Estimation of the hemodynamic response in event-related functional MRI: Bayesian networks as a framework for efficient Bayesian modeling and inference.

Guillaume Marrelec, Philippe Ciuciu, Mélanie Pélégrini-Issac, Habib Benali

Estimation of the Hemodynamic Response in Event-Related Functional MRI: Bayesian Networks as a Framework for Efficient Bayesian Modeling and Inference

Guillaume Marrelec*, Philippe Ciuciu, *Member, IEEE*, Mélanie Pélégrini-Issac, and Habib Benali, *Senior Member, IEEE*

***Abstract*—A convenient way to analyze blood-oxygen-level-dependent functional magnetic resonance imaging data consists of modeling the whole brain as a stationary, linear system characterized by its transfer function: the hemodynamic response function (HRF). HRF estimation, though of the greatest interest, is still under investigation, for the problem is ill-conditioned. In this paper, we recall the most general Bayesian model for HRF estimation and show how it can beneficially be translated in terms of Bayesian graphical models, leading to 1) a clear and efficient representation of all structural and functional relationships entailed by the model, and 2) a straightforward numerical scheme to approximate the joint posterior distribution, allowing for estimation of the HRF, as well as all other model parameters. We finally apply this novel technique on both simulations and real data.**

***Index Terms*—Bayesian inference, Bayesian networks, functional MRI, hemodynamic response function.**

I. Introduction

FUNCTIONAL magnetic resonance imaging (fMRI) is a noninvasive technique allowing for the evolution of brain processes to be dynamically followed in various cognitive and behavioral tasks [1]. In the most common fMRI technique, based on the so-called blood-oxygen-level-dependent (BOLD) contrast, the measure is only indirectly related to neuronal activity through a process that is still under investigation [2]–[4]. For this reason, a convenient way to analyze BOLD fMRI data consists of modeling the whole brain as a stationary, linear "black box" system characterized by its transfer response function, also called hemodynamic response function (HRF) [5]. This model, called general linear model (GLM), fairly well accounts for the properties of the real system as long as the inter-stimulus interval does not decrease beyond about two seconds [6], [7]. When this constraint is not respected, other models have to be developed [8], [9].

Estimation of the HRF is of the greatest interest when analyzing fMRI data, since it can give a deep insight into the underlying dynamics of brain activation and the relationships between activated areas. HRFs are increasingly suspected to vary from region to region, from task to task, and from subject to subject [10]–[12]. Age and disease are also more and more believed to have a significant influence on the BOLD response [13], [14]. Nevertheless, accurate estimation of the response function still belongs to ongoing research, since the problem is badly conditioned. Various nonparametric methods have been developed so far in an attempt to infer the HRF at each time sample, such as selective averaging [6], averaging over regions [15], introduction of nondiagonal models for the temporal covariance of the noise [16], or temporal regularization [17].

In [18] and [19], we proposed a Bayesian nonparametric estimation of the HRF for event-related designs. Basic yet relevant physiological information was introduced to temporally constrain the problem and calculate robust estimators of the parameters of interest. In [20]–[22], the model was extended to account for asynchronous event-related designs, different trial types, and several fMRI sessions, further improving the estimation. For calculation reasons, all variants proposed so far have, however, the drawback of not integrating the hyperparameter uncertainty. Furthermore, probabilistic treatment of the drift parameters in the extended model was possible [23], [24], but at a significantly higher computational cost.

In this paper, we propose to cast a new light on the GLM. We still place ourselves in a Bayesian framework, permitting integration of information originating from various sources and efficient inference on the parameters of interest. A general model is set to account for most event-related fMRI data. In a conventional Bayesian approach, we would then calculate the joint posterior distribution of all parameters, which would be the pivotal quantity for all further inference. Since direct sampling from this probability density function (pdf) would prove impossible, Monte Carlo Markov chain (MCMC) sampling would be required, such as Gibbs sampling [25], [26]. In this case, posterior conditional pdfs should be derived. In this perspective,

Manuscript received January 5, 2004; revised May 3, 2004. The work of G. Marrelec was supported by the Fondation pour la Recherche Médicale. This paper is an extension of [33] that appeared in the IPMI'03 Proceedings. The Guest Editors responsible for coordinating the review of this paper and recommending its publication were C. J. Taylor and J. A. Noble. *Asterisk indicates corresponding author.Asterisk indicates corresponding author.*

*G. Marrelec is with INSERM U494, CHU Pitié-Salpêtrière, 91 boulevard de l'Hôpital, 75634 Paris Cedex 13, France and also with the IFR 49, Orsay, France.

H. Benali are with INSERM U494, CHU Pitié-Salpêtrière, 91 boulevard de l'Hôpital, 75634 Paris Cedex 13, France, and also with the IFR 49, Orsay, France.

P. Ciuciu is with SHFJ/CEA/INSERM U562, 91401 Orsay, France, and also with the IFR 49, Orsay, France.

M. Pélégrini-Issac is with INSERM U483, 9 quai Saint Bernard, 75005 Paris, and also with the IFR 49, Orsay, France.

Digital Object Identifier 10.1109/TMI.2004.831221

we advocate that calculation of the posterior pdf is unncessary. Instead, we resort to a novel approach that focuses on graphical modeling and that, once the model has been properly set, makes it possible to directly lead probabilistic inference about all parameters. More precisely, we utilize graph theory [27] to conveniently deal with the model. Indeed, graphs give a very simple and efficient representation of the model, however complex it may be. In this framework, we translate the model into a Bayesian network. Using Markov properties of such networks, drawing inference becomes straightforward, and Gibbs sampling finally provides us with a numerical approximation of the joint posterior pdf.

In the next part of this paper, we develop the general Bayesian framework for HRF estimation, presenting an extended version of the GLM. In Section III, the GLM is translated in terms of graphical model, and it is shown how inference can readily be performed from there. We briefly present simulations and finally apply our model to real data.

II. HRF Estimation in fMRI Data Analysis

A. Notations

In the following, x denotes a real number, $\boldsymbol{x}$ a vector, and $\boldsymbol{X}$ a matrix. For the sake of simplicity, (x_i)—between parentheses—is a shortcut for $(x_i)_{1\leq i\leq I}$. "^{t}" is the regular matrix transposition. $\boldsymbol{I}_N$ stands for the $N\times N$ identity matrix. "$\propto$" relates two expressions that are proportional. For two variables x and y, "$x\,|\,y$" stands for "x given y," and $\mathrm{p}(x)$ for the probability of x. $\mathcal{N}(\boldsymbol{m},\boldsymbol{V};\boldsymbol{x})$ is the Gaussian density function with mean $\boldsymbol{m}$ and covariance matrix $\boldsymbol{V}$ calculated at sample $\boldsymbol{x}$. Inv-$\chi^2(d,r^2;u)$ is the scaled inverse-chi-square density function[1] with d degrees of freedom and scale parameter r^2 evaluated at sample u.

B. General Linear Model

Let an fMRI experiment be composed of S sessions, each session involving I different stimulus types. Define $\boldsymbol{y}_s=(y_{s,t_{s,n}})_{1\leq n\leq N_s}$ as the BOLD fMRI time course of a voxel (i.e., volume element) at (not-necessarily uniformly sampled) times $(t_{s,n})$ for session s, and $\boldsymbol{x}_{s,i}=(x_{s,i,t})_{t_{s,0}\leq t\leq t_{s,N_s}}$ the corresponding binary time series, composed of the ith stimulus onsets. The following discrete linear convolution model (H) is assumed to hold between the stimuli and the data

$$y_{s,t_{s,n}}=\sum_{i=1}^{I}\sum_{k=0}^{K_i}x_{s,i,t_{s,n}-k\Delta t}h_{i,k\Delta t}+\sum_{m=1}^{M_s}d_{m,t_{s,n}}\lambda_{s,m}+e_{t_n}$$

