Variational Solution to the Joint Detection Estimation of Brain Activity in fMRI

Lotfi Chaari[1,2], Florence Forbes[1,2], Thomas Vincent[3], Michel Dojat[2,4], and Philippe Ciuciu[3]

[1] INRIA, MISTIS, Grenoble, France
[2] Grenoble University, LJK, Grenoble, France
[3] CEA/DSV/I²BM/Neurospin, LNAO, Gif-Sur-Yvette, France
[4] INSERM, U836, GIN, Grenoble, France

Abstract. We address the issue of jointly detecting brain activity and estimating underlying brain hemodynamics from functional MRI data. We adopt the so-called Joint Detection Estimation (JDE) framework that takes spatial dependencies between voxels into account. We recast the JDE into a missing data framework and derive a Variational Expectation-Maximization (VEM) algorithm for its inference. It follows a new algorithm that has interesting advantages over the previously used intensive simulation methods (Markov Chain Monte Carlo, MCMC): tests on artificial data show that the VEM-JDE is more robust to model mis-specification while additional tests on real data confirm that it achieves similar performance in much less computation time.

1 Introduction

Functional Magnetic Resonance Imaging (fMRI) is a powerful tool to non-invasively study the relation between a cognitive task and an evoked neural activity through neurovascular coupling and the BOLD effect [10]. To localize which parts of the brain are activated by a given stimulus type, most approaches assume a single canonical *a priori* model for the impulse response of the neurovascular coupling also known as the hemodynamic response function (HRF) [5]. However, there has been evidence that this response can vary in space and between subjects [6, 1] so that both issues of properly detecting evoked activity and estimating the HRF play a central role in fMRI data analysis. They are usually dealt with independently with no possible feedback although they are strongly connected. To account for these sources of hemodynamic variability, a novel approach referred to as the Joint Detection Estimation (JDE) framework has been introduced in [9] and extended in [13] to account for spatial correlation between neighboring voxels in the brain volume. Since robust and accurate HRF estimation can only be achieved in regions that elicit an evoked response to an experimental condition [7], the JDE approach has been defined at an intermediate spatial resolution corresponding to *parcels* in which a fair compromise between homogeneity of the BOLD signal and reproducibility of the HRF shape is achieved. The JDE approach mainly rests upon: *i.)* a non-parametric or

FIR modelling of the HRF at this parcel-level for an unconstrained HRF shape; *ii.)* prior information about the temporal smoothness of the HRF to guarantee a physiologically plausible shape; and *iii.)* the modelling of spatial correlation between neighboring voxels within each parcel using condition-specific discrete hidden Markov fields. In [9, 13], posterior inference is carried out in a Bayesian setting using Markov Chain Monte Carlo (MCMC) methods, which requires fine tuning and is time consuming. Several other attempts to segregate neurological and hemodynamic events from fMRI time series have been proposed (see references in [13]). Among them lies an interesting bilinear dynamical system formulation [8] that deals with unknown HRF and uses Variational Bayes (VB) approximation for tractability. However, from a spatial viewpoint, this work remains univariate and ignores spatial correlation between voxels.

In this paper, we reformulate the complete JDE framework [13] as a missing data problem and propose a simplification of its estimation procedure. Akin to [8], we resort to a variational approximation using a Variational Expectation Maximization (VEM) algorithm in order to derive estimates of the HRF and stimulus-related activity. Experiments on artificial and real data demonstrate the good performance of our VEM algorithm. In particular, we provide a comparison with its MCMC counterpart and show the advantages of VEM both in terms of computation time and robustness to noise model deviations. This potentially increases considerably the impact of the JDE framework and makes its application to fMRI studies in neuroscience easier and more valuable.

2   A Joint Detection-Estimation Model

Capital letters indicate random variables and lower case their realizations. Matrices are denoted with bold upper case letters and the transpose with ^{t}.

