Comparison of stochastic and variational solutions to ASL fMRI data analysis

Aina Frau-Pascual[1,3], Florence Forbes[1], and Philippe Ciuciu[2,3]

[1]INRIA, Univ. Grenoble Alpes, LJK, Grenoble, France
[2]CEA/DSV/I^{2}BM/NeuroSpin, Bât. 145, F-91191 Gif-sur-Yvette, France
[3]INRIA/CEA Parietal team, NeuroSpin, Bât. 145, F-91191 Gif-sur-Yvette, France

Abstract. Functional Arterial Spin Labeling (fASL) MRI can provide a quantitative measurement of changes of cerebral blood flow induced by stimulation or task performance. fASL data is commonly analysed using a general linear model (GLM) with regressors based on the canonical hemodynamic response function. In this work, we consider instead a joint detection-estimation (JDE) framework which has the advantage of allowing the extraction of both task-related perfusion and hemodynamic responses not restricted to canonical shapes. Previous JDE attempts for ASL have been based on computer intensive sampling (MCMC) methods. Our contribution is to provide a comparison with an alternative variational expectation-maximization (VEM) algorithm on synthetic and real data.

1 Introduction

Arterial Spin Labeling (ASL) [**?**] is a MRI modality that is able to provide a quantitative measurement of cerebral blood flow (CBF). ASL data consists of alternating pairs of control and magnetically tagged ("tag") images. Local CBF or perfusion changes can be measured by considering the "control-tag" difference. Many control-tag pairs (> 50) need to be acquired to compensate for the low Signal-to-Noise Ratio (SNR) of this difference ($\simeq$1-2%). Aside from its main use in static measurements, ASL has also been used in functional MRI (functional ASL or fASL) as an alternative quantitative imaging technique to the standard blood-oxygen-level-dependent (BOLD) [**?**] contrast imaging modality. fASL can provide more specific information about brain function but its lower SNR and temporal resolution make its analysis more challenging. The standard approach for fASL data analysis is the general linear model (GLM) [**?**,**?**]. It relies on the canonical hemodynamic response function (HRF) for defining the hemodynamic and perfusion-related regressors, although the HRF has been calibrated in BOLD experiments only. Moreover, there has been strong evidence in the literature for space-varying and subject-specific HRF shape [**?**,**?**]. To deal with this issue, a joint detection-estimation (JDE) framework, originally developed for BOLD data analysis [**?**,**?**], has been extended in [**?**] to allow for the extraction of both task-related perfusion and hemodynamic responses while recovering perfusion-related and BOLD-related maps of evoked activity. Previous JDE implementations for

ASL have been based on Markov Chain Monte Carlo (MCMC) techniques. In this work, following the spirit of [?], we provide an alternative solution based on the variational expectation-maximization (VEM) algorithm and compare its performance to its MCMC alternative. In both solutions, prior knowledge is introduced on the relationship between perfusion and hemodynamic responses (resp. PRF and HRF) derived from physiological models [?]. This prior allows us to benefit from a better estimation of the HRF due to higher SNR of the BOLD component in the ASL signal to inform the PRF estimation. As already observed in [?] for the BOLD case, JDE-VEM provides comparable results to JDE-MCMC for a lower computational load. On top of that, JDE-VEM is more convenient to handle identifiability constraints (eg, unit norm or positivity constraint of the response shapes for solving the scale ambiguity issue in Eq. (??)).

2   Joint Detection Estimation model for fASL data

The JDE formalism is a region-based approach that considers hemodynamically homogeneous regions. In a region $\mathcal{P}$ comprising J voxels, the ASL JDE model [?,?] defines voxel-specific ASL time series as the linear superposition of the M task-induced (or stimulus-specific) responses:

$$\forall j \in \mathcal{P}, \quad \boldsymbol{y}_j = \sum_{m=1}^{M} \Big[\underbrace{a_j^m \boldsymbol{X}^m \boldsymbol{h}}_{(a)} + \underbrace{c_j^m \boldsymbol{W} \boldsymbol{X}^m \boldsymbol{g}}_{(b)} \Big] + \underbrace{\alpha_j \boldsymbol{w}}_{(c)} + \underbrace{\boldsymbol{P} \boldsymbol{\ell}_j}_{(d)} + \underbrace{\boldsymbol{b}_j}_{(e)} \quad (1)$$