$n=n_s+1,\ldots,N_s$, where n_s is the largest integer so, that $t_{s,n}-K_i\Delta t<t_{s,1}$ for all i. The (K_i+1)-dimensional vector $\boldsymbol{h}_i=(h_{i,k\Delta t})^t$ represents the ith unknown HRF to be estimated, sampled every Δt. All HRFs are assumed to be constant across sessions. $L_s=N_s-n_s$ is the actual amount of data used in the calculation for each session. $\boldsymbol{X}_{s,i}=(x_{s,i,t_{s,n}-k\Delta t})$ is the regular L_s-by-(K_i+1) design matrix, consisting of the lagged stimulus covariates. In the L_s-by-M_s matrix $\boldsymbol{D}_s=(d_{m,t_{s,n}})$ are the values at times $(t_{s,n})$ of a basis of M_s functions that takes a potential drift and any other nuisance effect into account, and $\boldsymbol{\lambda}_s=(\lambda_{s,m})^t$ contains the corresponding coefficients. For the sake of simplicity, the basis is assumed to be orthonormal, so that $(1/L_s)\boldsymbol{D}_s\boldsymbol{D}_s^t=\boldsymbol{I}_{L_s}$. Vector $\boldsymbol{e}_s=(e_{s,t_{s,n}})^t$ accounts for noise and is supposed to consist of independent and identically distributed Gaussian variables of unknown variance σ_s^2, assumed to be independent from the HRFs. In matrix form, (H) boils down to

$$\boldsymbol{y}_s=\sum_{i=1}^{I}\boldsymbol{X}_{s,i}\boldsymbol{h}_i+\boldsymbol{D}_s\boldsymbol{\lambda}_s+\boldsymbol{e}_s,\quad s=1,\ldots,S$$

also called general linear model (GLM). In this model, the likelihood of the data yields

$$\mathrm{p}\left((\boldsymbol{y}_s)\,|\,H,(\boldsymbol{h}_i),\left(\sigma_s^2\right),(\boldsymbol{\lambda}_s)\right)=\prod_{s=1}^{S}\mathrm{p}\left(\boldsymbol{y}_s\,|\,H,(\boldsymbol{h}_i),\sigma_s^2,\boldsymbol{\lambda}_s\right)\quad(1)$$

with each term in the product reading

$$\mathrm{p}\left(\boldsymbol{y}_s\,|\,H,(\boldsymbol{h}_i),\sigma_s^2,\boldsymbol{\lambda}_s\right)=\mathcal{N}\left(\sum_{i=1}^{I}\boldsymbol{X}_{s,i}\boldsymbol{h}_i+\boldsymbol{D}_s\boldsymbol{\lambda}_s,\sigma_s^2\boldsymbol{I}_{L_s};\boldsymbol{y}_s\right).\quad(2)$$

C. HRFs and Hyperparameters

In its general form, the GLM is usually ill-conditioned, for there are too many parameters to estimate compared to the information brought by the data. Prior information must, hence, be incorporated in order to constrain the problem. Since the underlying physiological process of BOLD fMRI is as of yet only partially understood, we set the following soft constaints [18], [21].

P1) The HRFs start and end at 0. This amounts to setting the first and last samples of each HRF to 0, so that only K_i-1 parameters (instead of K_i+1) are now unknown.

P2) The HRFs are smooth. Quantification is achieved by setting Gaussian priors for the norm of the second derivative of the HRFs, whose variances are adjusted by hyperparameters ϵ_i's. More precisely, we assume that

$$\mathrm{p}\left(\boldsymbol{h}_i\,\middle|\,H,\epsilon_i^2\right)\propto\exp\left[\frac{1}{2\epsilon_i^2}\|\partial^2\boldsymbol{h}_i\|^2\right]\qquad i=1,\ldots,I.$$

Following usual practice, $\partial^2\boldsymbol{h}_i$ can be discretized as

$$(\partial^2\boldsymbol{h})_{i,k\Delta t}\approx\frac{h_{i,(k+1)\Delta t}-2h_{i,k\Delta t}+h_{i,(k-1\Delta t)}}{(\Delta t)^2}$$

for $k=1,\ldots,K_i-1$. Taking into account that $h_{i,0}=h_{i,K}=0$, we obtain in matrix form

$$\partial^2\boldsymbol{h}_i=\boldsymbol{D}_{2,i}\boldsymbol{h}_i$$

where $\boldsymbol{D}_{2,i}$ is the following $(K_i-1)\times(K_i-1)$ matrix:

$$\boldsymbol{D}_{2,i}=\frac{1}{(\Delta t)^2}\begin{pmatrix}-2&1&0&&&&\\1&-2&1&0&&&\\0&1&-2&1&0&&\\&\ddots&\ddots&\ddots&\ddots&\ddots&\\&&0&1&-2&1&0\\&&&0&1&-2&1\\&&&&0&1&-2\end{pmatrix}.$$

[1] If u is chi-square distributed with d degrees of freedom, then dr^2/u is scaled inverse-chi-square distributed with d degrees of freedom and scale parameter r^2. Equivalently, a scaled inverse-chi-square distribution Inv-$\chi^2(d,r^2)$ is a special case of the inverse Gamma distribution Inv-$\Gamma(\alpha,\beta)$, with $\alpha=d/2$ and $\beta=dr^2/2$. See the Appendix for the exact expression of the corresponding pdf.

Hence

$$\begin{aligned}\|\partial^2 \boldsymbol{h}_i\|^2 &= (\boldsymbol{D}_{2,i}\boldsymbol{h}_i)^t(\boldsymbol{D}_{2,i}\boldsymbol{h}_i)\\ &= \boldsymbol{h}_i^t\left(\boldsymbol{D}_{2,i}^t\boldsymbol{D}_{2,i}\right)\boldsymbol{h}_i.\end{aligned}$$

Finally, we obtain for $\boldsymbol{h}_i$:

$$\mathrm{p}\left(\boldsymbol{h}_i \,\middle|\, H, \epsilon_i^2\right) = \mathcal{N}\left(\boldsymbol{0}, \epsilon_i^2\boldsymbol{R}_i^{-1}; \boldsymbol{h}_i\right) \quad i = 1, \ldots, I \tag{3}$$

where $\boldsymbol{R}_i = \boldsymbol{D}_{2,i}^t\boldsymbol{D}_{2,i}$ is the following $(K_i - 1)$-by-$(K_i - 1)$ symmetrical positive definite matrix

$$\boldsymbol{R}_i = \frac{1}{(\Delta t)^4} \times \begin{pmatrix} 5 & -4 & 1 & 0 & & \cdots & & & 0 \\ -4 & 6 & -4 & 1 & 0 & & & & \\ 1 & -4 & 6 & -4 & 1 & 0 & & & \\ 0 & 1 & -4 & 6 & -4 & 1 & 0 & & \vdots \\ & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \\ \vdots & & 0 & 1 & -4 & 6 & -4 & 1 & 0 \\ & & & 0 & 1 & -4 & 6 & -4 & 1 \\ & & & & 0 & 1 & -4 & 6 & -4 \\ 0 & & & \cdots & & 0 & 1 & -4 & 5 \end{pmatrix}.$$

P3) No prior dependence is assumed between HRFs, so that

$$\mathrm{p}\left((\boldsymbol{h}_i), \left(\epsilon_i^2\right) \,\middle|\, H\right) = \prod_{i=1}^{I} \mathrm{p}\left(\boldsymbol{h}_i \,\middle|\, H, \epsilon_i^2\right) \cdot \mathrm{p}\left(\epsilon_i^2 \,\middle|\, H\right).$$

For convenience reasons, the priors for ϵ_i^2's are set as conjugate priors, i.e., these parameters are assumed to be *a priori* independent identically distributed with common pdf a scaled inverse-χ^2 with n_ϵ degrees of freedom and scale parameter r_ϵ^2 given in the model (n_ϵ set to a small value to obtain a "hardly" informative prior). This setting is further precised later and also analyzed in the discussion.

D. Drifts and Noise Variances

Unlike the HRFs, noise variances and drift parameters may vary across sessions. Each σ_s^2 is assumed to follow a scaled inverse-χ^2 distribution with $n_{\sigma,s}$ degrees of freedom and scale parameter $r_{\sigma,s}^2$. Each $\boldsymbol{\lambda}_s$ is assumed to be Gaussian distributed with mean $\boldsymbol{m}_{\lambda,s}$ and covariance matrix $\boldsymbol{V}_{\lambda,s}$.

