Computational Statistics and Data Analysis 163 (2021) 107297

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

www.elsevier.com/locate/csda

Assessing dynamic effects on a Bayesian matrix-variate dynamic linear model: An application to task-based fMRI data analysis

Johnatan Cardona Jiménez [a,b,*], Carlos A. de B. Pereira [a]

[a] Institute of Mathematics and Statistics-University of São Paulo, Rua do Matão, 1010 - ZIP 05508-090 - São Paulo - SP, Brazil
[b] Facultad de Ingeniería, Institución Universitaria Pascual Bravo, Cl 73 No. 73 A 226 - ZIP 050034, Medellín, Colombia

ARTICLE INFO

Article history:
Received 8 July 2020
Received in revised form 15 April 2021
Accepted 29 May 2021
Available online 7 June 2021

Keywords:
Bayesian inference
Dynamic linear models
Monte Carlo integration
Task-based fMRI

ABSTRACT

A modeling procedure for task-based functional magnetic resonance imaging (fMRI) data analysis using a Bayesian matrix-variate dynamic linear model (MVDLM) is presented. With this type of model, less complex than the more traditional temporal-spatial models, it is possible to take into account the temporal and, at least locally, the spatial structures that are usually present in this type of data. Thus, every voxel in the brain image is jointly modeled with its nearest neighbors, as defined by a Euclidean metric. MVDLM's have been widely used in applications where the interest lies in performing predictions and/or analysis of covariance structures among time series. However, in this context, the interest is rather to assess the dynamic effects related to voxel activation. In order to do so, two algorithms are developed and an already-existing one is adapted to simulate on-line trajectories related to the state parameter. With those curves or simulated trajectories, a Monte Carlo evidence for voxel activation is computed. Through two practical examples of auditory- and motor-cortex activations and two different types of assessments using resting-state and simulated fMRI data, it is shown that the proposed method can be viewed by practitioners as a reliable tool for task-based fMRI data analysis. Despite the assessments and applications being illustrated just for a single-subject analysis, a description is given of how general group analysis can be implemented, exemplified with a group analysis for 21 subjects.

1. Introduction

Magnetic resonance imaging (MRI) is a non-invasive technique that is used to create elaborate anatomical images of the human body. Specifically, this technique can be used to obtain detailed brain images that can help to identify different types of tissue such as, for example, white matter and gray matter, and can also be used to diagnose aneurysms and tumors. Another important facet of this technique is that it can be used to visualize dynamic or functional activity in the brain. Functional Magnetic Resonance Imaging (fMRI) can be described as a generalization of the MRI technique, where the focus is not just one high-resolution image of the brain, but rather a sequence of low-resolution images that allows for identification, at least in an indirect way, of neuronal activity through the blood-oxygen-level-dependent (BOLD) contrast (Poldrack et al., 2011). The sequence of MRI images is obtained using an MRI scanner, essentially a huge magnet, which detects activity in

* Corresponding author.
E-mail addresses: jcardonj@ime.usp.br, jcardonj@unal.edu.co (J. Cardona Jiménez), scpereira@ime.usp.br (C.A. de B. Pereira).

https://doi.org/10.1016/j.csda.2021.107297

the brain's different regions by tracing blood flows. Statistical models are very useful for analyzing the post-processed data that compose the sequence of images obtained in an fMRI experiment. The most popular statistical model used to identify a brain-region reaction to an external stimulus is the normal regression linear model, usually known as the General Linear Model (GLM) in the fMRI literature ((Friston et al., 1994), (Beckmann et al., 2003)), which assumes spatiotemporal independence among the data, an unrealistic assumption that can lead to incorrect inference. For instance, Eklund et al. (2016) evaluate the most common software packages (SPM, FSL and AFNI) for fMRI analysis, which have the GLM implemented as one of their analysis tools. Specifically, they evaluate that model for group analysis using real resting-state data and they find that, for a nominal family-wise error rate of 5%, parametric statistical methods are shown to be conservative for voxel-wise inference and invalid for cluster-wise inference. However, the GLM is not the only alternative to model task-based fMRI data, there are other alternatives such as Bayesian models. Those Bayesian alternatives usually account for the spatiotemporal structure present in this type of data. For example, Zhang et al. (2016) propose a spatiotemporal nonparametric Bayesian approach to model both individual and group stages in one-step modeling. For posterior inference, they implement both a Markov Chain Monte Carlo (MCMC) algorithm and a suitable variational Bayes algorithm. Eklund et al. (2017) propose a Bayesian heteroscedastic GLM with autoregressive linear noise and heteroscedastic noise innovations for single-subject fMRI analysis. They develop an efficient MCMC algorithm that allows for variable selection among the regressors. Bezener et al. (2018) propose a Bayesian spatiotemporal modeling for single-subject fMRI analysis. Their modeling includes a novel areal model to parcel voxels into clusters and also use an MCMC algorithm for posterior inference. Yu et al. (2018) propose a Bayesian variable-selection approach to model complex-valued fMRI data. They develop their models applying complex-valued spike-and-slab priors on the parameters associated with brain activation and perform posterior inference via MCMC algorithms. Those methods are just some of the most recent works published in the field. See Zhang et al. (2015) and Bowman (2014) for a more complete review of statistical methods for fMRI data analysis. In this work, we propose local spatiotemporal Bayesian modeling for single-subject task-based fMRI data analysis and describe how more general group analysis can be implemented. However, here we do not resort to the traditional spatiotemporal models that are usually used to model fMRI data in the Bayesian context. Instead, we use a matrix-variate dynamic linear model (MVDLM) (see (West and Harrison, 1997, p. 581) for more details about these types of models). Specifically, we employ the type of models proposed by Quintana (1985) and Quintana (1987) in order to model the BOLD response related to block- and event-related experiments (see Kashou (2014) for a detailed explanation of these kinds of designs). Then, with this type of modeling, we can easily account for the temporal dependence present in this kind of data through the state or evolution equation. It also allows us to obtain closed-form posterior distributions for the parameters of interest and avoid the burden of MCMC algorithms or approximations to the posterior distribution. In most of the literature of fMRI data analysis, when a researcher wants to model the spatial dependencies, slices of the brain image are taken as the cluster of voxels for that purpose (see, for example, Friston and Penny (2003) and some of the references cited above). This approach can fail to be appropriate because there are nearest voxels (from slices located above and below) that the spatial modeling excludes, therefore discarding important information. Bearing this in mind and even though we take a voxel-wise approach, we model every voxel jointly with its nearest neighbors. Thus, we take advantage of the resulting covariance structure on the simulation process of on-line trajectories related to the state parameter, which are used to perform the voxel-wise inference related to brain activation. This is what we call local spatiotemporal modeling. To define the clusters or neighborhoods related to every voxel, we just employ a Euclidean distance. Thus, the modeler has to define a fixed distance r and all the voxels lying within that distance will be considered part of the cluster. This approach is not absent from possible criticism for excluding distant voxels. However, modeling the covariance structure among distant voxels is usually addressed on connectivity problems, and in this work, our aim is modeling task-based fMRI data. Before explaining how this modeling is used to detect brain activation, it is worth mentioning that most of the work related to MVDLM has been devoted to forecasting and analyzing covariance structure across several time series (Quintana and West (1987), Aguilar and West (2000), Fei et al. (2011)), however not much work has been developed regarding the inference of effect sizes related to regressors in an MVDLM. A good example of (univariate) dynamic linear models (DLM) applied in the context of fMRI data analysis is the work of Ho et al. (2005), where they propose a modeling procedure to tackle the problem of connectivity between brain regions. Despite defining a prior measure on the regressors' coefficient related to the expected BOLD response, their approach is not fully Bayesian. For instance, the connectivity analysis parameters are estimated using the EM algorithm, and their standard errors are obtained via Bootstrap estimation, which makes the estimation process more challenging. Despite obtaining on-line estimated trajectories of the state parameter related to the task activation, they do not offer or suggest how to evaluate the significance of the dynamic regression coefficient in the context of a brain reaction. (West and Harrison, 1997, p. 280) briefly mention how to perform inference about the dynamic regression coefficient or state parameter, but only for a particular time t within the interval of observed time. In the case of an fMRI modeling, we would like to perform inference on the state parameter for an entire interval of time at once. For that purpose, we propose three algorithms to draw on-line estimated trajectories of the state parameter and thus compute a measure of evidence via Monte Carlo integration. With that measure of evidence, we can perform inference about the state parameter or equivalently, in the context of task-based fMRI modeling, about brain activation. One algorithm is just an implementation of a forward-filtering-backward-sampling algorithm that was suggested by Frühwirth-Schnatter (1994) to draw state parameters for univariate dynamic linear models. In this matrix-variate setting, we use the same idea, but instead of using data augmentation to estimate individual parameters for the covariance matrix related to the state or evolution equation, we use discount factors to deal with those quantities (Ameen and Harrison (1985)). The other two algorithms can be described as forward-sampling schemes, in the sense that here we do not resort to

the filtering distributions and just use the posterior distributions computed at each observed time t. Briefly describing our modeling, at the individual stage, an MVDLM is fitted for every subject in the sample, and analysis such as brain activation and/or contrast between tasks can be performed using any of the algorithms proposed in this work. At the group stage, the individual posterior distributions for the state parameters are combined in a suitable way, obtaining new distributions that represent the updated information for the group effect. In this way, all of the analysis usually performed in an fMRI experiment (brain activation, task comparison, and group comparison) can be performed again by using any of the three algorithms proposed in this work. In order to show the performance of our method, we present two real-data examples related to auditory and motor task experiments respectively. We also assess our method in two ways. In the first, we follow the same approach as in Eklund et al. (2016) using resting-state fMRI data from healthy controls, obtained from the 1000 Functional Connectomes Project (Biswal et al., 2010). Here we define a fictitious stimulus and seek the rate of false positives given the null hypothesis of no brain reaction. The second way we assess our method is using simulated data obtained from the R package **neuRosim** (Welvaert et al., 2011).

In the next section, we give a brief description of the matrix-variate dynamic linear model and its posterior distribution under a matrix normal/inverse Wishart prior distribution. We also describe the modeling of the BOLD response using this type model for single and group analysis. Section 3 describes the three algorithms and computation of Monte Carlo evidence for voxel activation. In section 4, two examples for block- and event-related designs are presented to illustrate the performance of our proposal. In section 5, an assessment is performed to evaluate the capacity of our method both to correctly identify true-positive activations and to deal with false-positive activations. In the final section, we present some concluding remarks and potential open problems for future development.