Each time series $\boldsymbol{y}_j \in \mathbb{R}^N$ is decomposed into (a) task-related hemodynamic and (b) perfusion components, (c) a perfusion baseline term $\alpha_j \boldsymbol{w}^1$ which completes the modelling of the perfusion component, (d) a drift component $\boldsymbol{P}\boldsymbol{\ell}_j$ and (e) a noise term, assumed white Gaussian with variance σ_j^2. The control/tag effect is modelled in Eq. (??) by making use of $\boldsymbol{w}$ and $\boldsymbol{W} = \mathrm{diag}(\boldsymbol{w})$. Vectors $\boldsymbol{h}$ and $\boldsymbol{g}$ represent the D-dimensional ($D < N$) unknown HRF and PRF shapes, constant within $\mathcal{P}$. The magnitudes of activation or response levels for hemodynamic and perfusion components are $\boldsymbol{a} = \{a_j^m\}$ and $\boldsymbol{c} = \{c_j^m\}$ and referred to as HRLs and PRLs (hemodynamic and perfusion response levels) hereafter. $\boldsymbol{X} \in \mathbb{R}^{N \times D}$ is a binary matrix that encodes the lagged onset stimuli. The response levels are assumed to follow different Gaussian mixture models but governed by M common binary hidden Markov random fields $\boldsymbol{q}^m$ with $\boldsymbol{q}^m = \{q_j^m, j \in \mathcal{P}\}$ encoding voxels activation states for each experimental condition m. HRLs and PRLs are assumed independent conditionally to these activation labels $\boldsymbol{q} = \{\boldsymbol{q}^m, m = 1 : M\}$. For further details, please refer to [?].

As already mentioned, the perfusion component in the ASL signal has a very low SNR owing to its small size captured by the "control-tag" subtraction. To address this issue, we make use of a link derived in [?] from physiological models between the two responses $\boldsymbol{g} = \boldsymbol{\Omega} \boldsymbol{h}$. As a difference with [?], we then consider

---

[1] Vector $\boldsymbol{w}$ is N-dimensional such that $w_{t_n} = 1/2$ if t_n is even (control) and $w_{t_n} = -1/2$ otherwise (tagged).

that HRF and PRF shapes follow prior Gaussian distributions $\boldsymbol{h} \sim \mathcal{N}(\boldsymbol{0}, v_{\boldsymbol{h}} \boldsymbol{\Sigma_h})$ and $\boldsymbol{g}|\boldsymbol{h} \sim \mathcal{N}(\boldsymbol{\Omega h}, v_{\boldsymbol{g}} \boldsymbol{\Sigma_g})$, with covariance matrices $\boldsymbol{\Sigma}_h$ and $\boldsymbol{\Sigma}_g$ encoding a constraint on the second order derivatives so as to account for temporal smoothness of $\boldsymbol{h}$ and $\boldsymbol{g}$, respectively. We also consider constraints on the response functions to enforce their L_2-norm to be unitary.