Observed and Missing Variables. We first recast the parcel-based JDE model of [9, 13] in a missing data framework. For a given parcel $\mathcal{P}$, the *observed* data is denoted by $Y = \{Y_i, i \in \mathcal{P}\}$ where $Y_i \in \mathbb{R}^N$ is the fMRI time series measured in voxel $i \in \mathcal{P}$ at times $(t_n)_{n=1:N}$, where $t_n = n\,TR$, N being the number of scans and TR, the time of repetition. Additional *unobserved* variables are introduced: 1) The Neural Response Levels (NRLs) $A = \{A_m, m = 1 : M\}$ with $A_m = \{A_{mi}, i \in \mathcal{P}\}$ and M the number of experimental conditions involved in the paradigm. We will also make use of $A_i = [A_{mi}, m = 1 : M]^t$. 2) The HRF shape $H = [H_{d\Delta t}, d = 0 : D]^t \in \mathbb{R}^{D+1}$; 3) The activation class assignments $Z = \{Z_m, m = 1 : M\}$ with $Z_m = \{Z_{mi}, i \in \mathcal{P}\}$ represent the *activation classes* for each voxel, in each of the M experimental conditions. $Z_{mi} = k$ means that voxel i lies in activation class k for the mth experimental condition. The number of classes is here $K = 2$ for activating and non activating voxels. An additional deactivation class ($K = 3$) may actually be added depending on the experiment and all provided formulas are general enough to cover this case. The observed and missing variables are then linked by the following relationship involving additional parameters to be estimated:

$$\forall i \in \mathcal{P}, \quad Y_i = \sum_{m=1}^{M} A_{mi} \boldsymbol{X}_m H + \boldsymbol{P}\ell_i + \varepsilon_i, \tag{1}$$

where $\boldsymbol{X}_m = (x^m_{t_n - d\Delta t})_{n=1:N, d=0:D}$ denotes the $N \times (D+1)$ binary matrix that codes the onsets of the mth experimental condition on a Δt-sampled grid, where Δt is the sampling period of the HRF ($\Delta t < TR$); the ε_i's stand for the noise and are independent and normally distributed in space with $\varepsilon_i \sim \mathcal{N}(0, \boldsymbol{\Gamma}_i^{-1})$, and $\boldsymbol{P}$ is the low frequency orthogonal $N \times L$ matrix which accounts for physiological artifacts. Let $\ell = \{\ell_i, i \in \mathcal{P}\}$ be the set of low frequency drifts, where $\ell_i \in \mathbb{R}^L$ have to be estimated and let $\boldsymbol{\Gamma} = \{\boldsymbol{\Gamma}_i, i \in \mathcal{P}\}$ be the set of all unknown precision matrices (see Section 4 for its definition).

Hierarchical Model of the Complete Data Distribution. Using standard additional assumptions and omitting the dependence on the parameters, the distribution of both the observed and unobserved data writes: $p(y, a, h, z) = p(y \,|\, a, h)\, p(a \,|\, z)\, p(h)\, p(z)$. To fully define the model, we now specify each term. **The $p(y \,|\, a, h)$ term.** From (1), it comes that $p(y \,|\, a, h) = \prod_{i \in \mathcal{P}} p(y_i \,|\, a_i, h)$ with

$$Y_i \,|\, (A_i = a_i, H = h) \sim \mathcal{N}\Big(\textstyle\sum_{m=1}^M a_{mi} \boldsymbol{X}_m h + \boldsymbol{P}\ell_i, \boldsymbol{\Gamma}_i^{-1}\Big).$$

The $p(a \,|\, z)$ term. Akin to [9,13], the NRLs A_{mi} are assumed statistically independent across condition types. The allocation variables Z_{mi} are then introduced to segregate activating voxels from non-activating ones in condition-specific mixture models. Also, the A_{mi}'s are supposed independent in space conditionally on Z_m so that putting together all conditions we get: $p(a \,|\, z) = \prod_{m=1}^M \prod_{i \in \mathcal{P}} p(a_{mi} \,|\, z_{mi})$, where we further assume that $p(A_{mi} \,|\, Z_{mi} = k) = \mathcal{N}(\mu_{mk}, \sigma_{mk}^2)$. The Gaussian parameters are unknown and denoted by $\mu = \{\mu_m, m = 1 : M\}$ with $\mu_m = [\mu_{m1} \ldots \mu_{mK}]^t$ and $\sigma = \{\sigma_m, m = 1 : M\}$ with $\sigma_m = [\sigma_{m1} \ldots \sigma_{mK}]^t$. Also, $k = 1$ is assigned to non-activating voxels with $\mu_{m1} = 0$.

