text{TTP}_{r,s}^{R,\psi}$, $\text{FWHM}_{r,s}^{R,\psi}$) were extracted from GLM- and JDE-based HRF estimates ($\psi \in \{\text{JDE}, \text{GLM}\}$) using Eqs. (5)-(6) for each subject $s = 1 : S$ in each region $r \in \Upsilon$ and for any acceleration factor $R = (2, 4)$. Then, their histograms were represented as box-and-whiskers plots in Figs. 13-14. These graphs confirmed our above mentioned results on group-average HRF time courses and provided additional information about the between-subject variability: a stronger homogeneity of hemodynamic parameter estimates around canonical values was found in occipital and temporal regions irrespective of the acceleration factor and the intra-subject analysis (JDE, GLM). In contrast, a larger discrepancy between JDE-based parameter estimates and canonical values was observed in the motor and parietal regions: these estimates appeared larger than the canonical parameters. A deviation also occurred for GLM-based inference but only with respect to the TTP parameter mainly in the motor region. Moreover, JDE-based TTP and FWHM estimates were systematically larger than the corresponding estimates retrieved by GLM-based inference, which
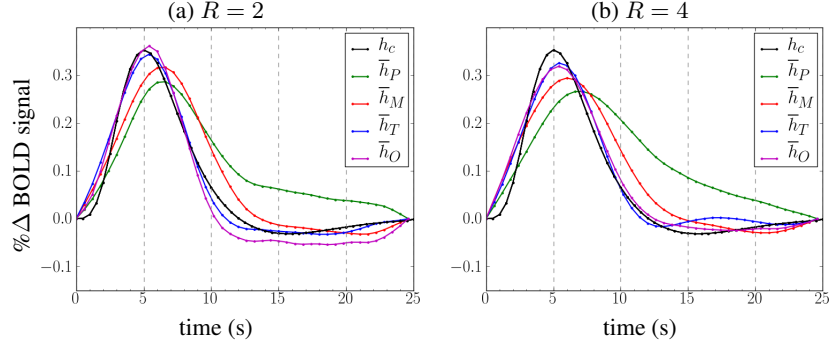
Figure 12: Group-average HRF estimates for acceleration factors $R = 2$ **(a)** and $R = 4$ **(b)**. $\overline{h}_P, \overline{h}_M, \overline{h}_T,$ $\overline{h}_O$ stand for HRF means in parietal, motor, temporal and occipital regions, respectively. h_c stands for the canonical HRF.

were distributed above and below the canonical values depending on the parameter type. Furthermore, a larger between-subject variability was observed in the parietal region, as outlined in Figs. 13-14 by larger boxes for both hemodynamic parameters and both SNR values. In the same figures, a reliable gradient of hemodynamic delay and activation duration was captured by JDE inference only across SNR values.Indeed, for $R = 2$, the fastest (occipital) and slowest (parietal) regions were associated with the less and more sustained activity, respectively, whereas for $R = 4$, the same assertion held except for the occipital region where the activation duration became more dilated than in the temporal cortex.
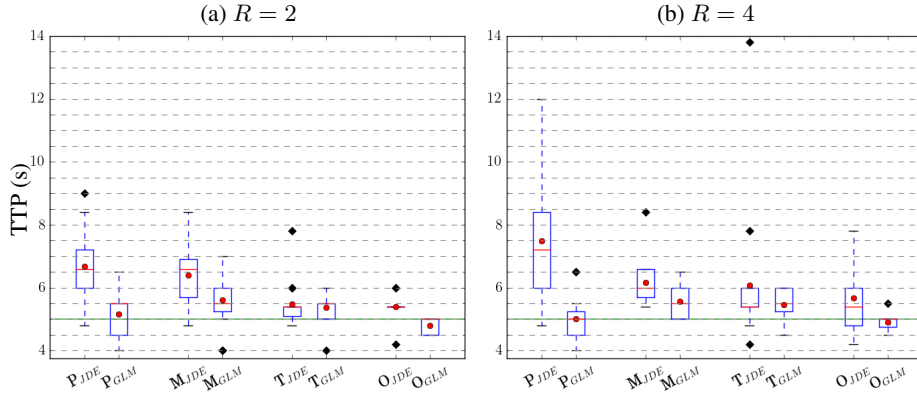

Figure 13: Box-and-whiskers plots (median and mean values shown as red line (–) and red dot (○), respectively) of TTP extracted from JDE-based and GLM-based HRF estimates according to Eq. (5), for $R = 2$ **(a)** and $R = 4$ **(b)**. TTP values are plotted next to another region for the parietal (P_ψ), motor (M_ψ), temporal (T_ψ) and occipital (O_ψ) regions with $\psi = \{JDE, GLM\}$. The green line corresponds to the canonical TTP i.e. to 5s.