E. Joint Posterior Distribution

Considering the model so constructed and assuming no further prior dependence between parameters, formal application of the chain rule yields

$$\begin{aligned}&\mathrm{p}\left((\boldsymbol{y}_s), \left(\epsilon_i^2\right), (\boldsymbol{h}_i), \left(\sigma_s^2\right), (\boldsymbol{\lambda}_s) \,|\, H\right)\\ &\quad= \prod_{s=1}^{S} \mathrm{p}\left(\boldsymbol{y}_s \,\middle|\, H, (\boldsymbol{h}_i), \sigma_s^2, \boldsymbol{\lambda}_s\right)\\ &\qquad\times \mathrm{p}(\boldsymbol{\lambda}_s \,|\, H) \cdot \mathrm{p}\left(\sigma_s^2 \,\middle|\, H\right)\\ &\qquad\cdot \prod_{i=1}^{I} \mathrm{p}\left(\boldsymbol{h}_i \,\middle|\, H, \epsilon_i^2\right) \cdot \mathrm{p}\left(\epsilon_i^2 \,\middle|\, H\right).\end{aligned} \tag{4}$$

Given data $(\boldsymbol{y}_s)$, our knowledge relative to the model parameters can easily be updated using the conditioning formula

$$\mathrm{p}\left(\left(\epsilon_i^2\right), (\boldsymbol{h}_i), \left(\sigma_s^2\right), (\boldsymbol{\lambda}_s) \,|\, H, (\boldsymbol{y}_s)\right) = \frac{\mathrm{p}\left((\boldsymbol{y}_s), \left(\epsilon_i^2\right), (\boldsymbol{h}_i), \left(\sigma_s^2\right), (\boldsymbol{\lambda}_s) \,|\, H\right)}{\mathrm{p}((\boldsymbol{y}_s) \,|\, H)}.$$

In words, the joint posterior probability distribution is proportional to the joint probability of (4). Replacing all distributions by their functional forms, this joint posterior pdf could be calculated in closed form, as is indeed done in most works applying Bayesian analysis. Since direct sampling from the joint posterior pdf is impossible, we must resort to MCMC, e.g., Gibbs sampling where the conditional pdfs should be derived. In this perspective, we propose to avoid calculating the joint posterior pdf to directly proceed to inference. In order to do so, we beforehand embed our model in a framework that allows for convenient representation, handling, and numerical inference: Bayesian graphical models.

III. Graphical Modeling

A. Directed Acyclic Graphs and Bayesian Networks

A graph G is a mathematical object that relates a set of vertices, or nodes, V, to a set of edges, E, consisting of pairs of elements taken from V. There is a directed edge or arrow between vertices $\boldsymbol{z}_n$ and $\boldsymbol{z}_m$ in V if the set E contains the ordered pair $(\boldsymbol{z}_n, \boldsymbol{z}_m)$; vertex $\boldsymbol{z}_n$ is a parent of vertex $\boldsymbol{z}_m$, and vertex $\boldsymbol{z}_m$ is a child of vertex $\boldsymbol{z}_n$. A directed graph is a graph whose edges are all directed. A path is a sequence of distinct vertices $\boldsymbol{z}_{n_1}, \ldots, \boldsymbol{z}_{n_m}$ for which $(\boldsymbol{z}_{n_l}, \boldsymbol{z}_{n_{l+1}})$ is in E for each $l = 1, \ldots, m-1$. The path is a cycle if the end points are allowed to be the same, $\boldsymbol{z}_{n_1} = \boldsymbol{z}_{n_m}$. An oriented graph with no cycle is called a directed acyclic graph (DAG).

A distribution p over $\boldsymbol{z}$ is compatible with a DAG G if it satisfies all independence relationships entailed by G. (G, p) is then called a Bayesian network. For more details, the reader is referred to [27]. The major feature of Bayesian networks is that $\mathrm{p}(\boldsymbol{z})$ must factorize according to the so-called factorization property

$$\mathrm{p}(\boldsymbol{z}) = \prod_{n=1}^{N} \mathrm{p}(\boldsymbol{z}_n \,|\, \mathrm{pa}(\boldsymbol{z}_n)) \tag{5}$$

where $\mathrm{pa}(n)$ is the set of parents of vertex $\boldsymbol{z}_n$. This is nothing but a multidimensional generalization of the Markov chain rule. Defining a Bayesian network, hence, amounts to 1) defining relevant variables (i.e., nodes) $\boldsymbol{z}_n$, 2) defining structural relationships (i.e., edges) $\boldsymbol{z}_n \rightarrow \boldsymbol{z}_m$, and 3) defining functional relationships $\mathrm{p}(\boldsymbol{z}_n \,|\, \mathrm{pa}(n))$. Pearl [28] showed a property that proves to be very efficient for numerical sampling, namely that nothing more is required to calculate the conditional probability of any node: the probability distribution of any variable $\boldsymbol{z}_n$ in the network, conditioned on the state of all other variables, is given by the product

$$\mathrm{p}(\boldsymbol{z}_n \,|\, \mathrm{r.v.}) \propto \mathrm{p}(\boldsymbol{z}_n \,|\, \mathrm{pa}(\boldsymbol{z}_n)) \cdot \prod_{\boldsymbol{z}_m \in \mathrm{ch}(\boldsymbol{z}_n)} \mathrm{p}(\boldsymbol{z}_m \,|\, \mathrm{pa}(\boldsymbol{z}_m)) \tag{6}$$

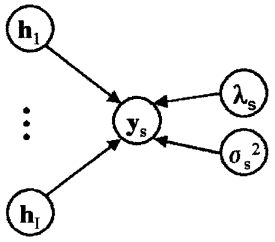

Fig. 1. Structural dependence of $\boldsymbol{y}_s$.

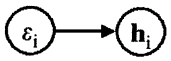

Fig. 2. Structural dependence of $\boldsymbol{h}_i$.

where r.v. stands for "remaining variables" and $\mathrm{ch}(\boldsymbol{z}_n)$ for the children nodes of $\boldsymbol{z}_n$. Note that $\boldsymbol{z}_m$ can have several parents and, hence, $\mathrm{pa}(\boldsymbol{z}_m)$ may not be restricted to $\boldsymbol{z}_n$. This formula states that the conditional probabilities can be derived from local quantities that are part of the model specification.

B. Constructing the GLM Graphical Model

The GLM can easily be expressed in terms of a Bayesian network. This translation requires two major steps: representing the structural relationships, and then the functional relationships.

Equation (1) states that each $\boldsymbol{y}_s$ depends on the corresponding $\boldsymbol{\lambda}_s$ and σ_s^2, as well as all $\boldsymbol{h}_i$'s. This equation can, hence, be represented by the graph depicted in Fig. 1.

As to the model relative to the prior information on the HRFs, it expresses that each $\boldsymbol{h}_i$ depends exclusively on ϵ_i^2. Fig. 2 is, hence, a good model for it.

Gathering all parts leads to the graph proposed in Fig. 3. Irrespective of the functional relationships between nodes, application of (5) to the graphical model states that the joint pdf for all variables decomposes as

$$\begin{aligned}
&\mathrm{p}\left((\boldsymbol{y}_s),\left(\epsilon_i^2\right),(\boldsymbol{h}_i),\left(\sigma_s^2\right),(\boldsymbol{\lambda}_s)\,|\,H\right)\\
&\quad=\prod_{s=1}^{S}\mathrm{p}(\boldsymbol{y}_s\,|\,H,\mathrm{pa}(\boldsymbol{y}_s))\\
&\qquad\times\mathrm{p}(\boldsymbol{\lambda}_s\,|\,H,\mathrm{pa}(\boldsymbol{\lambda}_s))\cdot\mathrm{p}\left(\sigma_s^2\,\middle|\,H,\mathrm{pa}\left(\sigma_s^2\right)\right)\\
&\qquad\times\prod_{i=1}^{I}\mathrm{p}\left(\boldsymbol{h}_i\,|\,H,\mathrm{pa}(\boldsymbol{h}_i)\right)\cdot\mathrm{p}\left(\epsilon_i^2\,\middle|\,H,\mathrm{pa}\left(\epsilon_i^2\right)\right)
\end{aligned}$$

which, once developed, is exactly (4). Our graphical model, thus, unambiguously embeds all structural relationships of the GLM. However, complicated (H) may be, it is still much simpler to conceptualize it in graph form than as it was presented before. Whereas determination of structural relationships between two given variables in model (H) remains a tough problem to tackle, the corresponding DAG clearly and unambiguously represents all possible independence relationships, that can be read off the graph using its Markov properties.