The $p(h)$ term. Akin to [9,13], we introduce constraints in the prior that favor smooth variations in h: $H \sim \mathcal{N}(0, \sigma_h^2 \boldsymbol{R})$ with $\boldsymbol{R} = (\Delta t)^4 \, (\boldsymbol{D}_2^t \boldsymbol{D}_2)^{-1}$ where $\boldsymbol{D}_2$ is the second-order finite difference matrix and σ_h^2 is a parameter to be estimated or fixed. Moreover, $H_0 = H_{D\Delta t} = 0$ as in [9,13].

The $p(z)$ term. As in [13], we assume prior independence between experimental conditions regarding the activation class assignments. It follows that $p(z) = \prod_{m=1}^M p(z_m; \beta_m)$ where we assumed in addition that $p(z_m; \beta_m)$ is a spatial Markov prior, namely a K-class Potts model with interaction parameter β_m [13]. The unknown parameters are then $\beta = \{\beta_m, m = 1 : M\}$. For the complete model, the whole set of parameters denoted by $\theta \in \Theta$ is $\theta = \{\boldsymbol{\Gamma}, \ell, \mu, \sigma, \sigma_h, \beta\}$.

3   Estimation by Variational EM

We propose to use an Expectation-Maximization (EM) framework to deal with the missing data namely, $A \in \mathcal{A}$, $H \in \mathcal{H}$, $Z \in \mathcal{Z}$. At iteration (r), denoting current parameter values by $\theta^{(r)}$, the E-step involves the posterior $p(a, h, z \,|\, y; \theta^{(r)})$, which is intractable for our model. Hence, we resort to a variational EM variant in which the intractable posterior is approximated as a product of three pdfs on $\mathcal{A}$, $\mathcal{H}$ and $\mathcal{Z}$ respectively. Previous attempts to use VB inference [2]

in fMRI [14, 11] have been successful with this type of approximations usually validated by assessing its fidelity to its MCMC counterpart. In Section 4, we will also provide such a comparison. It follows then that our E-step becomes an approximate E-step, which can be further decomposed into three stages consisting in updating the three pdfs, denoted by q_H, q_A and q_Z, in turn. Let $q_A^{(r-1)}, q_Z^{(r-1)}$ and $\theta^{(r)}$, be the current estimates at the rth iteration and $\text{E}_q[.]$ denotes the expectation with respect to (wrt) some pdf q, the first E-H step reads:

$$\texttt{E-H:} \quad q_H^{(r)}(h) \propto \exp\left(\text{E}_{q_A^{(r-1)} q_Z^{(r-1)}}\left[\log p(h\,|\,y, A, Z; \theta^{(r)})\right]\right) .$$

The following E-A and E-Z steps have similar expressions obtained by exchanging the role of H and A (resp. of H and Z) and replacing $q_A^{(r-1)}$ by $q_H^{(r)}$ (resp. $q_A^{(r-1)} q_Z^{(r-1)}$ by $q_A^{(r)} q_H^{(r)}$). For the E-H and E-A steps, it follows from standard algebra that $q_H^{(r)}$ and $q_A^{(r)}$ are both Gaussian pdfs: $q_H^{(r)} \sim \mathcal{N}(m_H^{(r)}, \boldsymbol{\Sigma}_H^{(r)})$ and $q_A^{(r)} = \prod_{i\in\mathcal{P}} \mathcal{N}(m_{A_i}^{(r)}, \boldsymbol{\Sigma}_{A_i}^{(r)})$.