These gradients across regions were not retrieved by GLM-based hemodynamic estimation since it allowed only few fluctuations around the canonical values of the activation delay ($TTP_c = 5s$) and duration ($FWHM_c = 5.39s$) depicted as green lines in
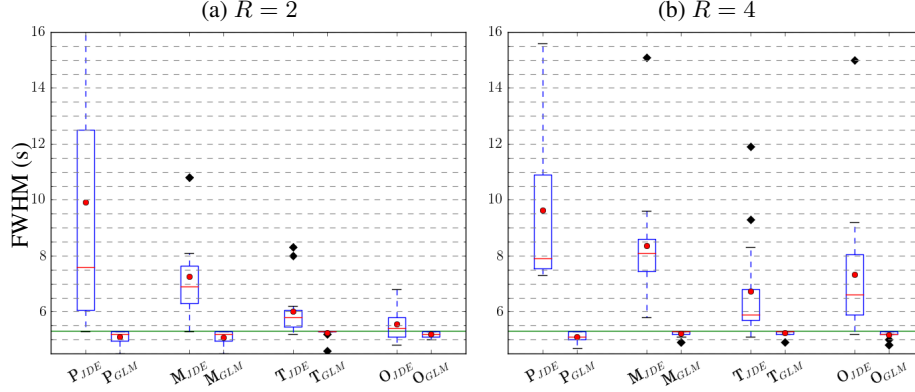
Figure 14: Box-and-whiskers plots (median and mean values shown as red line (–) and red dot (○), respectively) of FWHM extracted from JDE-based and GLM-based HRF estimates according to Eq. (6), for $R = 2$ **(a)** and $R = 4$ **(b)**. FWHM values are plotted next to another region for the parietal (P_ψ), motor (M_ψ), temporal (T_ψ) and occipital (O_ψ) regions with $\psi = \{\text{JDE}, \text{GLM}\}$. The green line corresponds to the canonical FWHM i.e. to 5.39s.

Figs. 13-14. As regards the activation duration, this result was expected since the GLM used here did not include any dispersion-related regressor to account for some FWHM fluctuation. Hence, the GLM-based FWHM estimates showed only little variability between regions. In contrast, JDE-based FWHM estimates were much more variable since FIR modeling allowed fluctuations of activation duration. The reader will find in the Supplementary Materials (cf. Tab. C.4) the group-average TTP and FWHM parameters reported for each region, intra-subject analysis and both acceleration factors, as well as grand mean and standard deviation over regions.

5.2.3. Inference-specific statistical analysis of hemodynamics parameters

In a first analysis of the subject-dependent parameters estimates $(\text{TTP}_{r,s}^{R,\psi})$ and $(\text{FWHM}_{r,s}^{R,\psi})$, we performed several 2-way repeated measures ANOVAs involving two factors (acceleration factor $R = (2, 4)$ and region $r \in \Upsilon$).These results are shown in the Supplementary Materials (see Appendix C). We observed a significant region effect for all intra-subject analyses and all parameters of interest (TTP or FWHM). To localize which brain region(s) drove this significant region in our ANOVA results, we performed two-tailed one-sample non-parametric t-tests on the inference-specific hemodynamic parameter estimates $\text{TTP}_{r,s}^{R,\psi}$ and $\text{FWHM}_{r,s}^{R,\psi}$ in which the null hypothesis was given by the corresponding hemodynamic canonical parameter:

$$
\begin{aligned}
&\forall r \in \Upsilon, R = (2, 4), \\
&\psi \in \{\text{JDE}, \text{GLM}\},
\end{aligned}
\quad
\left\{
\begin{array}{lll}
\text{H}_{0\,r,\text{TTP}}^{R,\psi} & : & \overline{\text{TTP}}_r^{R,\psi} = \text{TTP}_\text{c} = 5\,\text{s} \\
\text{H}_{0\,r,\text{FWHM}}^{R,\psi} & : & \overline{\text{FWHM}}_r^{R,\psi} = \text{FWHM}_\text{c} = 5.39\,\text{s}\,.
\end{array}
\right.
\tag{8}
$$

As the sample size is relatively small ($S = 15$ subjects), we calibrated the group-level distribution under the null hypothesis by permutations to avoid normality assumptions on each hemodynamic parameter (see Appendix D in Supplemen-

Table 4: Corrected p-values of nonparametric two-tailed one-sample t-test associated with the null hypotheses formulated in Eq. (8) for $\psi = \{\text{JDE}, \text{GLM}\}$, $r \in \Upsilon$ and $R = (2, 4)$. Significant p-values ($p_{\text{val}} < 0.05$) appear in **bold font**.