Among others, a direct consequence of this modeling is that it is now possible to apply (6) to the GLM graph model, expressing the conditional pdfs of all variables, $\mathrm{p}(\epsilon_i^2\,|\,H,\mathrm{r.v.}),\mathrm{p}(\boldsymbol{h}_i\,|\,H,\mathrm{r.v.}),\mathrm{p}(\sigma_s^2\,|\,H,\mathrm{r.v.})$, and $\mathrm{p}(\boldsymbol{\lambda}_s\,|\,H,\mathrm{r.v.})$ as a function of the distributions already defined

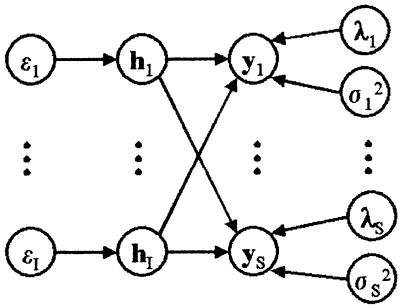

Fig. 3. DAG corresponding to the GLM.

$$\mathrm{p}\left(\epsilon_i^2\,\middle|\,H,\mathrm{r.v.}\right)\propto\mathrm{p}\left(\epsilon_i^2\,\middle|\,H\right)\cdot\mathrm{p}\left(\boldsymbol{h}_i\,\middle|\,H,\epsilon_i^2\right)\tag{7a}$$

$$\mathrm{p}(\boldsymbol{h}_i\,|\,H,\mathrm{r.v.})\propto\mathrm{p}\left(\boldsymbol{h}_i\,\middle|\,H,\epsilon_i^2\right)\times\prod_{s=1}^{S}\mathrm{p}\left(\boldsymbol{y}_s\,|\,H,(\boldsymbol{h}_i),\sigma_s^2,\boldsymbol{\lambda}_s\right)\tag{7b}$$

$$\mathrm{p}\left(\sigma_s^2\,\middle|\,H,\mathrm{r.v.}\right)\propto\mathrm{p}\left(\sigma_s^2\,\middle|\,H\right)\cdot\mathrm{p}\left(\boldsymbol{y}_s\,|\,H,(\boldsymbol{h}_i),\sigma_s^2,\boldsymbol{\lambda}_s\right)\tag{7c}$$

$$\mathrm{p}(\boldsymbol{\lambda}_s\,|\,H,\mathrm{r.v.})\propto\mathrm{p}(\boldsymbol{\lambda}_s\,|\,H)\cdot\mathrm{p}\left(\boldsymbol{y}_s\,|\,H,(\boldsymbol{h}_i),\sigma_s^2,\boldsymbol{\lambda}_s\right).\tag{7d}$$

Note that the functional relationships have not been defined yet—all the properties abovementioned are entailed by the sole structural relationships. As a matter of fact, the graphical representation is much more general than the GLM. The only constraints set by the graph is that the functional relationships be of the form $\mathrm{p}(\epsilon_i^2),\mathrm{p}(\boldsymbol{h}_i\,|\,\epsilon_i^2),\mathrm{p}(\sigma_s^2),\mathrm{p}(\boldsymbol{\lambda}_s)$, and $\mathrm{p}(\boldsymbol{y}_s\,|\,H,(\boldsymbol{h}_i),\sigma_s^2,\boldsymbol{\lambda}_s)$. But these conditional distributions can, in turn, be chosen at will. On the other hand, the graphical model can be made more specific for our purpose, so that it exactly fits the GLM. Identifying all functional relationships of the network to their counterparts for model (H) then makes the DAG a perfect representation of the GLM. $\mathrm{p}(\epsilon_i^2)$ and $\mathrm{p}(\boldsymbol{h}_i\,|\,\epsilon_i^2)$ can be chosen to a scaled inverse-χ^2 and a Gaussian distribution, respectively, as detailed in Section II-C. $\mathrm{p}(\sigma_s^2)$ and $\mathrm{p}(\boldsymbol{\lambda}_s)$ can be set as in Section II-D. Finally, $\mathrm{p}(\boldsymbol{y}_s\,|\,H,(\boldsymbol{h}_i),\sigma_s^2,\boldsymbol{\lambda}_s)$ can be set as in (2).

C. Numerical Inference

To obtain a numerical approximation of the joint posterior pdf, we apply Gibbs sampling. This consists of starting with a seed vector and sequentially modifying one vector component at a time by sampling according to the conditional pdf of that component given the remaining variables. Samples are composed of the set of all vectors whose components have been updated an equal amount of times.

A key issue with Gibbs sampling is to partition the vector of all parameters into components whose conditional sampling can easily be performed. Another one is derivation of the conditional pdfs corresponding to the chosen clustering. In our case, both questions are answered at once, thanks to the previous step of graph modeling. As a matter of fact, it first allows us to decompose the parameter vector onto its $2I+2S$ canonical components: I ϵ_i^2's and $\boldsymbol{h}_i$'s, S σ_s^2's, and $\boldsymbol{\lambda}_s$'s. All $\boldsymbol{y}_s$'s being given, no sampling needs to be done on these variables. The updating steps are performed on these

variables; we, therefore, need access to the following conditional pdfs: $\mathrm{p}(\epsilon_i^2 \,|\, H, \mathrm{r.v.}), \mathrm{p}(\boldsymbol{h}_i \,|\, H, \mathrm{r.v.}), \mathrm{p}(\sigma_s^2 \,|\, H, \mathrm{r.v.})$, and $\mathrm{p}(\boldsymbol{\lambda}_s \,|\, H, \mathrm{r.v.})$. But these are just the conditional distributions given by Pearl's theorem and structurally developed in (7). Integration of the exact functional relationships into (7) leads to (see the Appendix for a summary of the properties used)