• **E-H step.** The expressions for $m_H^{(r)}$ and $\boldsymbol{\Sigma}_H^{(r)}$ are similar to those derived in the MCMC case [9, Eq. (B.1)] with expressions involving the a_{mi}'s replaced by their expectations wrt $q_{A_i}^{(r-1)}$: $m_H^{(r)} = \boldsymbol{\Sigma}_H^{(r)}\left(\sum_{i\in\mathcal{P}} \boldsymbol{S}_i^{(r-1)\,t} \widetilde{y}_i^{(r)}\right)$ and

$$\boldsymbol{\Sigma}_H^{(r)-1} = \boldsymbol{R}^{-1}/\sigma_h^{2(r)} + \sum_{i\in\mathcal{P}}\left(\sum_{m,m'} \sigma_{A_{mi}A_{m'i}}^{(r-1)} \boldsymbol{X}_m^t \boldsymbol{\Gamma}_i^{(r)} \boldsymbol{X}_{m'} + \boldsymbol{S}_i^{(r-1)\,t} \boldsymbol{\Gamma}_i^{(r)} \boldsymbol{S}_i^{(r-1)}\right),$$

with $\boldsymbol{S}_i^{(r-1)} = \sum_{m=1}^{M} m_{A_{mi}}^{(r-1)} \boldsymbol{X}_m$ and $\widetilde{y}_i^{(r)} = \boldsymbol{\Gamma}_i^{(r)}(y_i - \boldsymbol{P}\ell_i^{(r)})$. Here, $m_{A_{mi}}^{(r-1)}$ and $\sigma_{A_{mi}A_{m'i}}^{(r-1)}$ denote the mth and $(m, m')^{\text{th}}$ entries of the mean vector and covariance matrix of the current $q_{A_i}^{(r-1)}$, respectively.

• **E-A step.** Here, the relationship with the MCMC update of a is not straightforward. In [9,13], the a_{mi}'s are sampled independently and conditionally on the z_{mi}'s. This is not the case in the VEM framework while some similarity appears if we set the probabilities $q_{Z_{mi}}(k)$'s either to 0 or 1 and consider only the diagonal part of $\boldsymbol{\Sigma}_{A_i}$. The update of q_A reads:

$$m_{A_i}^{(r)} = \boldsymbol{\Sigma}_{A_i}^{(r)}\left(\sum_{k=1}^{K} \boldsymbol{\Delta}_{ki}^{(r)} \mu_k^{(r)} + \widetilde{\boldsymbol{X}}_i^{(r)\,t} m_H^{(r)}\right) \text{ and } \boldsymbol{\Sigma}_{A_i}^{(r)} = \left(\sum_{k=1}^{K} \boldsymbol{\Delta}_{ki}^{(r)} + \widetilde{\boldsymbol{H}}_i^{(r)}\right)^{-1}$$

where $\mu_k^{(r)} = \left[\mu_{1k}^{(r)} \ldots \mu_{Mk}^{(r)}\right]^t$, $\widetilde{\boldsymbol{X}}_i^{(r)}$ is a $D+1 \times M$ matrix whose m^{th} column is $\boldsymbol{X}_m^t \widetilde{y}_i^{(r)}$, $\boldsymbol{\Delta}_{ki}^{(r)}$ is a $M \times M$ diagonal matrix whose $(m, m)^{\text{th}}$ entry is $q_{Z_{mi}}^{(r-1)}(k)/\sigma_{mk}^{2(r)}$ and $\widetilde{\boldsymbol{H}}_i^{(r)}$ is a $M \times M$ matrix whose $(m, m')^{\text{th}}$ entry is $\text{tr}\left((\boldsymbol{\Sigma}_H^{(r)} + m_H^{(r)} m_H^{(r)\,t}) \boldsymbol{X}_m \boldsymbol{\Gamma}_i^{(r)} \boldsymbol{X}_{m'}\right)$.