		TTP (s)		FWHM (s)	
	Region r	$\overline{\text{TTP}}_r^{R,\text{JDE}}$	$\overline{\text{TTP}}_r^{R,\text{GLM}}$	$\overline{\text{FWHM}}_r^{R,\text{JDE}}$	$\overline{\text{FWHM}}_r^{R,\text{GLM}}$
$R = 2$	Parietal	**4.88 10^{-4}**	1	**< 10^{-5}**	0.03
	Motor	**4.88 10^{-4}**	0.15	**< 10^{-5}**	0.03
	Temporal	0.064	0.16	**0.002**	1
	Occipital	**0.039**	0.13	0.95	0.06
$R = 4$	Parietal	**4.88 10^{-4}**	1	**< 10^{-5}**	0.03
	Motor	**< 10^{-5}**	**0.008**	**< 10^{-5}**	0.5
	Temporal	0.075	**0.049**	**4.88 10^{-4}**	0.25
	Occipital	0.14	1	**4.88 10^{-4}**	0.06

tary Materials). Two-tailed corrected p-values were considered in the following tests which eventually amounts to comparing $p_{\text{val}}^{\text{right}} = \Pr(T > T_{\text{obs}})$ with 0.05 because $\Pr(T > T_{\text{obs}}) = \Pr(T < -T_{\text{obs}})$ owing to the symmetry of the distribution under the null hypothesis. Here, for the sake of simplicity, we used Bonferroni correction to correct for multiple comparisons over regions-of-interest (see Appendix D for details). Our results are reported in Tab. 4. To make sure that this correction was not too conservative, we have checked with *one-tailed tests* that the alternative consisting of calibrating the distribution of the maximum t-value under the null hypothesis over the four regions of interest gave consistent p-values.

As regards the JDE-based TTP estimates for $R = 2$, the motor and parietal regions showed significant deviations to canonical hemodynamic delay. However, the significance threshold was also exceeded in the occipital region mainly because of the very peaky TTP distribution (tiny variance) in this area (see Fig. 13(a)). Indeed, the group-level mean value (see Tab. C.4) indicated a delay (5.4s) close to the canonical one (5s). When degrading the SNR ($R = 4$), the JDE-based TTP distribution became wider and not surprisingly only the TTP in motor and parietal regions was significant. Concerning the JDE-based FWHM estimates, significance was achieved in the motor and parietal regions as shown in Fig. 14 for both R values. In contrast, at high SNR, the FWHM distribution in temporal and occipital regions remained compatible with the canonical activation duration and the null hypothesis in Eq. (8) was not rejected. However, for $R = 4$, as illustrated in Fig. 14(b), we observed a shift towards larger activation durations and reported more spread distributions making the tests significant in the temporal and occipital cortices. In summary, these results were consistent with the higher hemodynamic variability already reported for motor and parietal regions.

GLM-based inference of TTP parameters provided us with estimates close to the canonical value at least for $R = 2$. As a result, the statistical tests reported in Tab. 4 did not pass the significance threshold. However, for $R = 4$ a statistically significant discrepancy to the canonical hemodynamic delay was found in the motor and temporal regions because slower delays were estimated in most individuals. In contrast, GLM-based inference of activation duration did not depart from the expected canonical

FWHM since no test survived to thresholding at $p_{\text{val}} < 0.05$ after Bonferroni correction. The reason lies in the basis set we a priori selected to capture hemodynamic fluctuations, more with respect to the delay by embedding the temporal derivative of the canonical HRF than with respect to activation duration by excluding its dispersion derivative. Our results thus emphasize that parametric inference of hemodynamic parameters was biased when the basis set was not flexible enough.

To summarize, an accurate estimation of the TTP parameter is critical for activation detection since, by definition, its position is associated with the maximum BOLD effect, as shown for the parietal region in Tab. 3 on the **C - S** contrast. However, the situation is less critical with respect to a wrong extraction of the FWHM parameter if the TTP parameter is well recovered, as shown on the **Lc - Rc** in the motor region (Tab. 2).

5.2.4. Statistical comparison of JDE and GLM-based inference schemes

In a second analysis, we focus on the differences of parameter estimates provided by GLM and JDE inference schemes:

$$\begin{array}{cc} \forall r \in \Upsilon, R = (2,4) \\ \forall s = 1 : S \end{array} \quad \left\{ \begin{array}{ccc} \delta\text{TTP}_{r,s}^{R} & = & \text{TTP}_{r,s}^{R,\text{JDE}} - \text{TTP}_{r,s}^{R,\text{GLM}} \\ \delta\text{FWHM}_{r,s}^{R} & = & \text{FWHM}_{r,s}^{R,\text{JDE}} - \text{FWHM}_{r,s}^{R,\text{GLM}} \end{array} \right. .$$

In two-way repeated measures ANOVAs involving the acceleration parameter and the region as main factors, the region effect appeared to induce significant variations of TTP and FWHM differences across methods (see Appendix C for further details). Hereafter, to localize which regions were responsible for these significant differences, we performed two-tailed one-sample non-parametric t-tests on the intra-subject differences $\delta\text{TTP}_{r,s}^{R}$ and $\delta\text{FWHM}_{r,s}^{R}$ with the following null hypotheses on the group-average differences $\delta\overline{\text{TTP}}_{r}^{R}$ and $\delta\overline{\text{FWHM}}_{r}^{R}$, respectively:

$$\forall r \in \Upsilon, R = (2,4), \quad \left\{ \begin{array}{ccc} \widetilde{\text{H}}_{0r,\text{TTP}}^{R} & : & \delta\overline{\text{TTP}}_{r}^{R} = 0 \\ \widetilde{\text{H}}_{0r,\text{FWHM}}^{R} & : & \delta\overline{\text{FWHM}}_{r}^{R} = 0 \end{array} \right. . \qquad (9)$$