2. MVDLM and fRMI modeling

The general theory of the MVDLM is presented in Quintana (1985), Quintana (1987) and West and Harrison (1997). Despite this model having been conceived for forecasting, in this work it is used for a different purpose: modeling fMRI data related to block- and event-related designs through a regression structure in order to identify brain reaction. In other words, we focus our interest on the estimation of, and inference on, the state parameter. Suppose we have a $q \times 1$ vector $\mathbf{Y}_t$, which can be modeled in terms of observation and evolution or state equations as follows.

$$\begin{array}{lll} \text{Observation:} & \mathbf{Y}_t' & = \mathbf{F}_t'\boldsymbol{\Theta}_t + \boldsymbol{\nu}_t' \\ \text{Evolution:} & \boldsymbol{\Theta}_t & = \mathbf{G}_t\boldsymbol{\Theta}_{t-1} + \boldsymbol{\Omega}_t, \end{array} \tag{1}$$

where for each t we have a $q \times 1$ vector $\boldsymbol{\nu}_t$ of observational errors, a $p \times q$ matrix $\boldsymbol{\Theta}_t$ of state parameters, and a $p \times q$ matrix $\boldsymbol{\Omega}_t$ of evolution errors. It is also supposed that $\boldsymbol{\nu}_t'$, $\boldsymbol{\Omega}_t$ and $\boldsymbol{\Theta}_{t-1}$ are independent (with $\boldsymbol{\nu}_t'$ and $\boldsymbol{\Omega}_t$ independent over time). The $1 \times p$ and $p \times p$ matrices $\mathbf{F}_t'$ and $\mathbf{G}_t$ respectively are common to each of the q univariate DLMs. The covariates related to the design being used, either a block- or an event-related design, as well as other characteristics of the subjects, can be included in the columns of $\mathbf{F}_t'$. It is supposed that $\nu_t \sim N_q[\mathbf{0}, V_t\boldsymbol{\Sigma}]$, independently over time, where $\boldsymbol{\Sigma}$ defines the cross-sectional covariance structure for the multivariate model and V_t is a known observational scale factor. For the random matrix $\boldsymbol{\Omega}_t$ the distribution is given by $\boldsymbol{\Omega}_t \sim N_{pq}[\mathbf{0}, \mathbf{W}_t, \boldsymbol{\Sigma}]$, a matrix-variate normal distribution with mean matrix $\mathbf{0}$, left $p \times p$ variance matrix $\mathbf{W}_t$ and right variance matrix $\boldsymbol{\Sigma}$ (Dawid (1981)). It is also supposed that the initial prior for $\boldsymbol{\Theta}_0$ and $\boldsymbol{\Sigma}$ is matrix normal/inverse Wishart, $(\boldsymbol{\Theta}_0, \boldsymbol{\Sigma}|D_0) \sim NW_{n_0}^{-1}[\mathbf{m}_0, \mathbf{C}_0, \mathbf{S}_0]$, for some known defining parameters $\mathbf{m}_0$, $\mathbf{C}_0$, $\mathbf{S}_0$ and n_0 ((Quintana, 1987)[Chapter 3]). $D_t = \{\mathbf{Y}_0, \mathbf{Y}_1, \ldots, \mathbf{Y}_t\}$ are the data observed at time t. The main justification for the normal model for the response variable and normal/inverse Wishart for the unknown quantities is conjugacy, which allows us to obtain a closed-form expression for the posterior distribution. However, even in contexts where conjugacy is not attained, like in some of the references cited above, the Gaussian model for both responses and parameters is commonly assumed in the context of fMRI modeling and has been shown to work in practice. Nevertheless, it is worth mentioning there are non-Gaussian alternatives for both the model (1) and in the field of fMRI data analysis. The interested reader can find more about these models in West et al. (1985); Gamerman (1998) and Wegmann et al. (2017) respectively.

2.1. Voxel-wise individual analysis

One of the main objectives in an fMRI experiment is to identify a brain reaction in response to some controlled external stimuli, in other words, to look for changes in the BOLD signal related to some experimental manipulation. In order to do that, one can use a block design, an event-related design, or possibly even a new design from the composition of those two (Kashou (2014)). The modeling we propose here relies on something called the expected BOLD response, which is obtained as the convolution of the stimulus time series f, which is obtained from the design being used, and the hemodynamic response function h (HRF), which can be interpreted as the picture over time of the change of the BOLD signal. In the examples and analysis presented here we use the canonical HRF, which is defined as the difference between two gamma densities. However, it is worth mentioning that any other form from the available options for the HRF can be used. This subject is discussed in depth in Poldrack et al. (2011) and Penny et al. (2011). In general, the expected BOLD response can be obtained as follows:

$$x(t) = (h * f)(t) = \int h(\tau) f(t-\tau) d\tau. \tag{2}$$

Thus, the aim is to identify the fMRI time series that matches the expected BOLD response. For that purpose, we model the observed BOLD response as a linear function of the expected BOLD response using the model (1).

2.1.1. Individual modeling

Let $y^{*(z)}_{[i,j,k],t,1}$ and $x^{(z)}(t)$ be the observed and expected BOLD response, respectively, at brain position $\{i, j, k\}$, time t, and subject z, for $i = 1, \ldots, d_1$, $j = 1, \ldots, d_2$, $k = 1, \ldots, d_3$, $t = 1, \ldots, T$ and $z = 1 \ldots, N_g$. Something that can be noticed from this last definition is that $x^{(z)}(t)$ is supposed to be the same for all the locations in the brain image, which means it is assumed that the BOLD response is the same in all brain regions. This may be an unrealistic assumption, but it is one that works well in practice. Let the vector

$$\mathbf{Y}^{(z)}_{[i,j,k],t} = \begin{pmatrix} y^{*(z)}_{[i,j,k],t,1} \\ y^{(z)}_{[i+1,j,k],t,2} \\ y^{(z)}_{[i-1,j,k],t,3} \\ y^{(z)}_{[i,j+1,k],t,4} \\ y^{(z)}_{[i,j-1,k],t,5} \\ y^{(z)}_{[i,j,k+1],t,6} \\ y^{(z)}_{[i,j,k-1],t,7} \end{pmatrix}_{1\times q}, \tag{3}$$

represent the cluster or neighborhood of size $q = 7$ of the voxel at position $\{i, j, k\}$ and time t. Thus, we model $\mathbf{Y}^{(z)}_{[i,j,k],t}$ using the model (1), with $\mathbf{G}_t = \mathbf{I}_p$ and $\mathbf{F}^{\prime(z)}_t = (x^{(z)}_1(t), \ldots, x^{(z)}_p(t))$, where $x^{(z)}_1(t), \ldots, x^{(z)}_p(t)$ are related to different types of stimuli or tasks performed in the fMRI experiment. What we intend with this type of modeling is to capture the covariance structure within the cluster of voxels through the matrix $\mathbf{\Sigma}^{(z)}$. The criterion to define the cluster form is based on the Euclidean distance, where the distance between voxel V^* and neighboring voxel $V_\sim$ is given by $d(V^*, V_\sim) \leq r$. In Fig. 1 (right panel), we can see a graphical illustration of a neighborhood of voxels related to the observed BOLD response vector (3) and its corresponding matrix parameter on equation (4), where $r = 1$ and consequently $q = 7$. In an R package (R Core Team (2018)) where the method proposed in this work has been implemented, a range of different r values can be set by the users: $r = 1, 2, 3, 4$, which consequently leads to $q = 7, 19, 27, 33$. This package is available at https://github.com/JohnatanLAB/BayesDLMfMRI_DataAndArticle.

$$\mathbf{\Theta}^{(z)}_{[i,j,k],t} = \begin{pmatrix} \theta^{*(z)}_{[i,j,k],t,1,1} & \theta^{(z)}_{[i,j,k],t,1,2} & \theta^{(z)}_{[i,j,k],t,1,3} & \theta^{(z)}_{[i,j,k],t,1,4} & \theta^{(z)}_{[i,j,k],t,1,5} & \theta^{(z)}_{[i,j,k],t,1,6} & \theta^{(z)}_{[i,j,k],t,1,7} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \theta^{*(z)}_{[i,j,k],t,p,1} & \theta^{(z)}_{[i,j,k],t,p,2} & \theta^{(z)}_{[i,j,k],t,p,3} & \theta^{(z)}_{[i,j,k],t,p,4} & \theta^{(z)}_{[i,j,k],t,p,5} & \theta^{(z)}_{[i,j,k],t,p,6} & \theta^{(z)}_{[i,j,k],t,p,7} \end{pmatrix}. \tag{4}$$

For the rest of the article, we change the notation for the location subscript from $[i, j, k]$ to v.

Posterior inference

Under the assumptions and prior distributions defined above and following the results in Dawid (1981), it can be shown that the joint posterior distribution of the state parameter $\mathbf{\Theta}^{(z)}_{v,t}$ and the covariance matrix $\mathbf{\Sigma}^{(z)}_v$ is given by

$$(\mathbf{\Theta}^{(z)}_{v,t}, \mathbf{\Sigma}^{(z)}_v | D_{vt}) \sim NW^{-1}_{n_t}[\mathbf{m}^{(z)}_{v,t}, \mathbf{C}^{(z)}_{v,t}, \mathbf{S}^{(z)}_{v,t}], \tag{5}$$

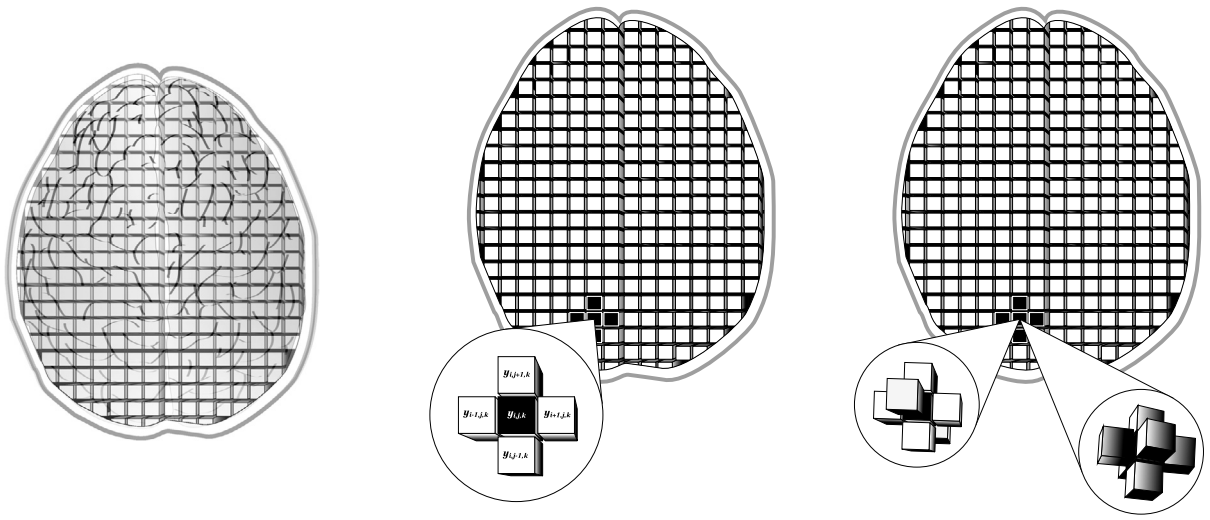
with

$$\begin{aligned} &\mathbf{m}^{(z)}_{v,t} = \mathbf{m}^{(z)}_{v,t-1} + \mathbf{A}^{(z)}_{v,t} \mathbf{e}^{\prime(z)}_{v,t}, \\ &\mathbf{C}^{(z)}_{v,t} = \mathbf{C}^{(z)}_{v,t-1} + \mathbf{W}^{(z)}_{v,t} - \mathbf{A}^{(z)}_{v,t} \mathbf{A}^{\prime(z)}_{v,t} Q^{(z)}_{v,t}, \\ &n_t = n_t + 1 \text{ and } \mathbf{S}^{(z)}_{v,t} = n_t^{-1}[n_{t-1}\mathbf{S}^{(z)}_{v,t-1} + \mathbf{e}^{\prime(z)}_{v,t} \mathbf{e}^{(z)}_{v,t} / Q^{(z)}_{v,t}], \end{aligned}$$

where

$$\begin{aligned} &Q^{(z)}_{v,t} = V_t + \mathbf{F}^{\prime(z)}_t [\mathbf{C}^{(z)}_{v,t-1} + \mathbf{W}^{(z)}_{v,t}] \mathbf{F}^{(z)}_t, \\ &\mathbf{A}^{(z)}_{v,t} = [\mathbf{C}^{(z)}_{v,t-1} + \mathbf{W}^{(z)}_{v,t}] F^{(z)}_t / Q^{(z)}_{v,t}, \end{aligned}$$

and



(a) Brain slice divided into voxels (b) BOLD response at voxel v (c) Cluster for $r = 1$

Fig. 1. Graphical illustration of a neighborhood of voxels, for $r = 1$.

$$\mathbf{e}_{v,t}^{(z)} = \mathbf{Y}_{v,t}^{(z)} - \mathbf{f}_{v,t}^{(z)}.$$

$\mathbf{f}_{v,t}^{(z)}$ is the mean of the predictive distribution $p(\mathbf{Y}_{v,t}^{(z)}|D_{v,t-1}^{(z)})$. Hence, the marginal posterior distribution for the state parameter is given by

$$(\mathbf{\Theta}_{v,t}^{(z)}|D_{v,t}^{(z)}) \sim T_{n_t}[\mathbf{m}_{v,t}^{(z)}, \mathbf{C}_{v,t}^{(z)}, \mathbf{S}_{v,t}^{(z)}], \tag{6}$$

which is known as a matrix T distribution with $p \times q$ mean matrix $\mathbf{m}_{v,t}^{(z)}$, $p \times p$ left variance matrix $\mathbf{C}_{v,t}^{(z)}$, and $q \times q$ right variance matrix $\mathbf{S}_{v,t}^{(z)}$. To deal with the unknown matrix parameter $\mathbf{W}_{v,t}^{(z)}$, we adopt the discount model approach proposed by Ameen and Harrison (1985) in the same manner as in West et al. (1985). We express the system variance matrix as $\mathbf{W}_{v,t}^{(z)} = \mathbf{B_t}\mathbf{C}_{v,t-1}^{(z)}\mathbf{B_t} - \mathbf{C}_{v,t-1}^{(z)}$, where $\mathbf{B_t}$ is a $p \times p$ diagonal matrix of positive discount factors $1/\beta_{it}^{1/2}$, for $i = 1, \ldots, p$ and $0 < \beta_{it} \leq 1$. As we show in one of the examples below, those discount factors are a key component of this modeling in the sense that they can affect the sensitivity of detection of brain activation.