• **E-Z step.** From $p(a|z)$ and $p(z)$ in Section 2, the (A_m, Z_m) couples correspond to independent hidden Potts models with Gaussian class distributions. It comes an approximation that factorizes over conditions: $q_Z^{(r)}(z) = \prod_{m=1}^{M} q_{Z_m}^{(r)}(z_m)$ where $q_{Z_m}^{(r)}(z_m) = p_m(z_m\,|\,A_m = m_{A_m}^{(r)}; \mu_m^{(r)}, \sigma_m^{(r)}, \beta_m^{(r)})$ is the posterior of Z_m in a modified hidden Potts model, p_m, in which the observations a_{mi}'s are replaced by their mean values $m_{A_{mi}}^{(r)}$ and an external field $\{\sigma_{A_{im}A_{im}}^{(r)}\left[1/\sigma_{m1}^{2(r)} \ldots 1/\sigma_{mK}^{2(r)}\right]^t, i \in \mathcal{P}\}$ is added to the prior Potts model $p(z_m; \beta_m^{(r)})$. The Potts expression above is intractable but we use a mean field-like technique (see [4] for details) to approximate $q_{Z_m}^{(r)}(z_m)$ by a factorized pdf $\widetilde{q}_{Z_m}^{(r)}(z_m) = \prod_{i\in\mathcal{P}} \widetilde{q}_{Z_{mi}}^{(r)}(z_{mi})$.

M-step. It is also divided into four sub-steps involving separately (μ, σ), σ_h, β and $(\ell, \boldsymbol{\Gamma})$. The first two maximizers admit closed-form expressions:

- M-(μ, σ) step: Let $\bar{q}_{mk}^{(r)} = \sum\limits_{i \in \mathcal{P}} q_{Z_{mi}}^{(r)}(k)$, then $\mu_{mk}^{(r+1)} = \sum\limits_{i \in \mathcal{P}} q_{Z_{mi}}^{(r)}(k)\, m_{A_{mi}}^{(r)} / \bar{q}_{mk}^{(r)}$

and $\sigma_{mk}^{2(r+1)} = \sum\limits_{i \in \mathcal{P}} q_{Z_{mi}}^{(r)}(k) \left((m_{A_{mi}}^{(r)} - \mu_{mk}^{(r+1)})^2 + \sigma_{A_{im}A_{im}}^{(r)} \right) / \bar{q}_{mk}^{(r)}$.

- M-σ_h^2 step: $\sigma_h^{2(r+1)} = (D-1)^{-1} \mathrm{tr}\left((\boldsymbol{\Sigma}_H^{(r)} + m_H^{(r)} m_H^{(r)^t}) \boldsymbol{R}^{-1} \right)$.

The two other M-steps require iterative maximization procedures. Updating β consists in making further use of a mean field-like approximation [4]. Regarding the pair $(\ell, \boldsymbol{\Gamma})$, it satisfies some fixed point equation, which simplifies in case of white noise. For autoregressive (AR) noise models, we found some similarity with [9, Eq. (B.2)] when replacing h and a by m_H and m_A, respectively.

4   Experiments

Simulation Results. We simulated data according to Eq. (1)[1] and $p(a \mid z)$ with a white Gaussian noise $\boldsymbol{\Gamma}_i^{-1} = 0.5\ \boldsymbol{I}_N$ ($\boldsymbol{I}_N$ is the $N \times N$ identity matrix), $M = 2$ experimental conditions and stimulus-varying contrast-to-noise ratios (CNR): $\mu_{12} = 2.8, \sigma_{12} = 0.5$ and $\mu_{22} = 1.8, \sigma_{22} = 0.6$ so that $\mu_{12}/\sigma_{12} > \mu_{22}/\sigma_{22}$. The *initial* artificial paradigm comprised 15 stimulus events for each condition. The simulation process finally yielded time-series lasting 152 scans. Condition-specific activating and non-activating voxels were defined as 20×20 2D slices shown in Fig. 1 (right) and superimposed to the estimated label probabilities in white solid line. The parameters β_1 and β_2 had been set to fixed values ($\beta_1 = \beta_2 = 0.8$ for the two algorithms). $\boldsymbol{\Gamma}$ and ℓ are estimated as in [9].
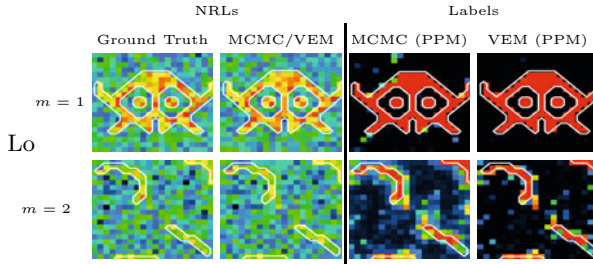