- According to (7a), $(\epsilon_i^2 \,|\, H, \mathrm{r.v.})$ is proportional to the product of two inverse-chi-square distributions in ϵ_i^2; it is, hence, also inverse-chi-square distributed
$$\mathrm{p}\left(\epsilon_i^2 \,\middle|\, H, \mathrm{r.v.}\right) = \text{Inv-}\chi^2\left(\mu_i, \tau_i^2; \epsilon_i\right)$$
with
$$\mu_i = n_\epsilon + (K_i - 1)$$
$$\tau_i^2 = \frac{n_\epsilon r_\epsilon^2 + \boldsymbol{h}_i^t \boldsymbol{R} \boldsymbol{h}_i}{n_\epsilon + (K_i - 1)}.$$
- According to (7b), $(\boldsymbol{h}_i \,|\, H, \mathrm{r.v.})$ is proportional to the product of two multivariate Normal distributions in $\boldsymbol{h}_i$; it is, hence, also multivariate Normal distributed
$$\mathrm{p}(\boldsymbol{h}_i \,|\, H, \mathrm{r.v.}) = \mathcal{N}(\boldsymbol{\delta}_i, \boldsymbol{\Delta}_i; \boldsymbol{h}_i)$$
with
$$\boldsymbol{\Delta}_i = \left(\frac{1}{\epsilon_i^2}\boldsymbol{R} + \sum_s \frac{1}{\sigma_s^2}\boldsymbol{X}_{s,i}^t \boldsymbol{X}_{s,i}\right)^{-1}$$
$$\boldsymbol{\delta}_i = \boldsymbol{\Delta}_i \left(\sum_s \frac{1}{\sigma_s^2}\boldsymbol{X}_{s,i}^t \cdot \left(\boldsymbol{y}_s - \sum_{j \neq i} \boldsymbol{X}_{s,j}\boldsymbol{h}_j - \boldsymbol{D}_s\boldsymbol{\lambda}_s\right)\right).$$
- According to (7c), $(\sigma_s^2 \,|\, H, \mathrm{r.v.})$ is proportional to the product of two inverse-chi-square distributions in σ_s^2; it is, hence, also inverse-chi-square distributed
$$\mathrm{p}\left(\sigma_s^2 \,\middle|\, H, \mathrm{r.v.}\right) = \text{Inv}\boldsymbol{-}\chi^2\left(\nu_s, \omega_s^2; \sigma_s^2\right)$$
with
$$\nu_s = n_{\sigma,i} + L_s$$
$$\omega_s^2 = \frac{n_{\sigma,s} r_{\sigma,i}^2 + \|\boldsymbol{y}_s - \sum_i \boldsymbol{X}_{s,i}\boldsymbol{h}_i - \boldsymbol{D}_s\boldsymbol{\lambda}_s\|^2}{n_{\sigma,i} + L_s}.$$
- According to (7d), $(\boldsymbol{\lambda}_s \,|\, H, \mathrm{r.v.})$ is proportional to the product of two multivariate Normal distributions in $\boldsymbol{\lambda}_s$; it is, hence, also multivariate Normal distributed
$$\mathrm{p}(\boldsymbol{\lambda}_s \,|\, H, \mathrm{r.v.}) = \mathcal{N}(\boldsymbol{\gamma}_s, \boldsymbol{\Gamma}_s; \boldsymbol{\lambda}_s)$$
with
$$\boldsymbol{\Gamma}_s = \left(\boldsymbol{V}_{\lambda,s}^{-1} + \frac{1}{\sigma_s^2}\boldsymbol{D}_s^t \boldsymbol{D}_s\right)^{-1}$$
$$\boldsymbol{\gamma}_s = \boldsymbol{\Gamma}_s \left(\boldsymbol{V}_{\lambda,s}^{-1}\boldsymbol{m}_{\lambda,s} + \frac{1}{\sigma_s^2}\boldsymbol{D}_s^t \left(\boldsymbol{y}_s - \sum_i \boldsymbol{X}_{s,i}\boldsymbol{h}_i\right)\right).$$

The sampling can then be performed by sequentially updating the ϵ_i^2's, the $\boldsymbol{h}_i$'s, then the σ_s^2's, and finally the $\boldsymbol{\lambda}_s$'s. Convergence monitoring is performed component-wise using parallel sampling as detailed in [29]. More precisely, we first take the logarithm of ϵ_i^2 and σ_s^2, so that all variables are spanned from $-\infty$ to $+\infty$. For each estimand $\phi, \phi = \log\epsilon_i^2, \log\sigma_s^2, h_{i,k}$, and $\lambda_{s,m}$, we draw B parallel sequences of length C (we typically took $B = 10$ and $C = 50$), each sample being denoted $\phi^{[bc]}$, with $b = 1, \ldots, B$ and $c = 1, \ldots, C$. We then compute the between-sequence variance BV, and the within-variance sequence WV as follows:

$$\mathrm{BV} = \frac{C}{B-1}\sum_{b=1}^{B}(\bar{\phi}^{[b\cdot]} - \bar{\phi}^{[\cdot\cdot]})^2$$

with

$$\bar{\phi}^{[b\cdot]} = \frac{1}{C}\sum_{c=1}^{C}\phi^{[bc]} \quad \text{and} \quad \bar{\phi}^{[\cdot\cdot]} = \frac{1}{B}\sum_{b=1}^{B}\phi^{[b\cdot]}$$

and

$$WV = \frac{1}{B}\sum_{b=1}^{B}(s^2)^{[b]}$$

where

$$(s^2)^{[b]} = \frac{1}{C-1}\sum_{c=1}^{C}\left(\phi^{[bc]} - \bar{\phi}^{[b\cdot]}\right)^2.$$

We then calculate

$$\sqrt{\hat{R}} = \sqrt{1 + \frac{1}{C}\left(\frac{\mathrm{BV}}{WV} - 1\right)}$$

for each scalar estimand. These quantities are supposed to decline to 1 as the sampling converges. We stop the algorithm when all $\sqrt{\hat{R}}$ are close enough to 1, e.g., smaller than 1.1, and remove α percent of each chain to account for a burn-in period.

We are admittedly mostly interested in the HRFs, but knowledge of the values taken by the other parameters are relevant as well for our analysis and a better understanding of brain processing. Gibbs sampling gives us access to estimates for all parameters or any quantity of interest related to them. For instance, in this paper, parameter estimators are given as

$$\text{posterior mean} \pm \text{posterior standard deviation}.$$

Once Gibbs sampling has converged, these quantities are approximated by their sample counterparts.

IV. Simulations

We simulated data with two HRFs ($I = 2$), as depicted in Fig. 4. To obtain the ϵ_i's corresponding to these HRFs, we calculated them as follows. If we knew that $\boldsymbol{h}_i = \boldsymbol{h}_i^0$, then we could infer ϵ_i^2 using Bayes' theorem

$$\mathrm{p}\left(\epsilon_i^2 \,|\, H', \boldsymbol{h}_i = \boldsymbol{h}_i^0\right) \propto \mathrm{p}\left(\epsilon_i^2 \,\middle|\, H'\right) \cdot \mathrm{p}\left(\boldsymbol{h}_i = \boldsymbol{h}_i^0 \,\middle|\, H', \epsilon_i^2\right)$$

where $\mathrm{p}(\boldsymbol{h}_i = \boldsymbol{h}_i^0 \,|\, H', \epsilon_i^2)$ would be given by (3), and $\mathrm{p}(\epsilon_i^2 \,|\, H')$ could be taken as a uniform prior, asuming no particular prior knowledge. $(\epsilon_i^2 \,|\, H', \boldsymbol{h}_i = \boldsymbol{h}_i^0)$ would then be Inv-χ^2 distributed and

$$\mathrm{E}\left[\epsilon_i^2\right] = \frac{1}{K-1} \cdot \boldsymbol{h}_i^{0t}\boldsymbol{R}\boldsymbol{h}_i^0.$$

This last relation can, in turn, be taken as estimate for ϵ_i^2, leading to

$$\epsilon_1^2 = 1.21$$
$$\epsilon_2^2 = 0.30.$$

For the simulation, we also took two sessions of $N = 100$ time samples. Δt and the sampling interval were both set to

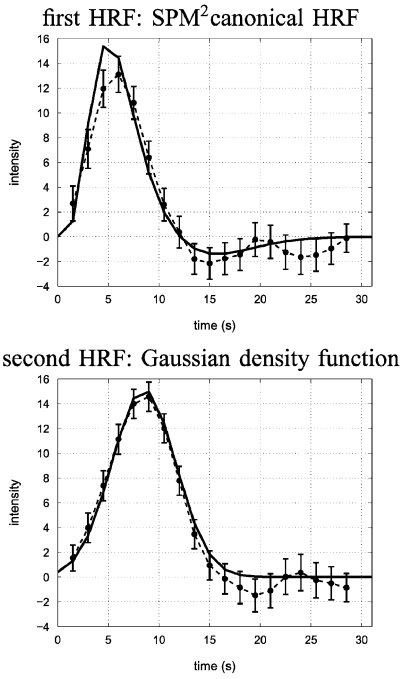

Fig. 4. Simulations. Estimated (*dashed line*) and simulated (*solid line*) HRF (Available: www.fil.ion.ucl.ac.uk/spm/spm99.html).