3   Variational EM estimation

In a first attempt to estimate the ASL JDE model, an intensive sampling MCMC procedure has been used in a Bayesian setting [?]. This provides an elegant way to estimate the missing model variables $\boldsymbol{a} \in \mathcal{A}$, $\boldsymbol{h} \in \mathcal{H}$, $\boldsymbol{c} \in \mathcal{C}$, $\boldsymbol{g} \in \mathcal{G}$, $\boldsymbol{q} \in \mathcal{Q}$ via the sampling of the posterior distribution $p(\boldsymbol{a}, \boldsymbol{h}, \boldsymbol{c}, \boldsymbol{g}, \boldsymbol{q}|\boldsymbol{y})$ whose direct computation or maximization is intractable. When the model involves a lot of missing variables, such an approach is not easy to monitor and not that flexible. Additional information or constraints on the response function shapes (*i.e.,* unit L_2-norm) cannot be easily integrated. Following the lead of [?], we propose an alternative Expectation-Maximization (EM) framework. Let $\mathcal{D}$ be the set of all probability distributions on $\mathcal{A} \times \mathcal{H} \times \mathcal{C} \times \mathcal{G} \times \mathcal{Q}$. EM can be viewed as an alternating maximization procedure of a function F such that for any $\tilde{p} \in \mathcal{D}$, $F(\tilde{p}, \boldsymbol{\theta}) = \mathrm{E}_{\tilde{p}}\big[\log p(\boldsymbol{y}, \boldsymbol{a}, \boldsymbol{h}, \boldsymbol{c}, \boldsymbol{g}, \boldsymbol{q} \; ; \boldsymbol{\theta})\big] + I[\tilde{p}]$ where $I[\tilde{p}] = -\mathrm{E}_{\tilde{p}}\big[\log \tilde{p}(\boldsymbol{a}, \boldsymbol{h}, \boldsymbol{c}, \boldsymbol{g}, \boldsymbol{q})\big]$ is the entropy of $\tilde{p}$, $\mathrm{E}_{\tilde{p}}\big[\big]$ denotes the expectation with respect to $\tilde{p}$ and $\boldsymbol{\theta}$ is the set of parameters. Maximizing function F is equivalent to minimizing the Kullback-Leibler divergence between $\tilde{p}$ and the true posterior of interest $p(\boldsymbol{a}, \boldsymbol{h}, \boldsymbol{c}, \boldsymbol{g}, \boldsymbol{q}|\boldsymbol{y})$. This view of EM has led to a number of variants in which the E-step is solved over a restricted class of probability distributions, $\tilde{\mathcal{D}}$. The variational approach corresponds to choosing $\tilde{\mathcal{D}}$ as the set of distributions that factorize over the set of missing variables: $\tilde{p}(\boldsymbol{a}, \boldsymbol{h}, \boldsymbol{c}, \boldsymbol{g}, \boldsymbol{q}) = \tilde{p}_a(\boldsymbol{a}) \; \tilde{p}_h(\boldsymbol{h}) \; \tilde{p}_c(\boldsymbol{c}) \; \tilde{p}_g(\boldsymbol{g}) \; \tilde{p}_q(\boldsymbol{q})$ where $\tilde{p}_a \in \mathcal{D}_A$, $\tilde{p}_h \in \mathcal{D}_H$, $\tilde{p}_c \in \mathcal{D}_C$, $\tilde{p}_g \in \mathcal{D}_G$ and $\tilde{p}_q \in \mathcal{D}_Q$, the sets of probability distributions on $\mathcal{A}, \mathcal{H}, \mathcal{C}, \mathcal{G}, \mathcal{Q}$ respectively. The E-step becomes an approximate E-step that can be further decomposed into five stages updating the different variables in turn. At iteration (r), with current estimates denoted by $\tilde{p}_a^{(r-1)}, \tilde{p}_h^{(r-1)}, \tilde{p}_c^{(r-1)}, \tilde{p}_g^{(r-1)}, \tilde{p}_q^{(r-1)}$ and $\boldsymbol{\theta}^{(r-1)}$, where $\boldsymbol{\theta} = \big\{\boldsymbol{\alpha}, \boldsymbol{\ell}, \boldsymbol{\sigma}^2, \boldsymbol{\mu}_{a,c}, \boldsymbol{\sigma}_{a,c}, v_h, v_g, \boldsymbol{\beta}\big\}$, the updating formulae are of the form:

$$\textbf{E-H-step:} \;\; \tilde{p}_h^{(r)} = \underset{\tilde{p}_h \in \mathcal{D}_H}{\arg\max} \, F(\tilde{p}_a^{(r-1)} \; \tilde{p}_h \; \tilde{p}_c^{(r-1)} \; \tilde{p}_g^{(r-1)} \; \tilde{p}_q^{(r-1)}; \boldsymbol{\theta}^{(r-1)}) \quad (2)$$

$$\textbf{E-G-step:} \;\; \tilde{p}_g^{(r)} = \underset{\tilde{p}_g \in \mathcal{D}_G}{\arg\max} \, F(\tilde{p}_a^{(r-1)} \; \tilde{p}_h^{(r)} \; \tilde{p}_c^{(r-1)} \; \tilde{p}_g \; \tilde{p}_q^{(r-1)}; \boldsymbol{\theta}^{(r-1)}) \quad (3)$$