A reasonable approximation for the posterior distribution (6) when $n_t \geq 30$ is given by

$$(\mathbf{\Theta}_{vt}^{(z)}|D_{v,t}^{(z)}) \overset{approx}{\sim} N_{pq}[\mathbf{m}_{v,t}^{(z)}, \mathbf{C}_{v,t}^{(z)}, \mathbf{S}_{v,t}^{(z)}]. \tag{7}$$

We have two main considerations to justify working only with posterior distributions for $n_t \geq 30$. The first is that as we use vague prior distributions at $t = 0$, the sequential update process takes some period of time before reaching a posterior distribution dominated by the data. Thus, the first posterior distributions (i.e., for $t < 30$) could be considered irrelevant for the analysis. The second consideration is simply that dealing with normal distributions simplifies the mathematical work when dealing with linear combinations, which are quite common in this work when inference is performed on the matrix parameter $\mathbf{\Theta}_{v,t}^{(z)}$ to detect brain activation.

As mentioned above, a brain activation related to the stimulus or task l (for $l = 1, \ldots, p$) is identified when the expected BOLD response ($x_l^{(z)}(t)$) matches the observed BOLD response ($\mathbf{Y}_{v,t}^{(z)}$). In other words, a brain activation in the cluster of voxels v related to the l-th task or stimulus is equivalent to the l-th row of $\mathbf{\Theta}_{v,t}^{(z)}$ being positive. In this sense, we define three different ways to evaluate brain activation, which are related to the following variables obtained from the components of $\mathbf{\Theta}_{v,t}^{(z)}$

$$\text{Marginal effect: } \theta_{v,t,l}^{*(z)} = \theta_{v,t,l,1}^{*(z)}, \tag{8}$$

$$\text{Average cluster effect: } \bar{\theta}_{v,t,l}^{(z)} = \frac{1}{q}\left[\theta_{v,t,l,1}^{*(z)} + \sum_{n=2}^{q} \theta_{v,t,l,n}^{(z)}\right], \tag{9}$$

$$\text{Joint effect: } \boldsymbol{\theta}_{v,t,l}^{(z)} = \left(\theta_{v,t,l,1}^{*(z)}, \theta_{v,t,l,2}^{(z)}, \ldots, \theta_{v,t,l,q}^{(z)}\right). \tag{10}$$

The marginal effect (8) is when just a marginal distribution from (7) is considered, in other words when the correlation captured by $\mathbf{S}_{v,t}^{(z)}$ is ignored. It can be shown that under particular considerations a DLM is equivalent to a traditional static

regression linear model (or GLM in the fMRI context), so an analysis using marginal effects for $t = T$ can yield results quite similar (or even the same when $\mathbf{W}_{v,t}^{(z)} = 0$) to those obtained when performing a voxel-wise approach using a GLM. The average (9) and joint (10) effects are more interesting in the sense that they take advantage of the information contained within the entire cluster of voxels. From the properties of the matrix normal distribution, the distributions of (8), (9) and (10) are given by

$$\theta_{v,t,l}^{*(z)} | D_{v,t}^{*(z)} \sim N(m_{v,t,l,1}^{*(z)}, C_{v,t,l,l}^{(z)} S_{v,t,1,1}^{(z)}), \tag{11}$$

$$\bar{\theta}_{v,t,l}^{(z)} | D_{v,t}^{(z)} \sim N(\bar{m}_{v,t,l}^{(z)}, \bar{S}_{v,t,l}^{(z)}), \tag{12}$$

$$\boldsymbol{\theta}_{v,t,l}^{(z)} | D_{v,t}^{(z)} \sim N_q(\boldsymbol{m}_{v,t,l}^{(z)}, C_{v,t,l,l}^{(z)} \boldsymbol{S}_{v,t}^{(z)}), \tag{13}$$

where $C_{v,t,l,l}^{(z)}$ and $S_{v,t,n,n}^{(z)}$ are the elements on the main diagonal of the matrices $\boldsymbol{C}_{v,t}^{(z)}$ and $\boldsymbol{S}_{v,t}^{(z)}$ respectively, $\bar{m}_{v,t,l} = \frac{1}{q}\sum_{n=1}^{q} m_{v,t,l,n}$, $\bar{S}_{v,t,l} = \frac{1}{q^2}\left[\sum_{n=1}^{q} C_{v,t,l,l} S_{v,t,n,n} + \sum_{n \neq n'} C_{v,t,l,l} S_{v,t,n,n'}\right]$, and $D_{v,t}^{*(z)}$ is the data related to the first component of the vector (3). With the distributions (11), (12) and (13), at least to our knowledge, it is only possible to perform inference (about (8), (9) and (10)) for a particular fixed time t, which in the context of fMRI means inference related to only one brain image. However, in task-based fMRI data analysis, it is more relevant to perform an inference procedure at once for the entire group of brain images just to understand brain dynamics related to the controlled stimulation presented on the fMRI experiment. In the next section, we present three algorithms that allow us to draw dynamic trajectories of the state parameter related to any of (8), (9) and (10) and use those trajectories to compute a measure of evidence for voxel activation on the entire observed interval of brain images. But before we get there, we first offer a brief explanation of how to get versions of (8), (9) and (10) for the group-stage analysis.

2.2. Voxel-wise group analysis

Now we describe the task-based fMRI group analysis where the focus is to detect an average group activation (single-group analysis) and/or compare voxel activation between two groups (e.g., patients vs. controls). Here we take any of the posterior distributions (11), (12), or (13), depending on the case, as an input for this stage. For instance, let us suppose that a sample of N_g subjects is part of an fMRI experiment where p stimuli are presented. Then, taking advantage of the properties of normal distribution, we obtain the following distributions for the group stage:

$$\text{Marginal group effect: } \bar{\theta}_{v,t,l}^{*(g)} \sim N(\bar{m}_{v,t,l}^{*(g)}, \bar{S}_{v,t,l}^{(g)}), \tag{14}$$

$$\text{Average cluster group effect: } \bar{\bar{\theta}}_{v,t,l}^{(g)} \sim N(\bar{\bar{m}}_{v,t,l}^{(g)}, \bar{\bar{S}}_{t,l}^{(g)}), \tag{15}$$

$$\text{Joint group effect: } \bar{\boldsymbol{\theta}}_{v,t,l}^{(g)} \sim N_q(\bar{\boldsymbol{m}}_{v,t,l}^{(g)}, \bar{\boldsymbol{S}}_{t}^{(g)}), \tag{16}$$

where

$$\bar{m}_{v,t,l}^{*(g)} = \frac{1}{N_g}\sum_{z=1}^{N_g} m_{v,t,l,1}^{*(z)}, \quad \bar{S}_{v,t,l}^{(g)} = \frac{1}{N_g^2}\sum_{z=1}^{n_g} C_{v,t,l,l}^{(z)} * S_{v,t,1,1}^{(z)},$$

$$\bar{\bar{m}}_{v,t,l}^{(g)} = \frac{1}{N_g}\sum_{z=1}^{n_g} \bar{m}_{v,t,l}^{(z)}, \quad \bar{\bar{S}}_{v,t,l}^{(g)} = \frac{1}{N_g^2}\sum_{z=1}^{N_g} \bar{S}_{v,t,l}^{(z)},$$

$$\bar{\boldsymbol{m}}_{v,t,l}^{(g)} = \frac{1}{N_g}\sum_{z=1}^{N_g} \boldsymbol{m}_{v,t,l}^{(z)}, \quad \text{and } \bar{\boldsymbol{S}}_{v,t}^{(g)} = \frac{1}{N_g^2}\sum_{z=1}^{n_g} C_{v,t,l,l}^{(z)} \boldsymbol{S}_{v,t}^{(z)}.$$

Let us suppose now that there are two groups we want to compare. For the patient group (group $\boldsymbol{A}$), we have N_A individuals and for the control group (group $\boldsymbol{B}$), we have N_B individuals. If, for instance, we decide to compare the joint effect between the two groups, the most obvious way to do so is by computing the distribution for the variable $\bar{\boldsymbol{\theta}}_{v,t,l}^{(AB)} = \bar{\boldsymbol{\theta}}_{v,t,l}^{(A)} - \bar{\boldsymbol{\theta}}_{v,t,l}^{(B)}$, which in this case is given by $N_q(\bar{\boldsymbol{m}}_{v,t,l}^{(A)} - \bar{\boldsymbol{m}}_{v,t,l}^{(B)}, \bar{\boldsymbol{S}}_{t}^{(A)} + \bar{\boldsymbol{S}}_{t}^{(B)})$. The analysis is analogous for the marginal and average effects. As mentioned previously, in this setting, it is only possible to perform any type of inference for a particular fixed time t. In the next section, we present three algorithms that help us to overcome this limitation.

3. Algorithms

In this section we present three different algorithms to draw on-line trajectories of the state parameter. The first two algorithms only depend on the posterior distribution obtained at each time t, whereas the third one is based on the filtered

or smoothing distribution $p(\boldsymbol{\Theta}^{(z)}_{v,t-j}, \boldsymbol{\Sigma}^{(z)}|D^{(z)}_{vt})$, for $1 < j \leq t$. To obtain this distribution, it is necessary to take the current information to be filtered back to previous time points. In other words, this distribution uses all current information available at time point t to smooth the information of the state parameter at the previous time points $t - j$. For more details about the concepts related to filtered or smoothing distributions in the context of DLM, see West and Harrison (1997) and Frühwirth-Schnatter (1994). For the examples and assessment presented in sections 4 and 5, we implement model (1) and these algorithms in the R package mentioned above, with the help of the parallel and RcppArmadillo (Eddelbuettel and Sanderson (2014)) packages to speed up computation time. We also make use of the oro.nifti (Whitcher et al., 2011) and neurobase (Muschelli, 2018) packages to import the fMRI data from and export them to .nii.gz format, and to plot the evidence activation maps, respectively. The FLS software package (Jenkinson et al. (2012)) is used when preprocessing of the fMRI data is needed.

3.1. Forward estimated trajectories sampler or FETS algorithm

Suppose we are interested in detecting brain activation using the distribution (12). For that purpose, we define the on-line estimated trajectory of $\boldsymbol{\theta}^{(z)}_{v,t,l}$ as the vector

$$\gamma^{(z)}_{v,l} = \{E(\boldsymbol{\theta}^{(z)}_{v,t_1,l}|D^{(z)}_{v,t_1}), E(\boldsymbol{\theta}^{(z)}_{v,t_2,l}|D^{(z)}_{v,t_2}), \ldots, E(\boldsymbol{\theta}^{(z)}_{v,T,l}|D^{(z)}_{v,T})\},$$

for $t_1 \geq 30$. In the case of a significant match between the expected and observed BOLD responses, one would expect every component of the vector $\gamma^{(z)}_{v,l}$ to be positive. In this sense, we must compute a measure of evidence for brain activation at voxel v, such as $p(\gamma^{(z)}_{v,l} > 0)$. For instance, if $p(\gamma^{(z)}_{v,l} > 0) > \alpha$, where α is a threshold defined by the user (e.g., $\alpha = 0.95$), then one would conclude that there is significant evidence of a match between the expected and the observed BOLD responses. In other words, there is a brain activation related to the l-th stimulus or task in the cluster related to the voxel v. The functional variable $\gamma^{(z)}_{v,l}$ can be defined in the same way when using either (11) or (13). Thus, taking advantage of the sequential update process related to the posterior distribution (7), we build an algorithm to draw $\gamma^{(z)}_{v,l}$ variables. We call it Forward Estimated Trajectory Sampler (FETS).
For $p(\theta^{(z)}_{v,t,l}|D^{(z)}_{vt})$ being any of (11), (12) or (13) and $\boldsymbol{s}^{(z)}_{v,t}$ a posterior estimate related to $\boldsymbol{\Sigma}_{v,t}$:

Algorithm FETS.