Our results are summarized in Tab. 5. As regards the tests involving the TTP estimates, we brought evidence that for both values of R only the parietal region contributed to the major region effect. This is consistent with the TTP histograms shown in Fig. 13 that depict close group-level mean values in the temporal and occipital regions as well as in the motor region for $R = 4$ only. In contrast, regarding the tests involving the FWHM estimates, the two intra-subject analysis schemes gave statistically significant differences except in the occipital region for $R = 2$. The reason lies in the ability of JDE-based inference to estimate longer activation durations that strongly depart from the canonical value in most regions. The reliability of this finding was confirmed by the reproducibility of the test significance for both values of R in three regions out of four.

To extend the investigation of the hemodynamic variability and assess the impact of the spatial aggregation hypothesis involved in the JDE formation compared to single temporal regularization of FIR HRF inference [11], we performed a thorough comparison in three parcels of interest, namely the temporal, motor and parietal re-

Table 5: Corrected p-values of nonparametric two-tailed one-sample t-tests associated with the null hypotheses formulated in Eq. (9) for all regions $r \in \Upsilon = \{P, M, T, O\}$ and for $R = (2, 4)$. Significant tests ($p_{\mathrm{val}} < 0.05$) appear in **bold font**.

	Region r	TTP (s)	FWHM (s)
$R = 2$	Parietal	**0.002**	$\mathbf{10^{-5}}$
	Motor	0.088	$\mathbf{2.4\ 10^{-4}}$
	Temporal	1	**0.003**
	Occipital	**0.014**	0.32
$R = 4$	Parietal	$\mathbf{2.4\ 10^{-3}}$	$\mathbf{10^{-5}}$
	Motor	0.22	$\mathbf{10^{-5}}$
	Temporal	1	$\mathbf{4.88\ 10^{-4}}$
	Occipital	0.06	$\mathbf{4.88\ 10^{-4}}$

gions (see Appendix B for details). The spatially average RFIR HRF estimates are close to the JDE-based HRF shape in the parietal region while some discrepancy was noticed in the temporal and motor regions. In any case, RFIR HRF shapes show a large spatial variability, especially in the parietal region.

6. Discussion

The current study has investigated different BOLD signal modeling and their consequences at the population-level in terms of sensitivity for detecting evoked brain activity as well as in terms of discrepancy between the estimated and canonical hemodynamic response. As shown by our results, both issues are tightly coupled since the latter impacts the former.

This question has been partially addressed in the literature [30, 29, 45] but not up to the extent we covered here by conducting group-level statistical analyses spanning multiple brain regions as well as assessing the impact of SNR fluctuations in parallel imaging. Indeed, the group-level hemodynamic study [45] was restricted to visual areas. In [29], a performance comparison of HRF estimation techniques has been performed by looking at its population level impact on brain activity detection but only for two brain regions. Among the compared models, these authors have shown that the smooth FIR (or RFIR in Appendix B) and the parametric Inverse Logistic (IL) decomposition (introduced in [30] play a particular role since they guarantee an accurate and nearly unbiased recovery of various HRF shapes irrespective of the induced modulations of hemodynamic properties (duration, delay, return to baseline, etc.). As a consequence, they provide the most sensitive results for detecting evoked brain activity. On real fMRI data sets, the same authors have compared the canonical HRF with time derivative, IL and RFIR models in the right dorsal anterior cingulate cortex and second somato-sensory cortex [29]. Large mismodeling was observed in these regions using a classical GLM and only the IL models retrieved significant activation.

In the present paper, we made use of the temporally smooth and spatially aggregated FIR modeling involved in the JDE framework which ensures improved robustness for HRF recovery due to the increased SNR caused by aggregating all time series

in a given parcel, as emphasized in Appendix B. Compared to [29, 45], we have adopted a more general viewpoint since we have investigated regional hemodynamic variability more deeply using a Localizer experimental paradigm [42]. To this end, we have extracted activation delay and activation duration in occipital, temporal, motor and parietal regions. To go one step further and disentangle the effects of spatially adaptive regularization and parcelwise HRF identification, we have also considered the intermediate SAGLM approach, which sets the hemodynamic filter to its canonical version while allowing the automatic tuning of spatial regularization as provided by in the JDE formalism. This avoids any dependence on the experimenter's tuning but requires a huge computational burden in the Markov chain Monte Carlo context. To solve this issue, a more efficient inference scheme based on variational approximations has been developed within the JDE framework [9]. Moreover, parcelwise inference i.e. independent inference across all parcels enables parallel computing, as proposed in the PyHRF[6] package.