with similar expressions for the other steps obtained by permuting the roles of the variables. Hereafter, for the ease of presentation, the (r) and $(r-1)$ superscripts are omitted. In contrast to the standard setting, to introduce normalisation constraints on $\boldsymbol{h}$ and $\boldsymbol{g}$, we modify the sought variational approximation into $\tilde{p} = \tilde{p}_a \; \delta_{\tilde{\boldsymbol{h}}} \; \tilde{p}_c \; \delta_{\tilde{\boldsymbol{g}}} \; \tilde{p}_q$, where the distributions on $\boldsymbol{h}$ and $\boldsymbol{g}$ are replaced by Dirac functions. This reduces the search to pointwise estimates $\tilde{\boldsymbol{h}}$ and $\tilde{\boldsymbol{g}}$. The E-H and

E-G steps in Eqs. (**??**)-(**??**) then yield maximization problems which are easily constrained to account for normalisation:

$$\textbf{E-H:} \quad \tilde{\boldsymbol{h}} = \arg\max_{\boldsymbol{h}} \mathrm{E}_{\tilde{p}_a \tilde{p}_c \tilde{p}_q}\big[\log p(\boldsymbol{h} \mid \boldsymbol{y}, \boldsymbol{a}, \boldsymbol{c}, \tilde{\boldsymbol{g}}, \boldsymbol{q}; \boldsymbol{\theta})\big] \tag{4}$$

$$\textbf{E-G:} \quad \tilde{\boldsymbol{g}} = \arg\max_{\boldsymbol{g}} \mathrm{E}_{\tilde{p}_a \tilde{p}_c \tilde{p}_q}\big[\log p(\boldsymbol{g} \mid \boldsymbol{y}, \boldsymbol{a}, \tilde{\boldsymbol{h}}, \boldsymbol{c}, \boldsymbol{q}; \boldsymbol{\theta})\big] \tag{5}$$

Solving Eqs. (**??**)-(**??**) amounts to minimizing a quadratic function under a quadratic constraint, namely $\|\boldsymbol{h}\|_2^2 = 1$ and $\|\boldsymbol{g}\|_2^2 = 1$ respectively. The other E-steps can be derived from standard expressions in [**?**] replacing expectations over $\boldsymbol{h}$ and $\boldsymbol{g}$ by $\tilde{\boldsymbol{h}}$ and $\tilde{\boldsymbol{g}}$, *e.g.*

$$\textbf{E-Q:} \quad \tilde{p}_q(\boldsymbol{q}) \propto \exp\left(\mathrm{E}_{\tilde{p}_a \tilde{p}_c}\big[\log p(\boldsymbol{q} \mid \boldsymbol{y}, \boldsymbol{a}, \tilde{\boldsymbol{h}}, \boldsymbol{c}, \tilde{\boldsymbol{g}}; \boldsymbol{\theta})\big]\right), \tag{6}$$

with similar expressions for the E-A and E-C steps. The corresponding **M-step** for the update of $\boldsymbol{\theta}$ can be divided into separate M-substeps, as in [**?**]:

$$\boldsymbol{\theta} = \arg\max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}}\big[\mathrm{E}_{\tilde{p}_a \tilde{p}_c}\big[\log p(\boldsymbol{y} \mid \boldsymbol{a}, \tilde{\boldsymbol{h}}, \boldsymbol{c}, \tilde{\boldsymbol{g}}; \boldsymbol{\alpha}, \boldsymbol{\ell}, \boldsymbol{\sigma}^2)\big] + \mathrm{E}_{\tilde{p}_a \tilde{p}_q}\big[\log p(\boldsymbol{a} \mid \boldsymbol{q}; \boldsymbol{\mu}_a, \boldsymbol{\sigma}_a)\big]$$
$$+ \log p(\tilde{\boldsymbol{h}}; v_h) + \mathrm{E}_{\tilde{p}_c \tilde{p}_q}\big[\log p(\boldsymbol{c} \mid \boldsymbol{q}; \boldsymbol{\mu}_c, \boldsymbol{\sigma}_c)\big] + \log p(\tilde{\boldsymbol{g}}; v_g) + \mathrm{E}_{\tilde{p}_q}\big[\log p(\boldsymbol{q}; \boldsymbol{\beta})\big]\big] \,.$$

It follows a VEM procedure in which missing quantities are updated in turn. Compare to an MCMC solution, the variational approach is based on an approximation of the full posterior distribution. The dependencies between te random variables in the JDE model are reduced to dependencies between their moments that appear in the successive E-steps. No theoretical results exist that would guarantee the quality of such an approximation but the performance comparison provided in the next section suggests that VEM is still able to capture enough information amongst the original dependencies.