1: **procedure**
2: for $k = 1 \ldots N$,
3: Draw $\theta^{(z,k)}_{v,t,l}$ from $p(\theta^{(z)}_{v,t,l}|D^{(z)}_{v,t})$ for $t = 1, \ldots, T$ and $l = 1, \ldots, p$
4: Draw $\nu^{(z,k)}_{v,t}$ from $N_q[0, \boldsymbol{s}^{(z)}_{\boldsymbol{v},t}]$ for $t = 1, \ldots, T$
5: Compute $\tilde{Y}^{(z,k)}_{v,t} = \sum_{l=1}^{p} \theta^{(z,k)}_{v,t,l} x^{(z)}_l(t) + \nu^{(z,k)}_{v,t}$ for $t = 1, \ldots, T$
6: Compute $p(\theta^{(z,k)}_{v,t,l}|\tilde{D}^{(z)}_{v,t})$ for $t = 1, \ldots, T$, where $\tilde{D}^{(z)}_{v,t} = \{\tilde{Y}^{(z,k)}_{v,1}, \ldots, \tilde{Y}^{(z,k)}_{v,t}\}$
7: Let $\gamma^{(z,k)}_{v,l} = \{E(\theta^{(z,k)}_{v,1,l}|\tilde{D}^{(z)}_{v,1}), \ldots, E(\theta^{(z,k)}_{v,T,l}|\tilde{D}^{(z)}_{v,T})\}$ for $l = 1, \ldots, p$
8: **end procedure**

Then a measure of evidence for the activation of voxel v is computed as

$$p(\gamma^{(z)}_{v,l} > 0) = E(\mathbb{1}_{(\gamma^{(z)}_{v,l} > 0)}) \approx \frac{\sum_{k=1}^{N} \mathbb{1}_{(\gamma^{(z,k)}_{v,l} > 0)}}{N}, \tag{17}$$

where $\mathbb{1}_{(A)}$ is the indicator function related to the event A.
The estimate $\boldsymbol{s}^{(z)}_{v,t}$ depends on the choice of $p(\theta^{(z)}_{v,t,l}|D^{(z)}_{v,t})$. Thus, for any of (11), (12) or (13), $\boldsymbol{s}^{(z)}_{v,t}$ can be replaced by $S^{(z)}_{v,t,1,1}$, $\frac{1}{q^2}\left[\sum_{n=1}^{q} S^{(z)}_{v,t,n,n} + \sum_{n \neq n'} S^{(z)}_{v,t,n,n'}\right]$ or $\mathbf{S}^{(z)}_{v,t}$, respectively. It is also possible to compute a measure of evidence for a contrast between two different tasks or stimuli. For instance, suppose one is interested in the comparison $\gamma^{(z)}_{v,l} > \gamma^{(z)}_{v,l'}$. In other words, one wants to investigate whether the brain activation related to the task l is stronger than that related to task l'. To answer that question, one can compute $p(\gamma^{(z)}_{v,l} - \gamma^{(z)}_{v,l'} > 0)$ taking the algorithm's output in step 6 and using Monte Carlo integration analogously to (17). In addition to being used to compute measures of evidence for brain activation, the **FETS** algorithm also generates other useful outputs such as $\{\tilde{Y}^{(z,k)}_{v,1}, \ldots, \tilde{Y}^{(z,k)}_{v,t}\}$ and $\{\mathbf{S}^{(z,k)}_{v,1}, \ldots, \mathbf{S}^{(z,k)}_{v,t}\}$, which could eventually be used to build model-assessment methods and/or to approach the connectivity problem between different brain regions. These issues are discussed in more detail in the Future Work section. The use of the **FETS** algorithm for the group case is straightforward and requires only the use of any of the distributions (14), (15), or (16) playing the role of $p(\theta^{(z)}_{v,t,l}|D^{(z)}_{v,t})$ and

an appropriate estimate for observational noise variance $\mathbf{\Sigma}_{v,t}^{(z)}$, which in this case can be computed as $\bar{\boldsymbol{S}}_{v,t}^{(g)} = \frac{1}{N_g^2}\sum_{z=1}^{n_g} \boldsymbol{S}_{v,t}^{(z)}$. However, the covariates related to $\mathbf{F}_{v,t}^{\prime(z)}$ must be the same for all the subjects. This is a common feature in many fMRI experiments as we will see in one of the examples presented below. However, some designs, such as event-related designs, occasionally require random sequences of stimuli, which lead to having different $\mathbf{F}_{v,t}^{\prime(z)}$ matrices for each subject in the sample. To overcome this limitation, we propose a simpler second algorithm, which does not depend on $\mathbf{F}_{v,t}^{\prime(z)}$ and can also be used to perform inference about the state parameter.

3.2. FSTS algorithm

The evolution equation in (1) relates to the state parameter at time t with its own value at time $t-1$. We take advantage of that Markovian property, some prior-posterior estimates, and the concept of discount factors to develop a forward sampler algorithm, which we call Forward State Trajectories Sampler (FSTS).

Algorithm Forward state trajectories sampler.

procedure
  for $k = 1 \dots N$
  Compute $\mathbf{W}_{v,t}^{(z)} = \mathbf{B_t}\mathbf{C}_{v,t-1}^{(z)}\mathbf{B_t} - \mathbf{C}_{v,t-1}^{(z)}$ for $t = 1, \dots, T$
  Draw $\mathbf{\Omega}_{v,t}^{(z,k)}$ from $N_{pq}[\mathbf{0}, \mathbf{W}_{v,t}^{(z)}, \boldsymbol{S}_{v,t}^{(z)}]$ for $t = 1, \dots, T$
  Draw $\mathbf{\Theta}_{v,t-1}^{(z,k)}$ from $N_{pq}[\mathbf{m}_{v,t-1}^{(z)}, \mathbf{C}_{v,t-1}^{(z)}, \mathbf{S}_{v,t-1}^{(z)}]$ for $t = 1, \dots, T$
  Compute $\mathbf{\Theta}_{v,t}^{(z,k)} = \mathbf{\Theta}_{v,t-1}^{(z,k)} + \mathbf{\Omega}_{v,t}^{(z,k)}$ for $t = 1, \dots, T$
  Let $\hat{\mathbf{\Theta}}_{v}^{(z,k)} = \{\mathbf{\Theta}_{v,1}^{(z,k)}, \dots, \mathbf{\Theta}_{v,T}^{(z,k)}\}$
end procedure

From the vector $\hat{\mathbf{\Theta}}_{v}^{(z,k)} = \{\mathbf{\Theta}_{v,1}^{(z,k)}, \dots, \mathbf{\Theta}_{v,T}^{(z,k)}\}$, we can define three different types of sub-vectors related to the variables in (8), (9) and (10). For instance, if one is interested in the joint effect (10), then $\hat{\boldsymbol{\theta}}_{v,l}^{(z)} = \{\boldsymbol{\theta}_{v,1,l}^{(z)}, \dots, \boldsymbol{\theta}_{v,T,l}^{(z)}\}$, for $l = 1, \dots, p$. Then for the task l one can compute an activation probability for the voxel v as

$$p(\hat{\boldsymbol{\theta}}_{v,l}^{(z)} > 0) = E(\mathbb{1}_{(\hat{\boldsymbol{\theta}}_{v,l}^{(z)} > 0)}) \approx \frac{\sum_{k=1}^{N} \mathbb{1}_{(\hat{\boldsymbol{\theta}}_{v,l}^{(z,k)} > 0)}}{N}. \tag{18}$$

The probability of activation in (18) can be computed analogously for (8) and (9) just by taking the appropriate components from $\mathbf{\Theta}_{v}^{(z)}$.

3.3. Forward-filtering-backward-sampling algorithm

Following the same idea as in Frühwirth-Schnatter (1994), we now present a matrix-variate version of the forward-filtering-backward-sampling algorithm. It is worth mentioning that in Frühwirth-Schnatter (1994), a data-augmentation approach is used to address problems related to the estimation of the covariance matrix $\mathbf{W}_{v,t}^{(z)}$. Here, instead, we apply a discount-factor approach as it is described above. Then using Bayes's theorem and given some assumptions of conditional independence related to the MVDLM, it can be shown that

$$p(\mathbf{\Theta}_{v,t-j}^{(z)} | \mathbf{\Theta}_{v,t-j+1}^{(z)}, \mathbf{\Sigma}^{(z)}, D_{v,t}^{(z)}) \propto p(\mathbf{\Theta}_{v,t-j}^{(z)} | \mathbf{\Sigma}^{(z)}, D_{v,t-j}^{(z)}) * p(\mathbf{\Theta}_{v,t-j+1}^{(z)} | \mathbf{\Theta}_{v,t-j}^{(z)}, \mathbf{\Sigma}^{(z)}, D_{v,t-j}^{(z)}),$$

which leads to

$$(\mathbf{\Theta}_{v,t-j}^{(z)} | \mathbf{\Theta}_{v,t-j+1}^{(z)}, \mathbf{\Sigma}^{(z)}, D_{v,t}^{(z)}) \sim N_{pq}(\mathbf{m}_{v,j}^{*(z)}, \mathbf{C}_{v,j}^{*(z)}, \mathbf{\Sigma}^{(z)}),$$

where

$$\mathbf{m}_{v,j}^{*(z)} = \mathbf{m}_{v,t-j}^{(z)} + \mathbf{C}_{v,t-j}^{(z)}(\mathbf{B_t}\mathbf{C}_{v,t-j}^{(z)}\mathbf{B_t})^{-1}(\mathbf{\Theta}_{v,t-j+1}^{(z)} - \mathbf{m}_{v,t-j}^{(z)})$$

and

$$\mathbf{C}_{v,j}^{*(z)} = \mathbf{C}_{v,t-j}^{(z)} - \mathbf{C}_{v,t-j}^{(z)}(\mathbf{B_t}\mathbf{C}_{v,t-j}^{(z)}\mathbf{B_t})^{-1}\mathbf{C}_{v,t-j}^{(z)}.$$

A proof of this result is given in Appendix A.

The forward-filtering-backward-sampling algorithm for the matrix-variate case is given by

Algorithm Forward-filtering-backward-sampling.

1: **procedure**
2: for $k = 1 \dots N$
3: Draw $\boldsymbol{\Sigma}^{(z,k)}$ from $W_{n_t}^{-1}(\mathbf{S}_t)$
4: For $j = 0$ draw $\boldsymbol{\Theta}_{v,t}^{(z,k)}$ from $N_{pq}(\mathbf{m}_{v,j}^{(z)}, \mathbf{C}_{v,j}^{(z)}, \boldsymbol{\Sigma}^{(z,k)})$
5: For $1 \leq j < t$ draw $\boldsymbol{\Theta}_{v,t-j}^{(z,k)}$ from $(\boldsymbol{\Theta}_{v,t-j}^{(z)} | \boldsymbol{\Theta}_{v,t-j+1}^{(z,k)}, \boldsymbol{\Sigma}^{(z,k)}, D_{v,t}^{(z)})$
6: Let $\hat{\boldsymbol{\Theta}}_v^{(z,k)} = \{\boldsymbol{\Theta}_{v,1}^{(z,k)}, \dots, \boldsymbol{\Theta}_{v,t}^{(z,k)}\}$
7: **end procedure**

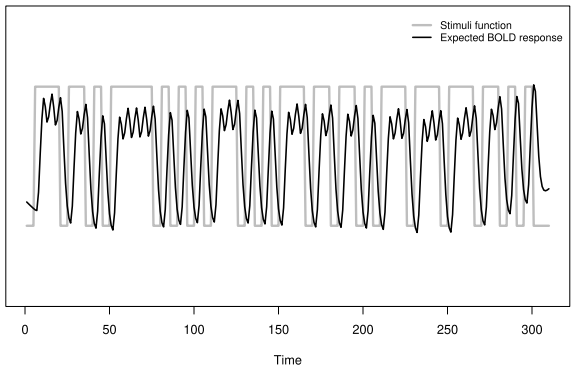
Using Monte Carlo integration as in (18), one can test brain activation by simply taking the appropriate components from $\boldsymbol{\Theta}_v^{(z)}$.