First, the comparison of the three evaluated methods on group-level artificial fMRI data involving canonical or non-canonical HRFs highlighted an efficiency gradient. At a low SNR, the GLM approach performed better in terms of detection sensitivity in the canonical case, SAGLM being slightly less efficient and JDE being the worse. In the non-canonical case, the sorting was the converse: JDE-based analysis performed better than the GLM-based one and SAGLM-based inference still yielding intermediate efficiency. These simulations illustrated on a controlled situation the impact of estimating the HRF and also the sole impact of the adaptive spatial regularization. More thorough artificial setups could be investigated and explore different noise levels or inter-subject anatomo-functional variability.

On real fMRI data sets and with regards to HRF estimation, our results have suggested a gradient between sensory, motor and more cognitive regions in terms of discrepancy with respect to the canonical shape. Indeed, in occipital and temporal areas, the most powerful intra-subject BOLD signal modeling relies on the canonical HRF shape (and potentially on its temporal derivative), in agreement with the literature [4] and as confirmed in Appendix A.Our analysis confirmed that the fitted nonparametric HRF in JDE closely matches the canonical filter in these regions. Hence, the estimation in the JDE framework was performed at the expense of the detection power. In the motor region, we observed an intermediate hemodynamic variability that can still be accounted for by the GLM including its temporal derivative. Indeed, the SAGLM approach yielded the most significant evoked activity for the subject's motor response (see Subsection 5.1.1). In contrast to [29, 45], these results were obtained with a subject-specific and parcel-dependent tuning of spatial regularization which guarantees a fair adaptability to the intrinsic spatial correlation of the data, as well as an automatic setting of the temporal regularization parameter involved in the smooth FIR fitting, as initially proposed in [11, 35]. Last, in the parietal region involved in the processing of mental calculations, we found a statistically significant discrepancy with respect to the canonical HRF shape in conjunction with a large between-subject variability. We demonstrated that the JDE-based modeling and estimation of the HRF time course have

---

[6]http://pyhrf.org

a positive impact on the recovery of evoked brain activity in response to cognitive tasks like mental calculations. Indeed, the JDE approach recovered the most significant results in this area, noticeably when the SNR is degraded by a larger acceleration factor in parallel imaging ($R = 4$), whereas the GLM-based hemodynamic parameters did not capture this variability.

By extracting TTP and FWHM parameter estimates from JDE-based HRF inference, we have analyzed several sources of hemodynamic variability among which the region and subject factors explain the vast majority of hemodynamic parameter variance as compared to the change induced by varying the SNR. This confirms the specialization of hemodynamic features with respect to the brain region. Complementary nonparametric statistical tests were performed considering the canonical TTP and FWHM parameters as the ground truth under the null hypothesis. We found that these null hypotheses were rejected in the parietal and motor regions, but not in the occipital and temporal ones where the actual HRF is close to the canonical shape.

Our results were obtained assuming that the fMRI data sets have been correctly decomposed into functionally homogeneous parcels, i.e. into territories where the parcel-specific HRF shape assumption is tenable. Our parcellation procedure was carried out prior to parcelwise JDE inference and its outcome may change the sensitivity of detection, as illustrated in [48]. The anatomo-functional parcellation method we used here has proved its robustness and optimality in the GLM framework [47]. However, many improvements could be incorporated in the near future. First, the use of a multi-subject parcellation could contribute to a gain in group-level reliability as shown in [46] and would make parcelwise group-level analysis of hemodynamics properties feasible. Second, the parcellation could also be updated according to the estimated hemodynamic parameters by grouping neighboring parcels with similar properties. A first attempt in this direction has recently been proposed in [7] and shown the gain in statistical sensitivity.

The current study relied on the modeling choices performed in the JDE framework. For this reason, our study did not offer the possibility to probe the between-condition hemodynamic variability within a given parcel. To address this issue, we performed some comparisons with condition-specific RFIR HRF shape estimates in Appendix B, as implemented in [11]. We have shown that the assumption of within-parcel homogeneous HRF across experimental conditions is tenable is the parietal region while it is more questionable in the motor one. However, this assumption allows the recovery of physiologically-plausible HRF profiles. Indeed, voxelwise RFIR estimates may deliver artifactual shapes for some experimental conditions, noticeably in the tail part because of the poor SNR. Nonetheless, it is still interesting to extract TTP or FWHM differences between conditions since the vascular part remains the same across stimulus types. As a consequence, small temporal differences in condition-specific TTP or FWHM might reflect a differential neural activity either due to the adaptive tuning of neuronal efficacy or to the recruitment of different neuronal populations. The JDE should be extended to answer this issue and for doing so one future direction might consist of first detecting parcel-specific *relevant* conditions as proposed in [3] and then associating each relevant condition with a specific HRF shape.