4   Results

Different data sets have been analysed to assess the performance of the VEM and MCMC approaches: first, artificial data synthesized with the generative model (**??**), and second real data acquired on different individuals from the AINSI initiative (`http://thalie.ujf-grenoble.fr/ainsi`).

4.1   Artificial data

$N = 292$ ASL artificial images (i.e., 146 control/tag pairs) have been simulated using a realistic SNR ($\simeq 3$ dB) according to Eq. (**??**) considering $\boldsymbol{h}$ and $\boldsymbol{g}$ as depicted in Fig.**??**(a)-(b) by dashed lines. To emulate the slow sampling rate of ASL images, Eq. (**??**) was synthesized at $\Delta t = 1$ sec and then undersampled at $TR = 3$ sec. Here, we considered a fast event-related paradigm comprising two conditions ($M = 2$). Drift coefficients and noise realizations were drawn according to $\boldsymbol{\ell}_j \sim \mathcal{N}(0, 10\boldsymbol{I}_O)$ and $\boldsymbol{b}_j \sim \mathcal{N}(0, 2\boldsymbol{I}_N)$, respectively. HRLs were

generated with $(a_j^m | q_j^m = 1) \sim \mathcal{N}(2.2, 0.3)$ for active voxels and $(a_j^m | q_j^m = 0) \sim \mathcal{N}(0, 0.3)$ for inactive ones. To make this synthetic setting realistic, PRLs were generated with a lower contrast than HRLs: $(c_j^m | q_j^m = 1) \sim \mathcal{N}(1.6, 0.3)$ and $(c_j^m | q_j^m = 0) \sim \mathcal{N}(0, 0.3)$. Activation states (assignment variables $\boldsymbol{q}$) are set by a hand-drawn map, as illustrated on the first column maps of Fig. **??**.

Fig. **??**(a-d) shows the HRF and PRF estimates obtained for two different noise levels. Both response functions were well recovered with MCMC and VEM at 3 dB SNR with an acceptable degradation at lower SNR (i.e. 0.5 dB). In the latter case, MCMC recovers slightly better the peak. The labels (activated/non-activated) in Fig. **??** are well recovered with both MCMC and VEM at the higher SNR. At the lower one, both solutions fail to recover accurate label maps. As typical of VEM, labels maps are more *contrasted* than with MCMC which is likely to better estimate variability. Fig. **??** shows the mean squared errors (MSE) for a range of SNR levels. Response functions are well recovered with small root MSE in all cases (Fig. **??**(a)) but with better estimations with MCMC. In contrast, response levels are better recovered with VEM (Fig. **??**(b)). This is consistent with previous comparisons between VEM and MCMC on BOLD signals [**?**].
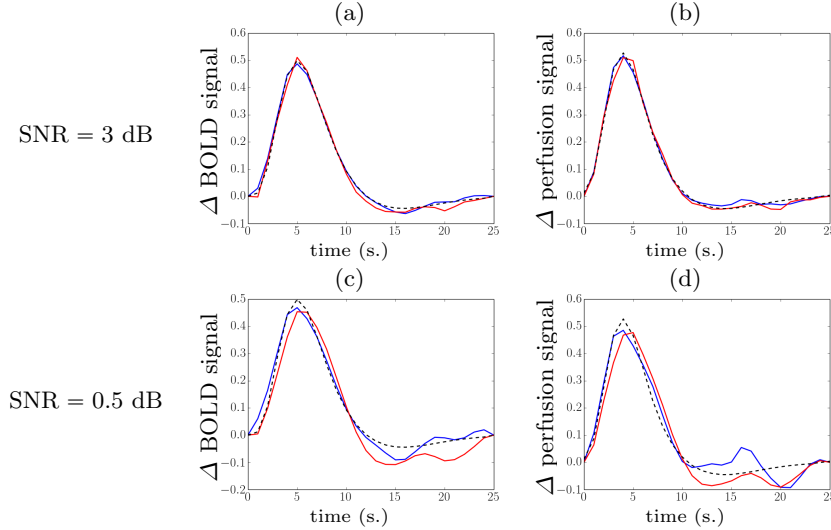

Fig. 1. Artificial data with 2 noise settings: (a, b) SNR = 3 dB, (c, d) SNR = 0.5 dB. Ground-truth response curves (black dashed lines) and estimated hemodynamic (a, c) and perfusion (b, d) response functions with MCMC in blue and VEM in red.