4. Examples

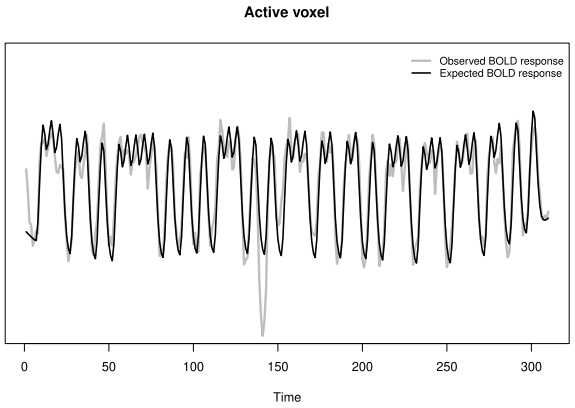
In this section we present two examples of modeling real fMRI data using model (1) and applying the three algorithms presented above in order to detect brain reaction. In the first example, Pernet et al. (2015) perform an experiment in which an auditory stimulus, following a block design, is presented in order to find the locations of "temporal voice areas". In that work, Pernet et al. (2015) model the voice and non-voice sounds separately, whereas here, for the sake of simplicity, we merge both vocal and non-vocal sounds in a single block design as it is shown in Fig. 2. With the aim of showing the performance of the three algorithms proposed in this work, we present the analysis for two single voxels (an active and a non-active voxel) and activation maps for the entire brain. In the second example, Bazán et al. (2015) perform a finger-tapping task experiment, following an event-related design, which should lead to an activation of the motor cortex. In both examples, we run an analysis for one single subject and use the same fixed value for the discount factor $\beta_{it} = 0.95$ and vague priors (zero mean and large variance) for the state parameter $\boldsymbol{\Theta}_{v,0}^{(z)}$, as well as for $\boldsymbol{\Sigma}_v^{(z)}$ with $\boldsymbol{S}_{v0}^{(z)} = \boldsymbol{I}_q$ and $n_0 = 1$. Nevertheless, it is possible to define informative prior distributions by taking advantage of the sequential updating process in the MVDLM. For instance, let us suppose you have an fMRI database for one individual from a pilot sample and fit the model (1) for that data. Then you get the posterior distribution (6) for $t_p = 1, \dots, T_p$ and set the hyperparameters of the new experiment as $\boldsymbol{\Theta}_{v,0}^{(z)} = \boldsymbol{\Theta}_{v,T_p}^{(z)}$, $\boldsymbol{S}_{v,0}^{(z)} = \boldsymbol{S}_{v,T_p}^{(z)}$ and $n_0 = n_p$, where n_p is the number of MRI images in the pilot sample. This way of including prior information in the analysis is actually useful in cases with small samples (or just a few MRI images) where the prior information can play a relevant role in the analysis. However, this is not a common case in fMRI data analysis, where there are usually hundreds of images available, as in the examples presented below.

Auditory cortex activation

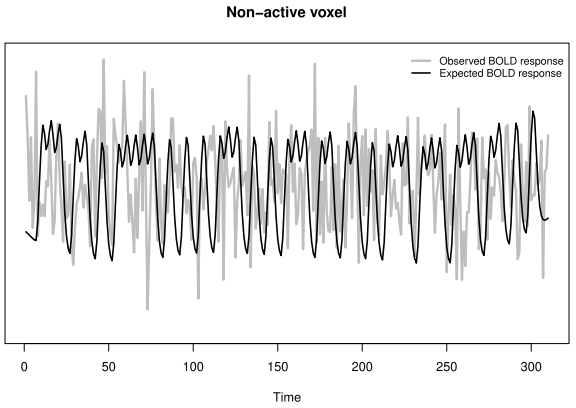
In Figs. 3 and B.9, we can see the results after applying the FETS, FSTS and FFBS algorithms to both active and non-active voxels, which are shown in Fig. 2. Considering that this is an experiment where only sound stimulation is applied, all three algorithms yield the results that are expected for voxels located in the temporal cortex and for voxels outside that region. For the active voxel, the on-line trajectories related to the state parameter lie above zero, which can be interpreted as a match between the expected and observed BOLD responses. For non-active voxels, the simulated on-line trajectories lie around zero, which yields lower evidence of activation (e.g., $<< 0.95$) according to the Monte Carlo evidence defined by (17) and/or (18). We also analyze extreme cases for non-active voxels as presented in Fig. 4, where the density presented in panel (b) represents the posterior distribution for the state parameter at time $t = T$. The inference obtained using only that distribution will yield results quite similar to those obtained when using the GLM (see (West and Harrison, 1997, page 80) and (Quintana, 1987, page 20)). Thus, the evidence of activation when it is computed using (17) correctly identifies that that voxel is non-active ($<< 0.90$), whereas when the last distribution at time $t = T$ is used, the evidence leads to the wrong conclusion, classifying that voxel as active ($>> 0.95$). In Fig. 5, we present the activation maps for a single subject, obtained when applying the FETS, FSTS and FFBS algorithms, respectively. For this example, we consider a voxel as active when its computed evidence obtained under any of the three algorithms is greater than 0.95. We also present a parametric map obtained when using the General Linear Model (with correction for multiple comparisons using Gaussian random-field theory and a cluster significance threshold equal to 0.01), which can be considered the most common approach employed by users of statistical techniques for task-based fMRI data analysis. From Figs. 3, 4, and 5, we can see the performance of our method in this example, with each of the three algorithms detecting the parietal activation due to sound stimulation. From our knowledge and after evaluating these images with some colleagues from the field of radiology, we conclude that the FETS algorithm yields better results in terms of a wider parietal activation and fewer ("false") activations from outside the temporal cortex. FFBS also yields good activation maps, but with more activations from outside the temporal cortex. FSTS also performs well, but has been shown to be conservative, with almost no possibly false activations detected and with a smaller parietal activation. For this example, only the average cluster effect (ACE) distribution (9) is employed as the sampling distribution for the state parameter. However, the results are similar when using either the marginal (8) or the joint (10) effect distribution.



(a)

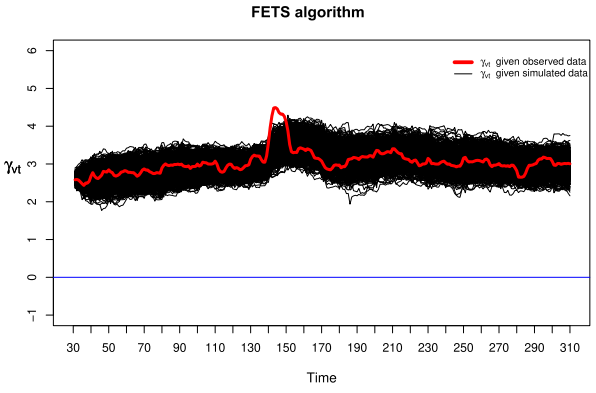


(b)

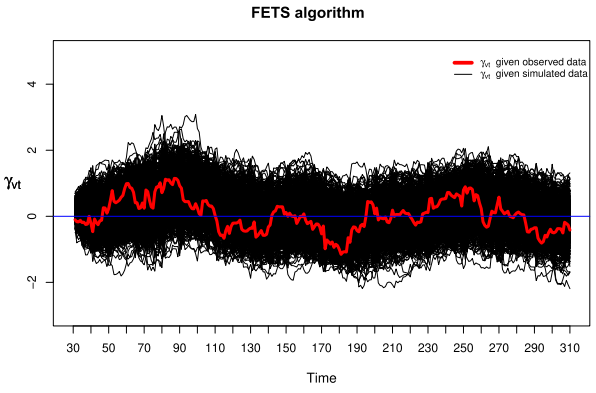


(c)

Fig. 2. The top Figure shows the block-design or stimulus function for voice and non-voice sounds and its expected BOLD response. The images on the lower left and right show the contrast (or possible match) between the expected and observed BOLD responses for both an active and non-active voxel, respectively, on a single subject.



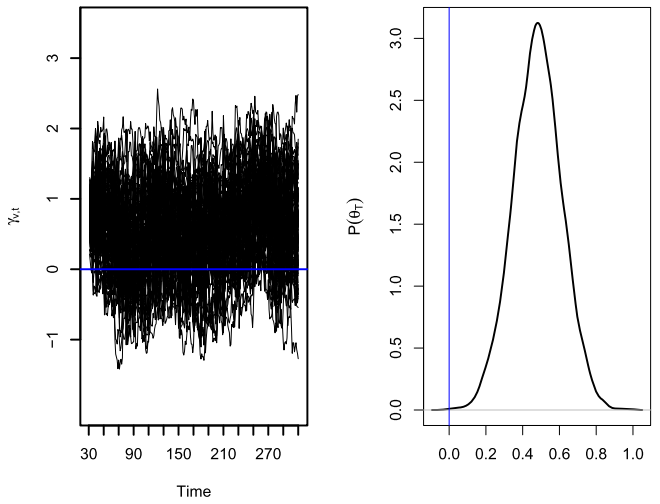
(a)



(b)

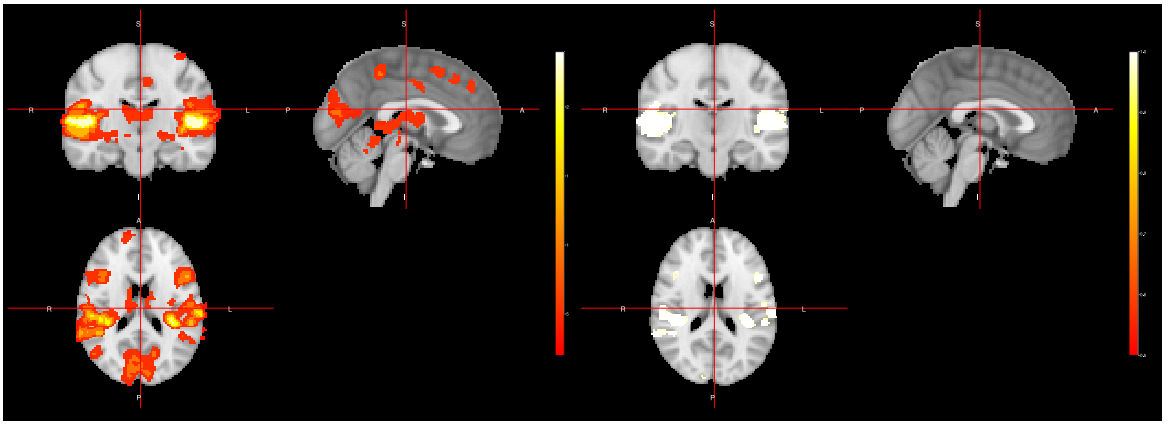
Fig. 3. Simulated on-line trajectories obtained when FETS algorithm is applied to both active (a) and non-active (b) voxels presented in Fig. 2.