Another debatable modeling choice concerns the linearity and time-invariance of the BOLD signal model, which is nevertheless justified by the type of paradigm that we

29

used and in which we did not a priori expect non-linearities and non-stationarities of the BOLD response between successive trials. In other experimental situations involving for instance the rapid repetition of the same stimulus, neurovascular habituation effects also known as the repetition-suppression phenomenon in the neuroscientific literature [15] may likely occur in the recorded fMRI time series. As shown in [12], The JDE framework adapts to such situation by resorting to a parsimonious time-varying modeling of the NRLs which depends on the inter-stimulus interval (ISI) between successive trials through an hyperbolic parametric model. A voxel-specific habituation speed governs the activity decrease across trials. However, this more complex modeling induces a more sophisticated inference scheme and thus a more expensive computation cost. In the same vein, a saturation effect could be modeled for short ISIs (< 1s) [50] using for instance a logarithmic relation across trials.

The results reported in this study concern young healthy adults, but in the near future other populations of interest will be targeted, such as infants, children, elderly volunteers or patients suffering from several neurodegenerative diseases involving the neurovascular system (e.g. Alzheimer, Stroke, CADASIL). These contexts will allow us to probe any deviation to the canonical hemodynamic properties in specific brain regions involved in the brain maturation of neurodegenerative process.

7. Conclusion

In this paper, we assessed the group-level impact of intra-subject BOLD signal modeling with respect to the issues of evoked activity detection and hemodynamics recovery using two SNR-varying parallel imaging fMRI data sets on the same cohort. Our study relied on a fair comparison between classical GLM-based inference and the recently proposed JDE framework, which optimally combines smooth non-parametric HRF estimation with adaptive spatial regularization of evoked activity. To disambiguate the effect of each contribution, we also investigated the performance of the intermediate SAGLM approach that discards HRF estimation. With regards to detection of evoked activity, our results emphasized the price to pay for robust HRF shape identification in primary sensory regions (visual, auditory) where the estimated hemodynamic filter closely matched the canonical HRF. In motor regions, the SAGLM approach outperformed its alternatives in terms of statistical sensitivity. Finally, JDE-based inference yielded most significant results in the parietal region. Also, we statistically demonstrated the link between such results and the underlying regional variability of hemodynamic parameters, which appeared stronger in the parietal region and thus explained our findings in detection.

References

[1] Afonso, D., Sanches, J., Lauterbach, M. H., Apr. 2008. Joint Bayesian detection of brain activated regions and local hrf estimation in functional MRI. In: 33th Proc. IEEE ICASSP. Las Vegas, Ultrasonics, pp. 437–440.

[2] Aguirre, G. K., Zarahn, E., D'Esposito, M., Nov. 1998. The variability of human BOLD hemodynamic responses. Neuroimage 8 (4), 360–369.

[3] Bakhous, C., Forbes, F., Vincent, T., Chaari, L., Dojat, M., Ciuciu, P., May 2012. Adaptive experimental condition selection in event-related fMRI. In: 9th Proc. IEEE ISBI. Barcelona, Spain, pp. 1755–1758.

[4] Boynton, G. M., Engel, S. A., Glover, G. H., Heeger, D. J., July 1996. Linear systems analysis of functional magnetic resonance imaging in human V1. J. Neurosci. 16 (13), 4207–4221.

[5] Casanova, R., Ryali, S., Serences, J., Yang, L., Kraft, R., Laurienti, P., Maldjian, J., May 2008. The impact of temporal regularization on estimates of the BOLD hemodynamic response function: a comparative analysis. Neuroimage 40 (4), 1606–1618.

[6] Casanova, R., Yang, L., Hairston, W. D., Laurienti, P. J., Maldjian, J. A., May 2009. Evaluating the impact of spatio-temporal smoothness constraints on the BOLD hemodynamic response function estimation: an analysis based on Tikhonov regularization. Physiol Meas 30 (5), N37–N51.

[7] Chaari, L., Forbes, F., Vincent, T., Ciuciu, P., Oct. 2012. Adaptive hemodynamic-informed parcellation of fMRI data in a variational joint detection estimation framework. In: 15th Proc. MICCAI, Part III. LNCS 7512. Springer Verlag, Nice, France, pp. 180–188.

[8] Chaari, L., Forbes, F., Vincent, T., Dojat, M., Ciuciu, P., Sep. 2011. Variational solution to the joint detection estimation of brain activity in fMRI. In: 14th Proc. MICCAI. LNCS 6892 (Part II). Springer Verlag Berlin Heidelberg, Toronto, Canada, pp. 260–268.

[9] Chaari, L., Vincent, T., Forbes, F., Dojat, M., Ciuciu, P., Dec. 2012. Fast joint detection-estimation of evoked brain activity in event-related fMRI using a variational approach. IEEE Trans. Med. Imag. , –.