1.5 s. Quadratic drifts ($p_1(t) = 846 + 0.2 \cdot t + 0.001 \cdot t^2$ and $p_2(t) = 950 + 0.15 \cdot t + 0.0011 \cdot t^2$) and Gaussian white noises ($\sigma_1^2 = 50, \sigma_2^2 = 100$) were also added. Note that the noise standard deviations are about of the same amplitude as the HRF. For the analysis, we set both orders to $K = 20$, and took a quadratic drift in consideration ($M = 3$) with each $\boldsymbol{m}_{\lambda,s}$ set to $(\mathrm{E}[\boldsymbol{y}_s], 0, 0)^t$, and $\boldsymbol{V}_{\lambda,s}$ to a diagonal matrix such, that $\mathrm{diag}(\boldsymbol{V}_{\lambda,s}) = (10\,000^2\ 1\,000^2\ 1\,000^2)$. All $n_{\sigma,s}$ and n_ϵ were set to 1 to implement vague priors for σ_s^2's and ϵ_i^2's. As to the remaining hyperparameters, the goal was to set them, so that they provide a magnitude order for the corresponding variables as follows:

- each $r_{\sigma,s}^2$ was set to $\mathrm{Var}[\boldsymbol{y}_s]$;
- r_ϵ^2 was set to

$$\frac{\mathrm{E}_s[\mathrm{Var}[\boldsymbol{y}_s]]}{\max[\boldsymbol{h}_{\mathrm{SPM}}]} \cdot \left[\frac{1}{K-1} \cdot \boldsymbol{h}_{\mathrm{SPM}}^t \boldsymbol{R} \boldsymbol{h}_{\mathrm{SPM}}\right]$$

  where $\boldsymbol{h}_{\mathrm{SPM}}$ is SPM canonical HRF and $\mathrm{E}_s[\mathrm{Var}[\boldsymbol{y}_s]]$ is a scaling factor.

Gibbs sampling took about 2 250 updates for each of the 10 parallel chains to converge, and we kept the last 10% of each chain. We obtained the following estimates:

$$\begin{aligned}
\epsilon_1^2 &\approx 1.29 \pm 0.80 \\
\epsilon_2^2 &\approx 0.936 \pm 0.524 \\
\sigma_1^2 &\approx 38.6 \pm 7.8 \\
\sigma_2^2 &\approx 109 \pm 20 \\
\lambda_{1,0} &\approx 858 \pm 4 \\
\lambda_{2,0} &\approx 956 \pm 4.
\end{aligned}$$

Estimates for ϵ_i^2, σ_s^2, and $\boldsymbol{\lambda}_s$ were accurate. As shown in Fig. 4, HRF estimates were also very accurate for the noise level considered.

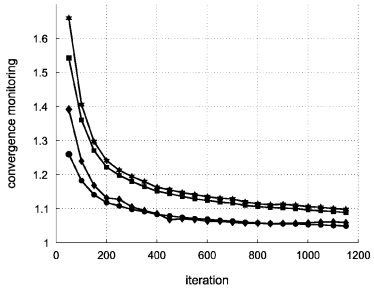

Fig. 5. Convergence monitoring. Every 50 steps is represented $\max[\sqrt{\hat{R}}]$ taken among all ϵ_i^2 (diamonds), σ_s^2 (circles), $h_{i,k}$ (squares), and $\lambda_{s,m}$ (stars).

V. Real Data

Eleven healthy subjects (age 18–40) were scanned while performing a motor sequence learning task. Using a joystick, they were asked to reach a target projected on a screen for 3 s, following an elliptic curve as precisely and rapidly as possible. They had to complete 64 trials of sequence (SEQ) mode (the targets appeared in a predefined order, unknown to the subject, to form a 8-item-long sequence) and 16 trials of random (RAN) mode (the targets appeared pseudorandomly). The time interval beween two consecutive trials, or inter-stimulus interval, was randomly selected to uniformly lie between 3 and 4 s. Functional T_2^*-weighted acquisitions were performed on a 3 T Bruker MEDSPEC 30/80 MR system (TR: 3 486 ms, TE: 35 ms, flip angle: 90°, matrix $64 \times 64 \times 42$, voxel size $3 \times 3 \times 3$ mm).

The data imply to work with $I = 2$ HRFs ($i = 1$ and $i = 2$ corresponding to SEQ and RAN, respectively) and $S = 5$ sessions. For the analysis, we first adjusted the stimulus on a grid of interval $\Delta t = \mathrm{TR}/5$. Both HRFs were assumed to have a common order $K = 5 \times 5$, for a total duration of 5 TRs. The prior hyperparameters were set as in Section IV. To illustrate the method, we selected two voxels, v_1 and v_2. v_1 was located in the right cerebellum and v_2 in the right inferotemporal lobe. Our goal was to estimate both HRFs corresponding to conditions SEQ and RAN, respectively.

Based on our monitoring system, convergence took around 1 300 iterations to occur. Fig. 5 shows a typical convergence curve. As showed in Fig. 6, the method was able to extract different HRF behaviors for different conditions, despite a very low signal-to-noise ratio. The high noise level was reflected in the large estimate error bars but did not prevent discrimination between conditions. The estimated values for the other parameters are given in Table I.

VI. Discussion

Our approach made it possible to associate the well-known GLM for HRF estimation in fMRI data analysis to a directed acyclic graph. This had the first advantage of making clear all modeling hypotheses. Moreover, in a Bayesian framework, the complex, yet central, step of calculating the joint posterior pdf was avoided. Instead, the graph provided us with a very convenient tool to first break down the set of all variables into coherent subsets, namely its nodes. Using the Markov properties, it was direct to derive all conditional pdfs that were required for Gibbs sampling as products of conditional pdfs that have been

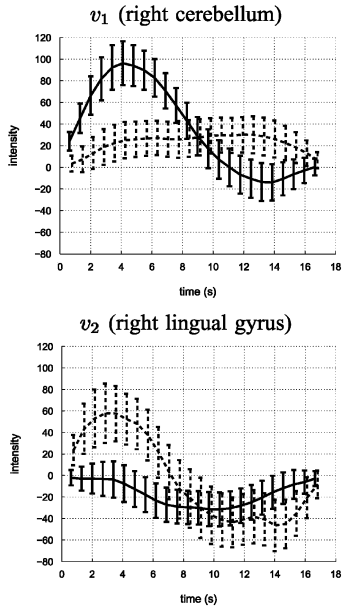

Fig. 6. Real data. HRFs corresponding to the SEQ (*solid line*) and RAN (*dashed line*) stimuli. Each sample has been slightly shifted to the left (SEQ) or to the right (RAN) for a better graphical rendering.

TABLE I
ESTIMATES OF THE SMOOTHING PARAMETERS ϵ_i, THE NOISE VARIANCES σ_s^2, AND THE BASELINES $\lambda_{s,1}$

		ϵ_i
v_1	$i=1$	$3.83e4 \pm 2.90e4$
	$i=2$	$2.76e4 \pm 5.30e4$
v_2	$i=1$	$2.43e4 \pm 1.90e4$
	$i=2$	$1.11e5 \pm 1.33e5$

		σ_s^2
v_1	$s=1$	$1.75e4 \pm 2.67e3$
	$s=2$	$5.02e3 \pm 8.24e2$
	$s=3$	$6.82e3 \pm 1.10e3$
	$s=4$	$5.38e3 \pm 8.55e2$
	$s=5$	$5.37e3 \pm 8.36e2$
v_2	$s=1$	$2.53e4 \pm 3.91e3$
	$s=2$	$4.53e3 \pm 7.08e2$
	$s=3$	$1.05e4 \pm 1.73e3$
	$s=4$	$9.11e3 \pm 1.46e3$
	$s=5$	$5.08e3 \pm 8.26e2$

		$\lambda_{s,1}$
v_1	$s=1$	$3.79e3 \pm 21.3$
	$s=2$	$3.86e3 \pm 18.3$
	$s=3$	$3.50e3 \pm 18.4$
	$s=4$	$3.65e3 \pm 18.1$
	$s=5$	$3.37e3 \pm 17.8$
v_2	$s=1$	$3.94e4 \pm 25.1$
	$s=2$	$3.48e3 \pm 20.2$
	$s=3$	$3.74e3 \pm 22.0$
	$s=4$	$3.52e3 \pm 21.4$
	$s=5$	$3.26e3 \pm 20.4$

specified with the modeling. Fully probabilistic numerical inference was then straightforward at a reasonable time cost. Furthermore, solving the same problem with several variables set to certain values was an easy matter, since all that must be done is removing these variables from the sampling scheme.