Fig. 6 shows the activation maps for a group of 21 subjects when using the FETS and FFBS algorithms. In this group scenario, the stimulated regions are much better identified, as is expected in a group analysis. The activation map obtained under the FSTS algorithm is similar to that obtained when using the FETS algorithm. Despite both the FETS and FFBS algorithms successfully identify the activated regions, there is an interesting aspect to mention from this particular example. It can be noticed that the FETS algorithm presents no activated voxels outside the stimulated region, i.e., an activation map with a zero rate of false-positive activations. However, this comes at a price because the stimulated region detected when

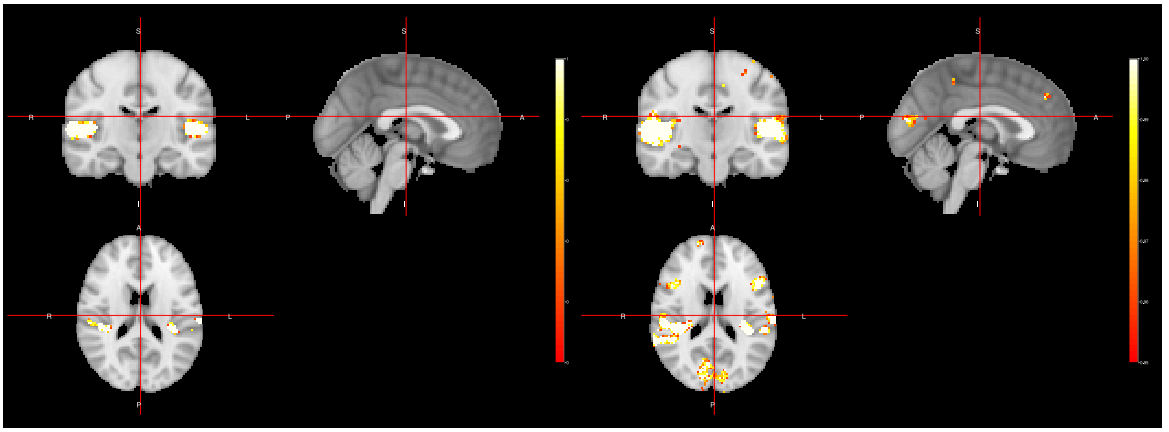


(a) (b)

Fig. 4. Simulated on-line trajectories obtained when the FETS algorithm is applied to a non-active voxel (a) and posterior distribution for the state parameter at time $t = T$ (b). Results are similar for FSTS and FFBS.

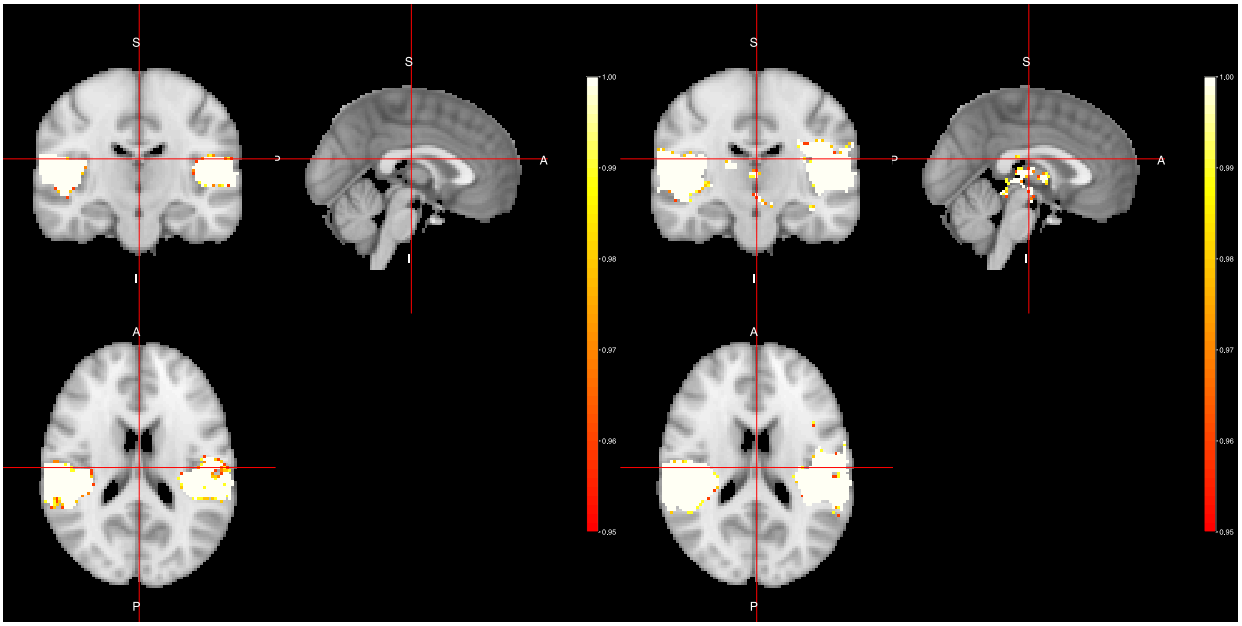


(a) GLM (b) FETS



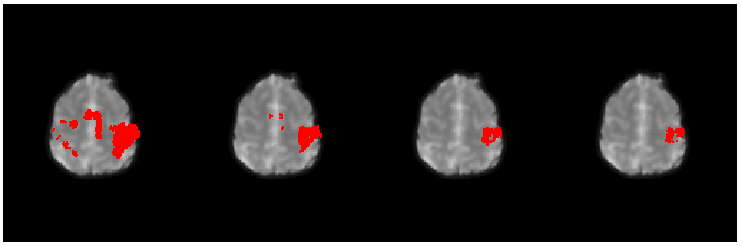
(c) FSTS (d) FFBS

Fig. 5. Voxel-wise analysis for one single subject. The top left figure shows a parametric map obtained via General Linear Model (GLM). The top right and lower left and right figures show activation maps related to each of the FETS, FSTS and FFBS algorithms respectively. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)



(a) FETS (b) FFBS

Fig. 6. Activation maps of 95% for a group of 21 subjects using the FETS and FFBS algorithms under the joint group effect distribution (16).



(a) $\beta = 0.95$ (b) $\beta = 0.90$ (c) $\beta = 0.87$ (d) $\beta = 0.85$

Fig. 7. Activation maps of 95% obtained using the FETS-ACE algorithm for different values of the discount factor β.

using the FETS is smaller than the region indicated as active when using the FFBS algorithm. This could mean that the mechanism that allows the FETS algorithm to discard false-positive activations sharply would also lead to discarding actual activations with a weak signal. This and other aspects to explore about the group stage are left for future work.

Motor cortex activation

In Fig. 7, we can see an activation map corresponding to the event-related finger-tapping experiment from Bazán et al. (2015), obtained using the FETS algorithm. When the FSTS and FFBS algorithms are applied, they yield similar results, successfully identifying the motor cortex activation expected from this experiment. What we aim to highlight with this example is the key role played by the discount parameter β. In the modeling we are proposing here, we leave the value of β as a free input value defined by the user. However, we can see that the activation map is sensitive to the value set for β. From our empirical experience, we recommend values $0.8 < \beta < 1$. However, one possible way to overcome this limitation is to run several maps for different β values and use a statistical criterion (e.g., Bayes factor) to select the "appropriate" map. In the final section, we discuss this issue in more detail.

5. Assessment of the method

In order to assess the activation maps obtained when the model (1) is fitted and any of the algorithms proposed in this work is used to detect voxel activation, we employ two different approaches. For the first type of assessment, we resort to simulated data using the R package **neuRosim** (Welvaert et al. (2011)). The second assessment approach is inspired by the work of Eklund et al. (2016), where real data from resting-state experiments are used in order to evaluate the rate of false

Table 1
Percentage of false-positive activations for the fictitious block paradigms B1 and B2.

	B1								
	FETS			FSTS			FFBS		
Source fMRI data	Marginal	LTT	Joint	Marginal	LTT	Joint	Marginal	LTT	Joint
Beijing (198 subjects)	0.00	0.00	0.00	0.00	0.00	0.00	0.25	0.41	0.00
Cambridge (198 subjects)	0.01	0.01	0.00	0.00	0.00	0.00	0.39	0.54	0.00
Oulu (103 subjects)	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.03	0.00
	B2								
	FETS			FSTS			FFBS		
Source fMRI data	Marginal	LTT	Joint	Marginal	LTT	Joint	Marginal	LTT	Joint
Beijing (198 subjects)	0.00	0.00	0.00	0.00	0.00	0.00	0.70	1.20	0.01
Cambridge (198 subjects)	0.14	0.21	0.00	0.00	0.00	0.00	1.00	1.40	0.02
Oulu (103 subjects)	0.00	0.00	0.00	0.00	0.00	0.00	1.20	2.10	0.04

positive activations in the entire brain. These two assessments aim to evaluate the performance of the method proposed in this article (in terms of false and true activations) rather than compare it with other alternative methods or results from other studies.

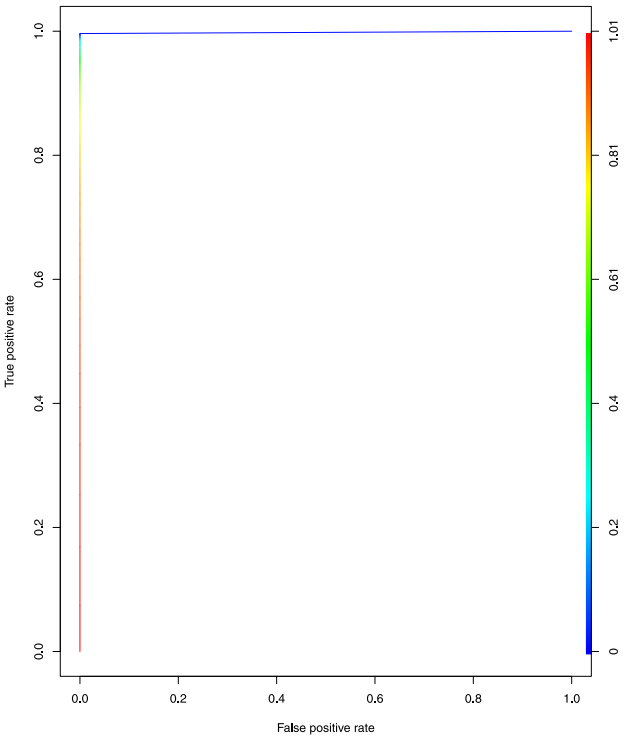
5.1. Using simulated data

The aim of using simulated data is to evaluate the capacity of our method to detect true voxel activations, which in this case are artificially created using the R package **neuRosim**. Specifically, we create two spherical-shape activations in the visual cortex related to a block-design experiment using the function `simVOLfmri` (see Fig. B.11 in Appendix B). In Welvaert et al. (2011), the interested reader can find a detailed explanation on how to simulate fMRI data using the **neuRosim** package. In this assessment, we set two different radius sizes (4 and 7) and the same effect size (250) over the two activated regions for three different signal-to-noise ratios going from weaker to stronger signals (SNR=3.7, 18, 30). In Fig. 8, the ROC graphs are shown for the scenarios studied in this evaluation when using the algorithm FSTS. It can be seen that the FSTS algorithm performs well in all the scenarios in terms of false-positive activations, but for the case of low SNR it would require lower threshold values than the common 0.95 value in order to identify significant patterns of true activations. For the case of FFBS (see Fig. B.10 in Appendix B), outstanding performance can be observed even for lower SNR values. The performance of the FETS algorithm over all the scenarios is quite similar to that observed with the FFBS algorithm. It is worth mentioning that from fMRI real-data problems it is not possible to achieve results such as those presented in Figs. 8 and B.10, because, as stated by Eklund et al. (2016), it is obviously very hard to simulate the complex spatiotemporal noise that arises from a living human subject in an MR scanner. However, this type of simulation study can give us some insight about the capability of the method related to correct identification of true voxel activations.