4.2   Real data

Real ASL data were recorded during an experiment designed to map auditory and visual brain functions, which consisted of $N = 291$ scans lasting TR = 3000 ms, with TE = 18 ms, FoV 192 mm, each yielding a 3-D volume composed of $64 \times 64 \times 22$ voxels (spatial resolution of $3 \times 3 \times 7$ mm^3). The tagging scheme used was PICORE Q2T, with $(\text{TI}_1, \text{TI}_2) = (700, 1700)$ ms. A fast event-related
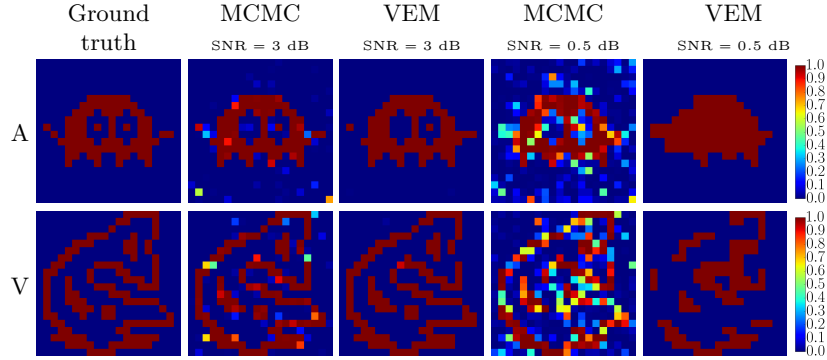
Fig. 2. Results on artificial data for labels q. The probability to be activated is shown for each voxel, for 2 experimental conditions, namely auditory (A) and visual (V) stimuli. The ground truth as well as the MCMC and VEM activation probability estimates are shown in two different SNR scenarios.
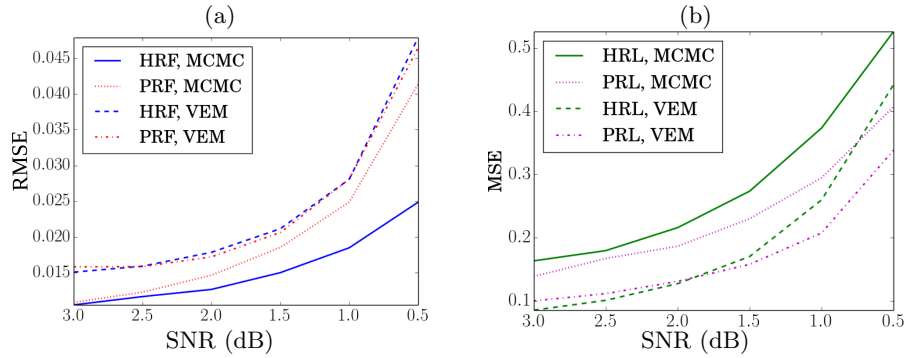

Fig. 3. MSE comparison between MCMC and VEM approaches. (a) Root MSE (RMSE) of the HRF and PRF. (b) Mean over conditions of the MSE of the response levels HRL and PRL.

paradigm comprising sixty auditory and visual stimuli, randomly distributed according to a mean inter-stimulus interval of 5.1 sec, was run during acquisition.

In Fig. **??**, the MCMC and VEM results are shown in the left and right visual and auditory cortices. The HRL maps in Fig. **??**(a) are very similar for the two approaches and for A and V conditions in contrast to the larger variability reported in the PRL maps owing to the lower effect size. Interestingly, the PRL maps yielded by the two algorithms are consistent for the V condition in contrast to what we observed for the A condition. The regions of interest (ROI) in Fig. **??**(b) correspond to the parcels with stronger mean HRL and PRL values for each condition respectively. The HRF and PRF estimates in these ROIs have plausible shapes and both approaches recover similar profiles. For both conditions, the PRF peaks before the HRF, as enforced by the physiological prior.