Fig. 1. Left: Ground truth and estimated Neural Response Levels (NRLs) by MCMC and VEM (same results); **Right:** Posterior probability maps (PPM) given by the approximation q_{Z_m} (VEM) and by the MMSE estimator (MCMC)

In Fig. 1, the VEM is compared to the MCMC alternative developed in [13]: both algorithms report similar NRL maps while some difference is exhibited on the posterior activation probability map (PPM) for the low CNR condition ($m = 2$, bottom row). This illustrates the gain in robustness achieved using the variational approximation under the true noise model.

---

[1] $\boldsymbol{P}$ was defined from a cosine transform basis.

To perform a quantitative comparison, several experiments with different stimuli densities (from 5 to 30), noise variance and autocorrelation ($\boldsymbol{\Gamma}_i^{-1}$) have been conducted. Fig. 2(a) illustrates the evolution of the Mean Square Error (MSE) of NRL estimates wrt the stimulus density in the experimental paradigm when a second order autoregressive noise (AR(2)) is considered. This figure shows that at low stimulus density (i.e. low Signal to Noise Ratio (SNR)[2]), the proposed VEM algorithm is more robust than the MCMC one to model discrepancy. Indeed, here the two inference algorithms were compared for a white and Gaussian noise modelling in Eq. (1). In contrast, at high stimulus density (≥ 20), the two methods perform similarly. Interestingly, Fig. 2(b)-(c) depict the shapes of the ground truth and estimated HRF shapes inferred by the VEM and MCMC schemes wrt the stimulus density: Note that the main HRF features (peak value (PV), time-to-peak (TTP) and time-to-undershoot (TTU)) remain well estimated by both methods. However, at low stimulus density, Fig. 2(b) shows that the VEM algorithm is less accurate than its MCMC counterpart close to the undershoot position. Similar experiments have been conducted while changing the ground truth HRF properties (PV, TTP, TTU), and coherent results have been obtained. Other comparisons performed under the true noise model did not reveal any significant difference between VEM and MCMC.
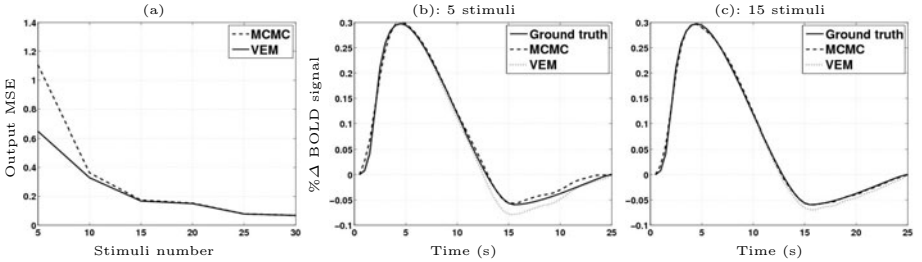

Fig. 2. (a): MSE evolution of estimated NRLs wrt stimuli number. **(b)-(c):** Ground truth and HRF estimates inferred by the VEM and MCMC schemes for two stimulus densities corrupted by AR(2) noise.