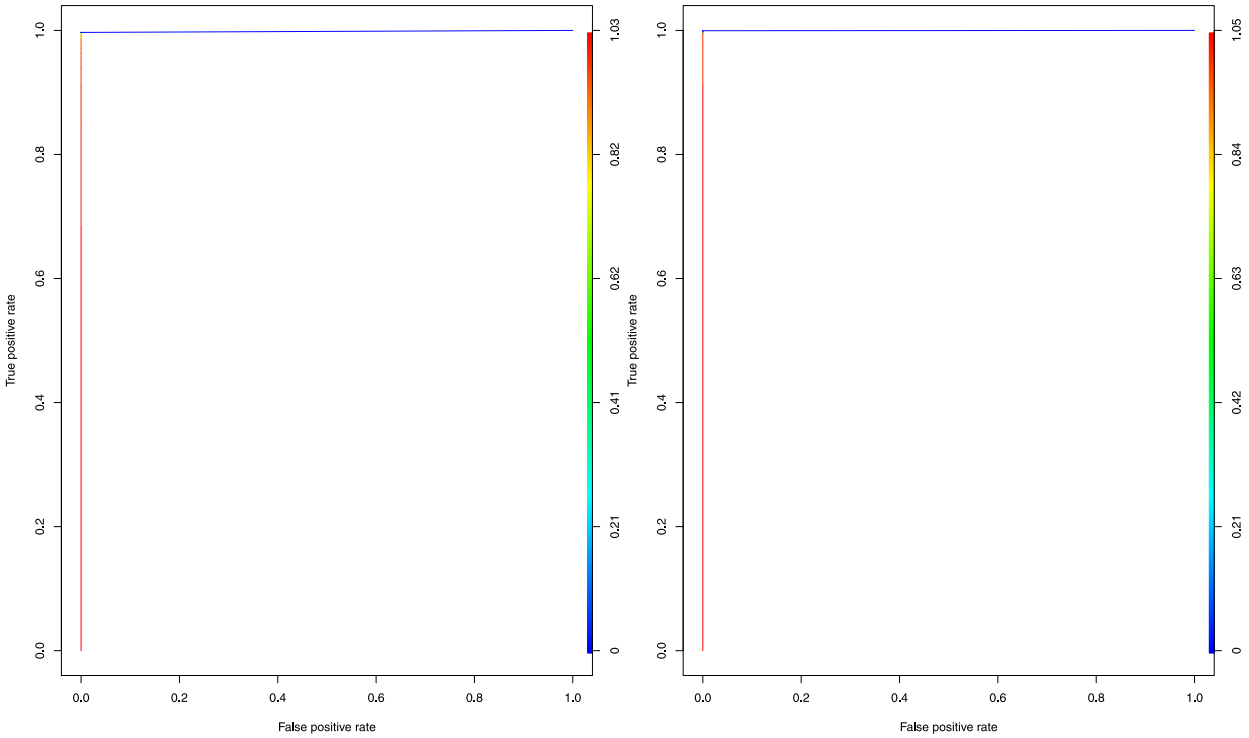
5.2. Using real data

One of the big challenges for statistical techniques for fMRI data analysis is controlling the rate of false-positive activations that usually appear on activation maps. For instance, in the case of the GLM, even using some sort of corrections for multiple comparisons, the rate of false positives can be as high as 70% (Eklund et al. (2012) and Eklund et al. (2016)). Thus, in order to assess the capacity of the method proposed in this work for controlling false activations, we follow the same approach as Eklund et al. (2016). Although this approach, like any other method intended to evaluate the performance of tools for task-based fMRI data analysis, is not without criticism (Slotnick, 2017), it can be considered a reliable method for this purpose (Nichols et al., 2017). The approach consists of the use of real fMRI data, related to resting-state experiments obtained from the 1000 Functional Connectomes Project (Biswal et al., 2010), which are correlated with covariates from fictitious stimuli. Hence, all the significant activations detected are interpreted as false-positive activations. This is a reasonable assumption in the sense that if one models the observed BOLD signal ($Y_{v,t}^{(z)}$) from a resting-state experiment as $E(Y_{v,t}^{(z)}) = \mathbf{F}'_{v,t}\boldsymbol{\Theta}_{v,t}$, where $\mathbf{F}'_{v,t}$ is an (fictitious) expected BOLD signal not related to $Y_{v,t}^{(z)}$, what one would expect to find is $E(\boldsymbol{\Theta}_{v,t}|D_{v,t})) \leq 0$, which can be interpreted as no existing match between the expected and observed BOLD signals. Therefore, we create four fictitious variables related to two block paradigms (B1 (10 s on off), B2 (30 s on off)) and to two event-related paradigms (E1 (2 s activation, 6s rest), E2(1 to 4 s activation, 3 to 6 s rest, randomized)). We correlate those variables with the fMRI resting data from three different labs (Beijing(198 subjects), Cambridge (198 subjects), Oulu (103 subjects)), and count the number of (false) significant activations. For the method proposed in this article, this assessment is only performed at the individual level to assess its own performance rather than compare it to other models or methods.

In Table 1, we show the assessment results for the variables B1 and B2. E1 and E2 yield similar outcomes. Those results are congruent with what is observed in the assessment with the simulated-data and real-data examples shown above. The three algorithms present a low rate of false-positive activations, but with less-conservative behavior in the case of FETS and FFBS allowing for the detection of weaker BOLD signals. From these assessments using real and simulated data, the examples



(a) SNR = 3.7



(b) SNR = 18

(c) SNR = 30

Fig. 8. Receiver operating characteristics (ROC) graphs obtained under different SNR when using the FSTS algorithm. The intensity bar at the right of the graphs represents the different values for the threshold.

presented above, and other analyses that it is not possible to present here, we would recommend FETS as the best option among the three algorithms, not only because of its good performance in terms of weak BOLD signals and false-positive activations, but also because its time of execution is 50% faster in comparison with the FFBS and FSTS algorithms.

6. Concluding remarks

In this work we develop an effective and reliable method for statistical analysis of task-based fMRI data, taking into account its temporal and (at least locally) spatial features, without the need for the more traditional and complex spatial-temporal models. We employ an MVDLM, which has shown to be a very flexible type of modeling because of the availability of closed-form posterior distributions, discount factors to deal with the variability of the state parameter and the possibility of combining the posterior distributions across multiple subjects to perform group analysis. Another important aspect of the MVDLM is its relationship with the traditional and broadly known GLM ((West and Harrison, 1997, page 80), (Quintana, 1987, page 20)), which can help non-statistician practitioners to become more familiar with this type of modeling. We also make an important contribution, not just in the context of fMRI modeling, but a general way to perform an assessment of dynamic effects on dynamic linear models. To our knowledge, this is something that has been neglected, as mentioned above, possibly because of the type of applications and motivations behind MVDLM. In this work, instead of making predictions or analyzing covariance structures among time series, our aim is to assess effect size related to neural activity. Thus in order to achieve that using this type of modeling, we propose two new algorithms (FETS, FSTS) and adapt an existing one (FFBS) to draw on-line trajectories related to the state parameter. Throughout the examples and assessments presented above, one can see that the performance of our method is remarkable, in the sense that it is able to both properly identify brain activation under different experimentation conditions and control the rate of false activations. Currently, we are working on the implementation of an R package for this method, which will be freely available for interested users from the fMRI community. The first version of this package will offer CPU-parallelized versions of its analysis to speed up the computation time in voxel-wise analysis. And it will add GPU-parallelized implementations for faster analysis in future versions.

It is worth mentioning that with this work we leave some open problems. One of them is the choice of β. In dealing with fMRI data from different types of experiments, we have found that setting values $0.8 < \beta < 1$ can work well. But instead of setting or tuning a particular value for β, a model selection approach for different β values could be easily implemented in order to select the "best" β value according to some statistical criterion such as Bayes factors. Another open problem is related to the distance parameter r, for which we believe a sensitivity analysis could be performed in order to have a better sense of the behavior of this parameter. In all of the examples and analysis performed in this work, we set $r = 1$. Additionally, in the case of the FETS algorithm, there are some outputs, such as the sequence of covariance matrices $\{\mathbf{S}_{v,1}^{(z,k)}, \ldots, \mathbf{S}_{v,t}^{(z,k)}\}$, that could be used to build some sort of dynamic graphs to tackle the problem of connectivity among brain regions for resting-state fMRI data (see Shen et al. (2010) for a better understanding of this type of analysis). Regarding the second stage in task-based fMRI studies, all the analyses presented in this work will be extended to group analyses following the ideas presented in section 2.2. Finally, we want to highlight that in this work only a tiny part of the rich possibilities offered by the general theory of the DLM's is explored. Our ideas could be extended to more complex model specifications or even to the case of exponential-family dynamic models and nonlinear dynamic models.

Acknowledgements

This paper and the research behind it would not have been possible without support from CAPES-Brazil - Finance Code 001 and Minciencias-Colombia - Finance Code 728. We thank Mark Andrew Gannon for the careful reading of the manuscript and his insightful suggestions that helped improve this article. Also, we thank Paulo Bazán for allows us to use the fMRI data related to the finger-tapping example. Finally, we would like to thank the network eScience from the Institute of Mathematics and Statistics at the University of Saõ Paulo to access their servers and laboratories where we have developed most of this project.

Appendix A

In this section, it is shown in more detail how to obtain the conditional distribution $p(\boldsymbol{\Theta}_{v,t-j}^{(z)}|\boldsymbol{\Theta}_{v,t-j+1}^{(z)}, \boldsymbol{\Sigma}^{(z)}, D_{v,t}^{(z)})$, which is necessary to run the matrix-variate version of the forward-filtering-backward-sampling algorithm. Using Bayes's theorem, that conditional distribution is defined as

$$p(\boldsymbol{\Theta}_{v,t-j}^{(z)}|\boldsymbol{\Theta}_{v,t-j+1}^{(z)}, \boldsymbol{\Sigma}^{(z)}, D_{v,t}^{(z)}) = \frac{p(\boldsymbol{\Theta}_{v,t-j}^{(z)}|\boldsymbol{\Theta}_{v,t-j+1}^{(z)}, \boldsymbol{\Sigma}^{(z)}, D_{v,t-j}^{(z)})p(\mathbf{Y}|\boldsymbol{\Theta}_{v,t-j}^{(z)}, \boldsymbol{\Theta}_{v,t-j+1}^{(z)}, \boldsymbol{\Sigma}^{(z)}, D_{v,t-j}^{(z)})}{p(\mathbf{Y}|\boldsymbol{\Theta}_{v,t-j+1}^{(z)}, \boldsymbol{\Sigma}^{(z)}, D_{v,t-j}^{(z)})},$$

where $D_{v,t}^{(z)} = \{D_{v,t-j}, \mathbf{Y}\}$. And given the conditional independence structure assumed in DLMs, $\mathbf{Y}$ is independent of the previous $\boldsymbol{\Theta}_{v,t-j}^{(z)}$, so the terms $p(\mathbf{Y}|\cdot)$ cancel. Then, applying Bayes's theorem to the remaining term

$$p(\boldsymbol{\Theta}_{v,t-j}^{(z)}|\boldsymbol{\Theta}_{v,t-j+1}^{(z)}, \boldsymbol{\Sigma}^{(z)}, D_{v,t}^{(z)}) \propto p(\boldsymbol{\Theta}_{v,t-j}^{(z)}|\boldsymbol{\Sigma}^{(z)}, D_{v,t-j}^{(z)})p(\boldsymbol{\Theta}_{v,t-j+1}^{(z)}|\boldsymbol{\Theta}_{v,t-j}^{(z)}, \boldsymbol{\Sigma}^{(z)}, D_{v,t-j}^{(z)}),$$

and assuming that $\boldsymbol{\Sigma}^{(z)}$ is given, it follows that $(\boldsymbol{\Theta}^{(z)}_{v,t-j}|\boldsymbol{\Sigma}^{(z)}, D^{(z)}_{v,t-j}) \sim N_{pq}[\mathbf{m}^{(z)}_{v,t-j}, \mathbf{C}^{(z)}_{v,t-j}, \boldsymbol{\Sigma}^{(z)}]$ and $(\boldsymbol{\Theta}^{(z)}_{v,t-j+1}|\boldsymbol{\Theta}^{(z)}_{v,t-j}, \boldsymbol{\Sigma}^{(z)}, D^{(z)}_{v,t-j}) \sim N_{pq}\left[\boldsymbol{\Theta}^{(z)}_{v,t-j}, \mathbf{W}^{(z)}_{v,t-j+1}, \boldsymbol{\Sigma}^{(z)}\right]$. Now, from the evolution equation (taking $\mathbf{G}^{(z)}_{v,t-j+1} = \mathbf{I}$) and assumptions in the model (1),

$$\begin{pmatrix} \boldsymbol{\Theta}^{(z)}_{v,t-j} \\ \boldsymbol{\Theta}^{(z)}_{v,t-j+1} \end{pmatrix} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{I} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \boldsymbol{\Theta}^{(z)}_{v,t-j} \\ \boldsymbol{\Omega}^{(z)}_{v,t-j+1} \end{pmatrix}, \tag{A.1}$$

$$\begin{pmatrix} \boldsymbol{\Theta}^{(z)}_{v,t-j} \\ \boldsymbol{\Omega}^{(z)}_{v,t-j+1} \end{pmatrix} \sim N\left(\begin{pmatrix} \mathbf{m}^{(z)}_{v,t-j} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{C}^{(z)}_{v,t-j} & \mathbf{0} \\ \mathbf{0} & \mathbf{W}^{(z)}_{v,t-j+1} \end{pmatrix}, \boldsymbol{\Sigma}^{(z)} \right). \tag{A.2}$$