Regarding computational times, a substantial decrease was observed for VEM as compared to the MCMC solution, which is consistent with previous reports [**?**].

5   Conclusion

A VEM algorithm has been proposed to address the issue of jointly detecting evoked activity and estimating the associated hemodynamic and perfusion responses from functional ASL data. Compared to MCMC, VEM delivers estimations in analytic form for each latent variable. Although the VEM setting remains an approximation, it facilitates the inclusion of additional information such as constraints. Our results demonstrate a good performance of VEM when compared to MCMC at a significantly lower computation time. This suggests VEM as a fast and valid alternative for functional ASL data analysis.
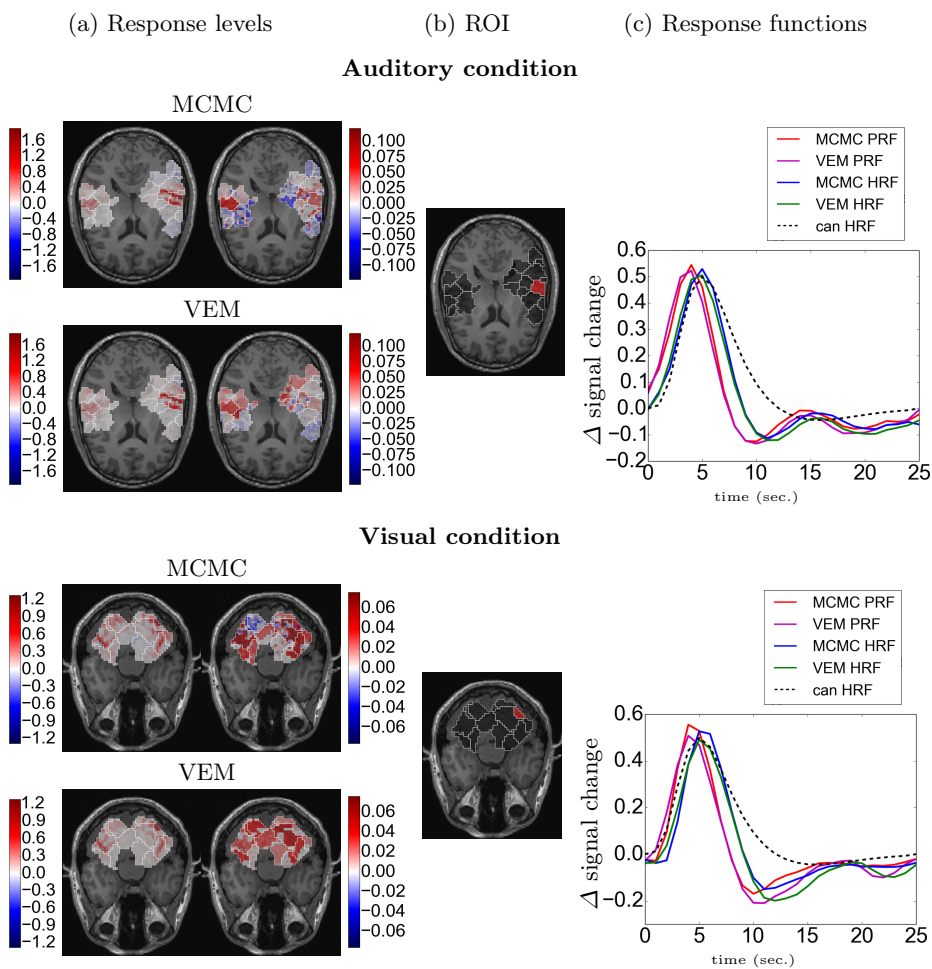
(a) Response levels   (b) ROI   (c) Response functions

Auditory condition

Visual condition

Fig. 4. Results on real fASL data for a single subject of the AINSI database for both conditions: Auditory (top) and Visual (bottom). (a) HRL on the left and PRL on the right; (b) region of interest (ROI) where the response functions in (c) are estimated. In (c) and as indicated in the legend, the red and blue curves represent the PRF and HRF respectively for the MCMC approach, and the magenta and green curves represent the PRF and HRF respectively for the VEM approach. As a reference, we depicted the canonical HRF with a black dashed line.