Because variable partitioning is implied by the graph structure, our application of Gibbs sampling differs from classical applications. For instance, we sequentially sampled each HRF, resp. drift vector, resp. noise variance, whereas a conventional procedure would simultaneously sample all HRFs, then all drift parameters, and so on. What influence this difference makes on the convergence speed of the Markov chain is a matter that needs to be further investigated. Another point would be to compare our method to a more conventional procedure where part, or all, of the nuisance parameters (e.g., drift parameters) are integrated out of the joint posterior pdf, and inference is done on the posterior marginal pdf of the HRFs. In the "marginal" scheme, each sample is performed on a lower dimension, and the Markov Chain is also of lower dimension. For these reasons, it is expected that convergence will be faster. This scheme is, however, very sensitive to model changes and makes inference on other parameters tedious.

Influence of the prior parameters is of importance to check sensitivity of the estimates. $\boldsymbol{m}_\lambda$ and $\boldsymbol{V}_\lambda$ were found to have very little weight on the inference, n_σ and r_σ some more. Selection of n_ϵ and r_ϵ did not much matter for drift and noise parameters, but had a dramatic influence on HRF estimation. The ad hockery proposed to manually set these hyperparameters seems to be useful, but its influence on the sampling scheme is still unclear. Indeed, except for this, we did not even have a prior idea of the magnitude order of ϵ_i^2, whereas scaled inverse-chi-square pdfs are relatively localized around their mode. To remedy this flaw, we suggest that setting priors for $\log(\epsilon_i^2)$ would prove more adapted to the state of ignorance that we are in relative to these parameters than conjugate priors. A general procedure is proposed in [30] to sample from pdfs that have the structure implied by (6) using rejection sampling. Another improvement would be to modify the model, so that ϵ_i becomes independent of the HRF intensity. This could be achieved by consideration of a normalized HRF and a scaling factor. Choice of noninformative, i.e., improper,[2] priors might also help to circumvent that problem. The reason why we did not consider this option here is that it is safe to use improper priors as long as the posterior pdf can be proved to be proper. Since one advantage of our method is to avoid calculation of the posterior pdf, this step cannot be performed in our setting without making it lose some of its appeal. Nonetheless, we are optimistic and believe that further investigation could justify use of improper priors in a restricted fashion.

In the framework of DAG modeling, the local properties of relationships renders the model very simple to structurally or functionally modify at a local level, either because it does not correctly explain the phenomenon under interest, or because a more complex model is sought. As a matter of fact, the proposed model for HRF estimation can already be seen as an improvement of the graphical model associated to the basic one-HRF, one-session linear model. Fig. 7 illustrates how this model can successively be expanded by introduction of smoothing priors, several sessions, and several stimulus types. These models can be dealt with as efficiently as the one detailled in this paper. Their joint posterior distribution would be given by application of (5), whereas all conditional distributions required by Gibbs sampling could easily be obtained through (6).

Besides, consideration of local spatial information, as in [23], [24], [31] could be achieved by gathering all voxel graphical models that were here assumed to be independent from each other and adding relationships between neighboring $\boldsymbol{h}_{v,i}$'s. Another point would be to relax the assumption that all HRFs are

[2]A prior pdf p is improper if its integral is not finite, i.e.,

$$\int p(x)\,\mathrm{d}x = \infty.$$

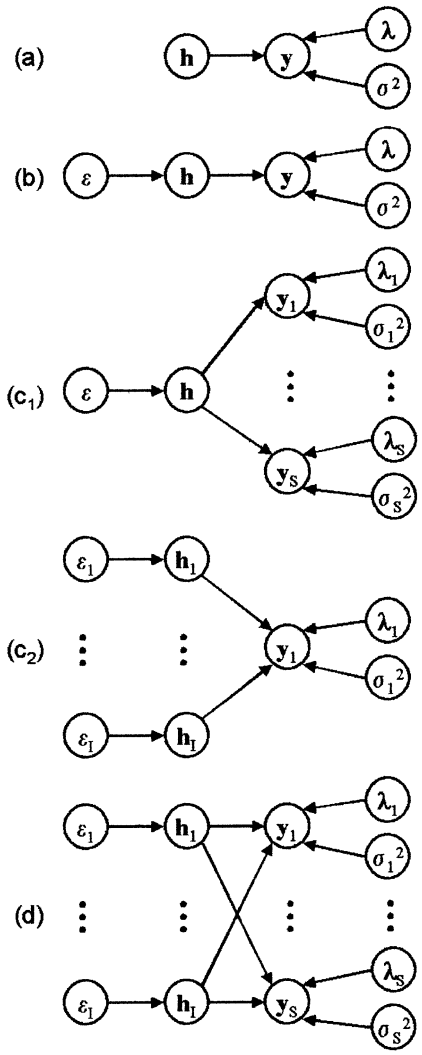

Fig. 7. Successives complexification of the HRF model. (a) basic model; (b) model with hyperparameter for smoothness regularization; (c_1) model accounting for one stimulus type and several sessions; (c_2) model accounting for several stimulus types and one sessions; (d) GLM considered in this paper.

constant across sessions and rather assume that they share the same shape accross sessions, with an amplitude that can vary, as discussed in [22]. As more and more information is incorporated into the model, the corresponding graph will become more and more complex. However, tools have been developed to deal with such graphs. Parallel processing of Gibbs sampling can be implemented. To avoid the problem of simultaneous updating of neighboring variables, one has to apply the so-called "edge reversal" control policy, as detailed in [28]. For huge graphs, [32] proposed an efficient variant of Gibbs sampling.

We finally believe that this novel approach has a much broader application range than just fMRI data analysis. Indeed, we are confident in the fact that any Bayesian model can be embedded in a graphical framework. This would allow to concentrate on the modeling, since efficient and automated inference would directly derive from the model.

VII. Conclusion

In this paper, we proposed a novel Bayesian inference framework for HRF estimation in fMRI data analysis, based on translating the existing Bayesian model into a Bayesian network to combine the features of graphical modeling and Bayesian analysis. This approach makes extensive use of Bayesian networks to 1) represent the model in a compact, yet efficient way, and 2) lead probabilistic inference through Gibbs sampling. This technique takes advantage of Markov properties of DAGs. Models can easily be designed, and both structural (i.e., of independence) and functional relationships are clearly presented. Moreover, using Gibbs sampling on the DAG, fully probabilistic numerical inference is straightforward. Ongoing research includes integration of more diffuse prior pdfs when necessary, as well as spatial constraints for the HRFs.

Appendix

The results detailed here originate from [29].

A. Multivariate Normal Distribution

If $\boldsymbol{x}$ is a d-dimensional multivariate Normal distributed variable with mean $\boldsymbol{m}$ and covariance matrix $\boldsymbol{V}$, then

$$\mathrm{p}(\boldsymbol{x}) = (2\pi)^{d/2}|\boldsymbol{V}|^{-1/2}\exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{m})^t\boldsymbol{V}^{-1}(\boldsymbol{x}-\boldsymbol{m})\right].$$

If $\boldsymbol{x}$ has a probability distribution defined by

$$\mathrm{p}(\boldsymbol{x}) \propto f_1(\boldsymbol{x})\cdot f_2(\boldsymbol{x})$$

where f_i is multivariate Gaussian with mean $\boldsymbol{m}_i$ and covariance matrix $\boldsymbol{V}_i$, then $\boldsymbol{x}$ is multivarite Gaussian with mean

$$\boldsymbol{m} = \left(\boldsymbol{V}_1^{-1}+\boldsymbol{V}_2^{-1}\right)^{-1}\left(\boldsymbol{V}_1^{-1}\boldsymbol{m}_1+\boldsymbol{V}_2^{-1}\boldsymbol{m}_2\right)$$

and covariance matrix

$$\boldsymbol{V} = \left(\boldsymbol{V}_1^{-1}+\boldsymbol{V}_2^{-1}\right)^{-1}.$$

B. Inverse-Chi-Square Distribution

If y is inverse-chi-square distributed with n degrees of freedom and scale s, then

$$\mathrm{p}(y) = \frac{(n/2)^{n/2}}{\Gamma(n/2)}s^n y^{-(n/2+1)}e^{-ns^2/(2y)}.$$

If y has a probability distribution defined by

$$\mathrm{p}(y) \propto f_1(y)\cdot f_2(y)$$

where f_i is inverse-chi-square distributed with n_i degrees of freedom scale s_i, then y is inverse-chi-square with

$$n = n_1 + n_2$$

degrees of freedom and scale s so, that

$$s^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1+n_2}.$$

Acknowledgment

The authors are grateful to Pr. J. Doyon (Institut de Gériatrie, Université de Montréal, Canada) for providing them with the data and to C. Posé for her technical support.