In Fig. 3(a)-(b) the output MSE is plotted against the input SNR when varying the noise variance and its amount of autocorrelation, respectively. In the latter case, the two AR parameters are varied while maintaining a stable AR(2) process. As already observed in [3], at fixed input SNR, the impact of large autocorrelation is stronger than that of large noise variance irrespective of the inference scheme. Moreover, the two inference methods perform very similarly on a large scale of input SNR (SNR > 5 dB). In terms of computational time, VEM results have been obtained 30 times faster than with MCMC on an Intel Core 2 - 2.26 GHz - 2 Gb RAM architecture.

---

[2] The SNR is given by: SNR $= 10 \log \sum_{i \in \mathcal{P}} \| \sum_{m=1}^{M} A_{mi} \boldsymbol{X}_m h \|^2 / \sum_{i \in \mathcal{P}} \| \varepsilon_i \|^2$.
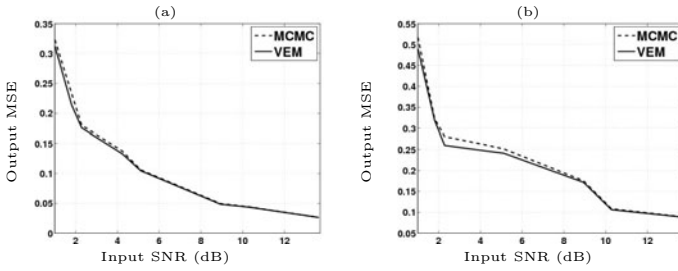
Fig. 3. MSE evolution of NRL estimates wrt input SNR (AR(2) noise) by varying the noise variance **(a)** and the amount of AR(2) noise autocorrelation **(b)**
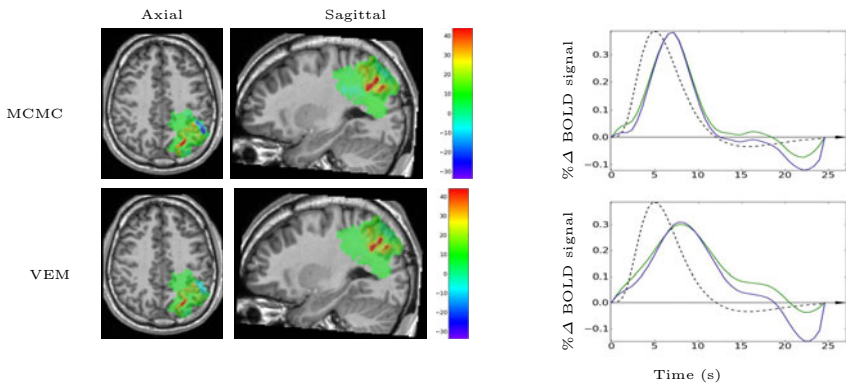

Fig. 4. Left: Estimated contrast **Computation-Sentences** by MCMC and VEM; **Right:** HRF estimates by MCMC (green) and VEM (blue) at maximum intensity peak (top) and in a neighboring parcel (bottom). Canonical HRF with dashed line.

Real Data Processing. fMRI data were recorded at 3 T (Siemens Trio) using a gradient-echo EPI sequence (TE=30 ms/TR=2.4 s/FOV=192 mm^2) during a *Localizer* experiment [12]. The acquisition consisted of a single session of $N = 128$ scans, yielding 3-D volumes with a spatial resolution of $2 \times 2 \times 3$ mm^3. The paradigm was a fast event-related design comprising sixty auditory, visual and motor stimuli, defined in ten experimental conditions (auditory and visual sentences, auditory and visual calculations, left/right auditory and visual clicks, horizontal and vertical checkerboards).

We focused on the `Computation-Sentences` contrast differentiating the activations induced by the calculation and sentence conditions in the left intraparietal sulcus, a region known to elicit hemodynamic response that departs from the canonical HRF. As shown in Fig. 4, the contrasted NRL estimates for the MCMC and VEM inference schemes are very similar and follow the underlying sulco-gyral anatomy. Note that only the most activating slice is considered for visualization purpose. The corresponding HRF estimates in the most acti-

vating parcel of about 200 voxels are also depicted in Fig. 4: they appear very similar and both quite different from the canonical shape regarding the TTP and TTU parameters. More oscillations arise in the VEM inference close to the undershoot, however we may have less confidence in the HRF tail than in its peak since it involves less signal strength. Moreover, the event-related nature of the paradigm is not suited to properly study the undershoot properties. Finally, in terms of computational efficiency, the variational approximation runs also 30 times faster than the MCMC inference in this parcel.