[10] Ciuciu, P., Idier, J., Roche, A., Pallier, C., Apr. 2004. Outlier detection for robust region-based estimation of the hemodynamic response function in event-related fMRI. In: 2th Proc. IEEE ISBI. Arlington, VA, pp. 392–395.

[11] Ciuciu, P., Poline, J.-B., Marrelec, G., Idier, J., Pallier, C., Benali, H., Oct. 2003. Unsupervised robust non-parametric estimation of the hemodynamic response function for any fMRI experiment. IEEE Trans. Med. Imag. 22 (10), 1235–1251.

[12] Ciuciu, P., Sockeel, S., Vincent, T., Idier, J., Apr. 2009. Modelling the neurovascular habituation effect on fMRI time series. In: 34th Proc. IEEE ICASSP. Taipei, Taiwan, pp. 433–436.

[13] de Pasquale, F., Del Gratta, C., Romani, G. L., Aug. 2008. Empirical Markov Chain Monte Carlo Bayesian analysis of MRI data. Neuroimage 42 (1), 99–111.

[14] Dehaene, S., Spelke, E., Pinel, P., Stanescu, R., Tsivkin, S., 1999. Sources of mathematical thinking: behavioral and brain-imaging evidence. Science 284, 970–974.

[15] Dehaene-Lambertz, G., Dehaene, S., Anton, J.-L., Campagne, A., Ciuciu, P., Dehaene, G. P., Denghien, I., Jobert, A., Le Bihan, D., Sigman, M., Pallier, C., Poline, J.-B., 2006. Functional segregation of cortical language areas by sentence repetition. Hum. Brain Mapp. 27, 360–371.

[16] Flandin, G., Penny, W. D., Feb. 2007. Bayesian fMRI data analysis with sparse spatial basis function priors. Neuroimage 34 (3), 1108–1125.

31

[17] Friston, K., 1998. Imaging neuroscience: Principles or maps? Proc. Natl. Acad. Sci. USA 95, 796–802.

[18] Friston, K., Penny, W., Phillips, C., Kiebel, S., Hinton, G., Ashburner, J., 2002. Classical and bayesian inference in neuroimaging: Theory. Neuroimage 16 (2), 465–483.

[19] Glover, G. H., 1999. Deconvolution of impulse response in event-related BOLD fMRI. Neuroimage 9, 416–429.

[20] Gössl, C., Auer, D. P., Fahrmeir, L., June 2001. Bayesian spatio-temporal modeling of the hemodynamic response function in BOLD fMRI. Biometrics 57, 554–562.

[21] Goutte, C., Nielsen, F. A., Hansen, L. K., Dec. 2000. Modeling the haemodynamic response in fMRI using smooth filters. IEEE Trans. Med. Imag. 19 (12), 1188–1201.

[22] Handwerker, D. A., Ollinger, J., D'Esposito, M., Apr. 2004. Variation of BOLD hemody-namic responses across subjects and brain regions and their effects on statistical analyses. Neuroimage 21, 1639–1651.

[23] Heeger, D., Ress, D., 2002. What does fMRI tell us about neuronal activity? Nature reviews neurosciences 3, 142–151.

[24] Henson, R., Price, C., Rugg, M., Turner, R., Friston, K., 2002. Detecting latency differ-ences in event-related BOLD responses: application to words versus nonwords and initial versus repeated face presentations. Neuroimage 15 (1), 83–97.

[25] Henson, R., Rugg, M., Friston, K., 2001. The choice of basis function in event-related fMRI. In: NeuroImage (HBM'01). Vol. 13. p. 149.

[26] Holmes, A., Blair, R., Watson, J., Ford, I., 1996. Nonparametric analysis of statistic images from functional mapping experiments. J. Cereb. Blood Flow Metab. 16, 7–22.

[27] J. Xia, F. L., Wang, Y. M., June 2009. fMRI analysis through Bayesian variable selection with a spatial prior. In: 6th Proc. IEEE ISBI. Boston, USA, pp. 714–717.

[28] Josephs, O., Turner, R., Friston, K. J., 1997. Event-related fMRI. Hum. Brain Mapp. 5, 243–248.

[29] Lindquist, M., Meng Loh, J., Atlas, L., Wager, T., Mar. 2009. Modeling the hemodynamic response function in fMRI: efficiency, bias and mis-modeling. Neuroimage 45 (1 Suppl), S187–98.

[30] Lindquist, M., Wager, T., Aug. 2006. Validity and power in hemodynamic response mod-eling: a comparison study and a new approach. Human brain mapping 28 (8), 764–84.

[31] Logothetis, N. K., June 2008. What we can do and what we cannot do with fMRI. Nature 453 (7197), 869–878.

[32] Logothetis, N. K., Wandell, B., Oct. 2004. Interpreting the BOLD signal. Annual Review of Physiology 66, 735–769.

[33] Makni, S., Ciuciu, P., Idier, J., Poline, J.-B., Sep. 2005. Joint detection-estimation of brain activity in functional MRI: a multichannel deconvolution solution. IEEE Trans. Signal Pro-cessing 53 (9), 3488–3502.