References

[1] W. Chen and S. Ogawa, "Principles of BOLD functional MRI," in *Functional MRI*, C. Moonen and P. Bandettini, Eds, Berlin: Springer, 1999, pp. 103–113.

[2] T.-Q. Li, T. N. Haefelin, B. Chan, A. Kastrup, T. Jonsson, G. H. Glover, and M. E. Moseley, "Assessment of hemodynamic response during focal neural activity in human using bolus tracking, arterial spin labeling and BOLD techniques," *Neuroimage*, vol. 12, pp. 442–451, 2000.

[3] A. Aubert and R. Costalat, "A model of the coupling between brain electrical activity, metabolism, and hemodynamics: Application to the interpretation of functional neuroimaging," *Neuroimage*, vol. 17, pp. 1162–1181, 2002.

[4] K. Uğurbil, L. Toth, and D.-S. Kim, "How accurate is magnetic resonance imaging of brain function?," *Trends Neurosci.*, vol. 26, pp. 108–114, 2003.

[5] K. J. Friston, P. Jezzard, and R. Turner, "Analysis of functional MRI time-series," *Hum. Brain Mapp.*, vol. 1, pp. 153–171, 1994.

[6] A. M. Dale and R. L. Buckner, "Selective averaging of rapidly presented individual trials using fMRI," *Hum. Brain Mapp.*, vol. 5, pp. 329–340, 1997.

[7] R. L. Buckner, "Event-related fMRI and the hemodynamic response," *Hum. Brain Mapp.*, vol. 6, pp. 373–377, 1998.

[8] C. R. Genovese, "A Bayesian time-course model for functional magnetic resonance imaging data," *J. Amer. Statist. Assoc.*, vol. 95, pp. 691–719, 2000.

[9] C. Gössl, L. Fahrmeir, and D. P. Auer, "Bayesian modeling of the hemodynamic response function in BOLD fMRI," *Neuroimage*, vol. 14, pp. 140–148, 2001.

[10] R. L. Buckner, W. Koutstaal, D. L. Schacter, A. D. Wagner, and B. R. Rosen, "Functional-anatomic study of episodic retrieval using fMRI (I)," *Neuroimage*, vol. 7, pp. 151–162, 1998.

[11] R. L. Buckner, W. Koutstaal, D. L. Schacter, A. M. Dale, M. Rotte, and B. R. Rosen, "Functional-anatomic study of episodic retrieval using fMRI (II)," *Neuroimage*, vol. 7, pp. 163–175, 1998.

[12] G. K. Aguirre, E. Zarahn, and M. D'Esposito, "The variability of human, BOLD hemodynamic responses," *Neuroimage*, vol. 8, pp. 360–369, 1998.

[13] M. D'Esposito, E. Zarahn, G. K. Aguirre, and B. Rypma, "The effect of normal aging on the coupling of neural activity to the BOLD hemodynamic response," *Neuroimage*, vol. 10, pp. 6–14, 1999.

[14] M. D'Esposito, L. Y. Deouel, and A. Gazzaley, "Alterations in the BOLD fMRI signal with ageing and disease: A challenge for neuroimaging," *Nat. Neurosci.*, vol. 4, pp. 863–872, 2003.

[15] J. Kershaw, S. Abe, K. Kashikura, X. Zhang, and I. Kanno, "A Bayesian approach to estimating the haemodynamic response function in event-related fMRI," *Neuroimage*, vol. 11, p. S474, 2000.

[16] M. A. Burock and A. M. Dale, "Estimation and detection of event-related fMRI signals with temporally correlated noise: A statistically efficient and unbiased approach," *Hum. Brain Mapp.*, vol. 11, pp. 249–260, 2000.

[17] C. Goutte, F. Å. Nielsen, and L. K. Hansen, "Modeling the haemodynamic response in fMRI using smooth FIR filters," *IEEE Trans. Med. Imag.*, vol. 19, pp. 1188–1201, Dec. 2000.

[18] G. Marrelec, H. Benali, P. Ciuciu, and J.-B. Poline, "Bayesian estimation of the hemodynamic response function in functional MRI," in *Proc. MaxEnt Workshop*, R. Fry, Ed., 2001, pp. 229–247.

[19] G. Marrelec, H. Benali, P. Ciuciu, M. Pélégrini-Issac, and J.-B. Poline, "Robust Bayesian estimation of the hemodynamic response function in event-related BOLD fMRI using basic physiological information," *Hum. Brain Mapp.*, vol. 19, pp. 1–17, 2003.

[20] P. Ciuciu, G. Marrelec, J. Idier, H. Benali, and J.-B. Poline. A general tool to estimate the hemodynamic response function in fMRI data. presented at OHBM Proc. [CD-ROM].

[21] P. Ciuciu, G. Marrelec, J.-B. Poline, J. Idier, and H. Benali, "Robust estimation of the hemodynamic response function in asynchronous multitasks multisessions event-related fMRI paradigms," in *Proc. IEEE ISBI*, 2002, pp. 847–850.

[22] P. Ciuciu, J.-B. Poline, G. Marrelec, J. Idier, C. Pallier, and H. Benali, "Unsupervised robust nonparametric estimation of the hemodynamic response function for any fMRI experiment," *IEEE Trans. Med. Imag.*, vol. 22, pp. 1235–1251, Oct. 2003.

[23] P. Ciuciu, J. Idier, G. Marrelec, and J.-B. Poline, "Estimation régularisée et non supervisée de la fonction de réponse hémodynamique en imagerie cérébrale fonctionnelle (IRMf)," in *Proc. GRETSI*, 2003, pp. 312–315.

[24] P. Ciuciu, J. Idier, and C. Pallier. (2003) Outlier detection for robust region-based estimation of the hemodynamic response function in event-related fMRI. [Online] SHFJ/CEA Tech. Rep.. Available: http://www.madic.org/biblio/en/Author/CIUCIU-P.php

[25] C. Robert, *Simulations par la méthode MCMC*, Paris: Economica, 1997.

[26] M.-H. Chen, Q.-M. Shao, and J. G. Ibrahim, *Monte Carlo Methods in Bayesian Computation*. New York: Springer, 2000.

[27] J. Pearl, *Causality: Models, Reasoning, and Inference*, Cambridge, U.K.: Cambridge Univ. Press, 2001.

[28] ——, "Evidential reasoning using stochastic simulation of causal models," *Artif. Intell.*, vol. 32, pp. 245–257, 1987.

[29] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis*, London, U.K.: Chapman & Hall, 1998.

[30] G. Marrelec and H. Benali, "Automated rejection sampling from product of distributions," *Comput. Statist.*, 2004, to be published.

[31] C. Gössl, D. Auer, and L. Fahrmeir, "Bayesian spatiotemporal inference in functional magnetic resonance imaging," *Biometrics*, vol. 57, pp. 554–562, 2001.

[32] C. S. Jensen, A. Kong, and U. Kjærulff, "Blocking Gibbs sampling in very large probabilistic expert systems," in *Int. J. Hum. Comput. Studies (Special Issue Real-World Applications of Uncertain Reasoning)*, vol. 42, 1993, Available: http://citeseer.nj.nec.com/jensen93blocking.html, pp. 647–666.

[33] G. Marrelec, P. Ciuciu, M. Pélégrini-Issac, and H. Benali, "Estimation of the hemodynamic response function in event-related functional MRI: Directed acyclic graphs for a general Bayesian framework," in *Lecture Notes in Computer Science*. ser. , C. Taylor and A. Noble, Eds, Berlin, Germany: Springer-Verlag, 2003, vol. 2732, Information Processing in Medical Imaging, pp. 635–646.