5   Conclusion

We proposed a Variational EM algorithm as an alternative solution to intensive stochastic sampling for inferring upon the JDE parameters. Illustrations on simulated data showed that our approach achieved similar and even better results than the MCMC-based inference scheme at low input SNR or stimuli density and was more robust to noise model mismatch. Also, in contrast to the hybrid MCMC in [13], the VEM algorithm only requires a simple stopping criterion as convergence diagnosis tool. Another advantage of the variational approach lies in its flexibility to adapt to more complex situations such as accounting for higher AR noise order, habituation modelling or including model selection steps using the log-evidence as information criterion. Other future work will focus on analyzing the impact of the VEM-JDE to group-level analysis as done in [1].

References

1. Badillo, S., Vincent, T., Ciuciu, P.: Impact of the joint detection-estimation approach on random effects group studies in fMRI. In: 7th International Symposium on Biomedical Imaging, Chicago, IL, pp. 376–380 (April 2011)
2. Beal, M., Ghahramani, Z.: The variational Bayesian EM algorithm for incomplete data: with application to scoring graphical model structures. Bayesian Statistics, vol. 7, pp. 453–464. University of Oxford Press, Oxford (2003)
3. Casanova, R., Ryali, S., Serences, J., Yang, L., Kraft, R., Laurienti, P., Maldjian, J.: The impact of temporal regularization on estimates of the BOLD hemodynamic response function: a comparative analysis. Neuroimage 40(4), 1606–1618 (2008)
4. Celeux, G., Forbes, F., Peyrard, N.: EM procedures using mean field-like approximations for Markov model-based image segmentation. Pat. Rec. 36, 131–144 (2003)
5. Friston, K., Jezzard, P., Turner, R.: Analysis of functional MRI time-series. Hum. Brain Mapp. 1, 153–171 (1994)
6. Handwerker, D.A., Ollinger, J.M., D'Esposito, M.: Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. Neuroimage 21(4), 1639–1651 (2004)
7. Kershaw, J., Ardekani, B.A., Kanno, I.: Application of Bayesian inference to fMRI data analysis. IEEE Trans. Med. Imag. 18(12), 1138–1153 (1999)
8. Makni, S., Beckmann, C., Smith, S., Woolrich, M.: Bayesian deconvolution of fMRI data using bilinear dynamical systems. Neuroimage 42(4), 1381–1396 (2008)

9. Makni, S., Idier, J., Vincent, T., Thirion, B., Dehaene-Lambertz, G., Ciuciu, P.: A fully Bayesian approach to the parcel-based detection-estimation of brain activity in fMRI. Neuroimage 41(3), 941–969 (2008)
10. Ogawa, S., Lee, T.M., Kay, A.R., Tank, D.W.: Brain magnetic resonance imaging with contrast dependent on blood oxygenation. Nat. Acad. Sci. 87, 9868–9872 (1990)
11. Penny, W.D., Kiebel, S., Friston, K.J.: Variational Bayesian inference for fMRI time series. Neuroimage 19(3), 727–741 (2003)
12. Pinel, P., Thirion, B., Mériaux, S., Jobert, A., Serres, J., Le Bihan, D., Poline, J.B., Dehaene, S.: Fast reproducible identification and large-scale databasing of individual functional cognitive networks. BMC Neurosci. 8(1), 91 (2007)
13. Vincent, T., Risser, L., Ciuciu, P.: Spatially adaptive mixture modeling for analysis of within-subject fMRI time series. IEEE Trans. Med. Imag. 29, 1059–1074 (2010)
14. Woolrich, M., Behrens, T.: Variational Bayes inference of spatial mixture models for segmentation. IEEE Trans. Med. Imag. 25(10), 1380–1391 (2006)