[34] Makni, S., Idier, J., Vincent, T., Thirion, B., Dehaene-Lambertz, G., Ciuciu, P., July 2008. A fully Bayesian approach to the parcel-based detection-estimation of brain activity in fMRI. Neuroimage 41 (3), 941–969.

[35] Marrelec, G., Ciuciu, P., Pélégrini-Issac, M., Benali, H., Aug. 2004. Estimation of the hemodynamic response function in event-related functional MRI: Bayesian networks as a framework for efficient Bayesian modeling and inference. IEEE Trans. Med. Imag. 23 (8), 959–967.

[36] Miezin, F. M., Maccotta, L., Ollinger, J. M., Petersen, S. E., Buckner, R. L., 2000. Characterizing the hemodynamic response: effects of presentation rate, sampling procedure, and the possibility of ordering brain activity based on relative timing. Neuroimage 11, 735–759.

[37] Mériaux, S., Roche, A., Dehaene-Lambertz, G., Thirion, B., Poline, J.-B., May 2006. Combined permutation test and mixed-effect model for group average analysis in fMRI. Hum. Brain Mapp. 27 (5), 402–410.

[38] Neumann, J., Lohmann, G., 2003. Bayesian second-level analysis of functional magnetic resonance images. Neuroimage 20 (2), 1346–1355.

[39] Nichols, T., Holmes, A., 2002. Nonparametric permutation tests for functional neuroimaging: A primer with examples. Hum. Brain Mapp. 15, 1–25.

[40] Ogawa, S., Lee, T., Kay, A., Tank, D., 1990. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. Proc. Natl. Acad. Sci. USA 87 (24), 9868–9872.

[41] Penny, W. D., Trujillo-Barreto, N., Friston, K. J., 2005. Bayesian fMRI time series analysis with spatial priors. Neuroimage 23 (2), 350–362.

[42] Pinel, P., Thirion, B., Mériaux, S., Jobert, A., Serres, J., Le Bihan, D., Poline, J.-B., Dehaene, S., Oct. 2007. Fast reproducible identification and large-scale databasing of individual functional cognitive networks. BMC Neurosci. 8 (1), 91.

[43] Pruessmann, K. P., Weiger, M., Scheidegger, M., Boesiger, P., Nov 1999. SENSE: Sensitivity encoding for fast MRI. Magn. Reson. Med. 42, 952–962.

[44] Roche, A., Mériaux, S., Keller, M., Thirion, B., 2007. Mixed-effects statistics for group analysis in fMRI: A nonparametric maximum likelihood approach. Neuroimage 38, 501–510.

[45] Steffener, J., Tabert, M., Reuben, A., Stern, Y., 2010. Investigating hemodynamic response variability at the group level using basis functions. NeuroImage 49 (3), 2113–22.

[46] Thirion, B., Flandin, G., Pinel, P., Roche, A., Ciuciu, P., Poline, J.-B., Aug. 2006. Dealing with the shortcomings of spatial normalization: Multi-subject parcellation of fMRI datasets. Hum. Brain Mapp. 27 (8), 678–693.

[47] Thyreau, B., Thirion, B., Flandin, G., Poline, J.-B., May 2006. Anatomo-functional description of the brain: a probabilistic approach. In: Proc. 31th Proc. IEEE ICASSP. Vol. V. Toulouse, France, pp. 1109–1112.

[48] Vincent, T., Ciuciu, P., Thirion, B., May 2008. Sensitivity analysis of parcellation in the joint detection-estimation of brain activity in fMRI. In: 5th Proc. IEEE ISBI. Paris, France, pp. 568–571.

[49] Vincent, T., Risser, L., Ciuciu, P., Apr. 2010. Spatially adaptive mixture modeling for analysis of within-subject fMRI time series. IEEE Trans. Med. Imag. 29 (4), 1059–1074.

[50] Wager, T. D., Vazquez, A., Hernandez, L., Noll, D. C., Mar. 2005. Accounting for nonlinear BOLD effects in fMRI: parameter estimates and a model for prediction in rapid event-related studies. Neuroimage 25 (1), 206–218.

[51] Wang, J., Zhu, H., Fan, J., Giovanello, K., Lin, W., Sep. 2011. Adaptively and spatially estimating hemodynamic response functions in fMRI. In: 14th Proc. MICCAI. LNCS (Part II). Springer Verlag Berlin Heidelberg, Toronto, Canada, pp. 269–276.

[52] Woolrich, M., Jenkinson, M., Brady, J., Smith, S., Feb. 2004. Fully Bayesian spatio-temporal modelling of fMRI data. IEEE Trans. Med. Imag. 23 (2), 213–231.

[53] Zarahn, E., Aguirre, G. K., D'Esposito, M., January 1997. A trial-based experimental design for fmr. Neuroimage 6 (2), 122–138.