From the linear transformation (A.1) and distribution (A.2) it follows that

$$\begin{pmatrix} \boldsymbol{\Theta}^{(z)}_{v,t-j} \\ \boldsymbol{\Theta}^{(z)}_{v,t-j+1} \end{pmatrix} \sim N\left(\begin{pmatrix} \mathbf{m}^{(z)}_{v,t-j} \\ \mathbf{m}^{(z)}_{v,t-j} \end{pmatrix}, \begin{pmatrix} \mathbf{C}^{(z)}_{v,t-j} & \mathbf{C}^{(z)}_{v,t-j} \\ \mathbf{C}^{(z)}_{v,t-j} & \mathbf{C}^{(z)}_{v,t-j} + \mathbf{W}^{(z)}_{v,t-j+1} \end{pmatrix}, \boldsymbol{\Sigma}^{(z)} \right). \tag{A.3}$$

Now, from the properties of matrix-variate normal distributions (Dawid, 1981; Quintana, 1987) and applying discount factors, it follows that

$$(\boldsymbol{\Theta}^{(z)}_{v,t-j}|\boldsymbol{\Theta}^{(z)}_{v,t-j+1}, \boldsymbol{\Sigma}^{(z)}, D^{(z)}_{v,t}) \sim N_{pq}(\mathbf{m}^{*(z)}_{v,j}, \mathbf{C}^{*(z)}_{v,j}, \boldsymbol{\Sigma}^{(z)}),$$

where $\mathbf{m}^{*(z)}_{v,j} = \mathbf{m}^{(z)}_{v,t-j} + \mathbf{C}^{(z)}_{v,t-j}(\mathbf{B}_t\mathbf{C}^{(z)}_{v,t-j}\mathbf{B}_t)^{-1}(\boldsymbol{\Theta}^{(z)}_{v,t-j+1} - \mathbf{m}^{(z)}_{v,t-j})$ and $\mathbf{C}^{*(z)}_{v,j} = \mathbf{C}_{v,t-j} - \mathbf{C}_{v,t-j}(\mathbf{B}_t\mathbf{C}^{(z)}_{v,t-j}\mathbf{B}_t)^{-1}\mathbf{C}_{v,t-j}$.

Appendix B

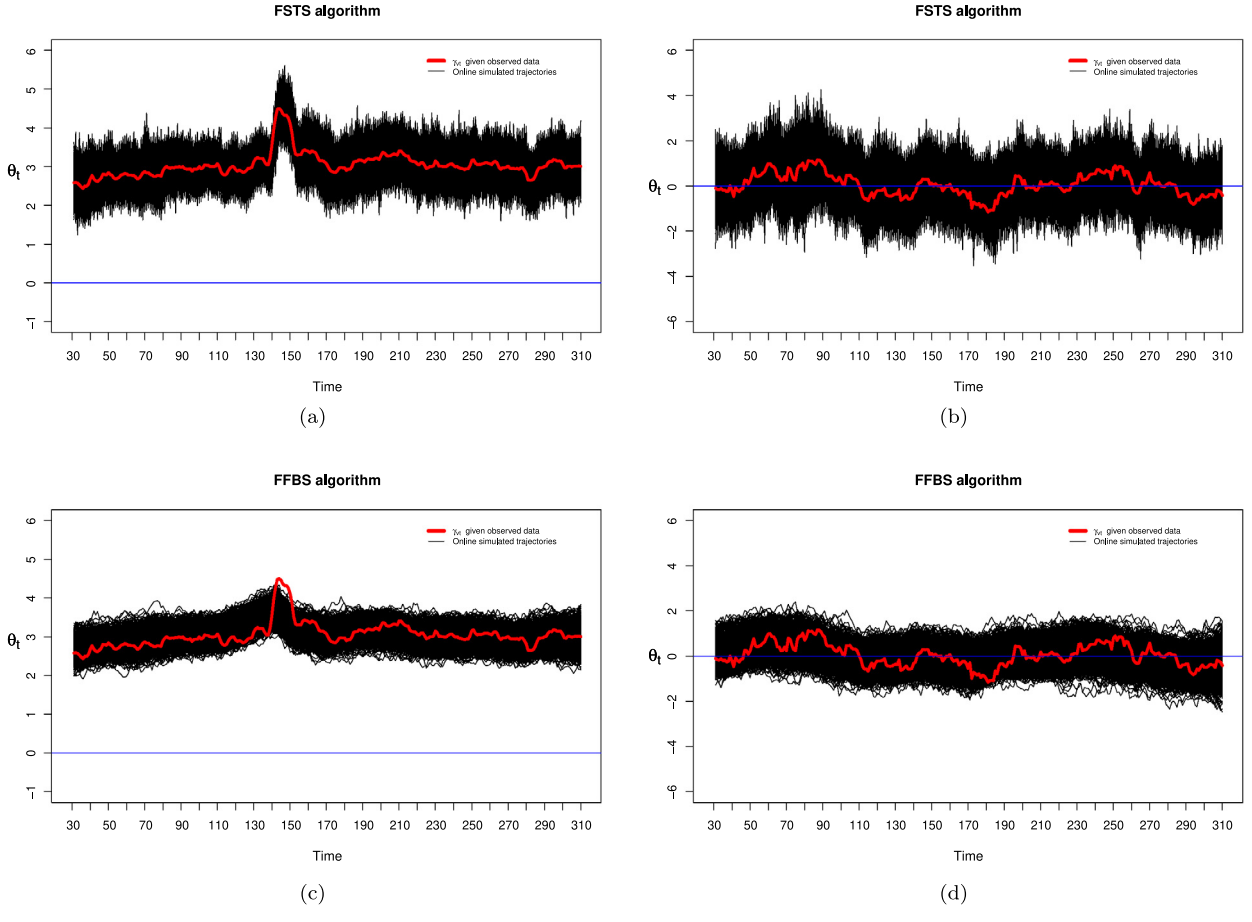
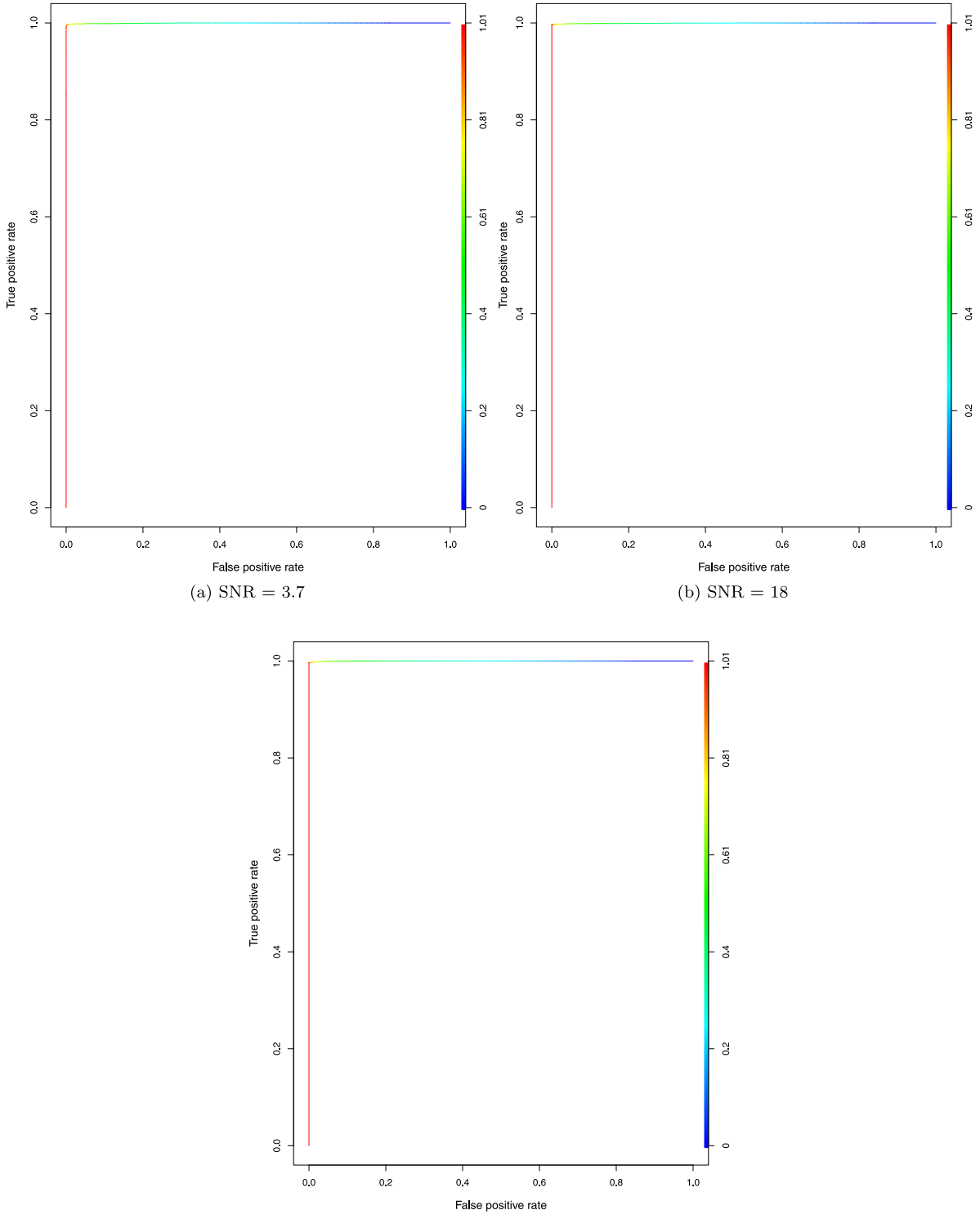


Fig. B.9. Simulated on-line trajectories obtained when FSTS, and FFBS algorithms are applied to both active ((a), (c)) and non-active ((b), (d)) voxels presented in Fig. 2.



(a) SNR = 3.7

(b) SNR = 18

(c) SNR = 30

Fig. B.10. Receiver operating characteristics (ROC) graphs obtained under different SNR when using the FFBS algorithm. The intensity bar at the right of the graphs represents the different values for the threshold.

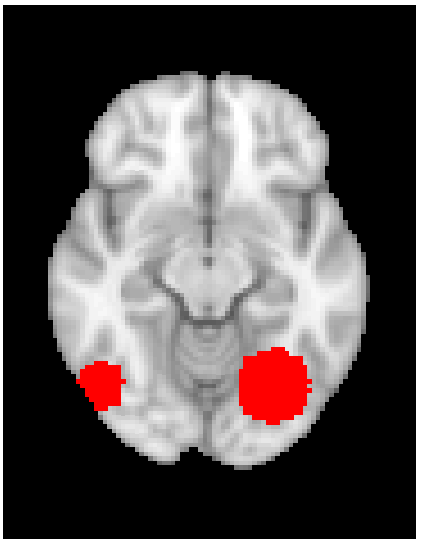


Fig. B.11. Spherical active regions simulated for the analysis of true activations.

References

Aguilar, O., West, M., 2000. Bayesian dynamic factor models and portfolio allocation. J. Bus. Econ. Stat. 18, 338–357.

Ameen, J., Harrison, P., 1985. Normal discount Bayesian models. In: Bayesian Statistics 2: Proceedings of the Second Valencia International Meeting. September 6/10, 1985, vol. 2. North Holland, pp. 271–298.

Bazán, P.R., Biazoli Jr., C.E., Sato, J.R., Amaro Jr., E., 2015. Motor readiness increases brain connectivity between default-mode network and motor cortex: impact on sampling resting periods from fMRI event-related studies. Brain Connect. 5, 631–640.

Beckmann, C.F., Jenkinson, M., Smith, S.M., 2003. General multilevel linear modeling for group analysis in fMRI. NeuroImage 20, 1052–1063.

Bezener, M., Hughes, J., Jones, G., et al., 2018. Bayesian spatiotemporal modeling using hierarchical spatial priors, with applications to functional magnetic resonance imaging. Bayesian Anal.

Biswal, B.B., Mennes, M., Zuo, X.-N., Gohel, S., Kelly, C., Smith, S.M., Beckmann, C.F., Adelstein, J.S., Buckner, R.L., Colcombe, S., et al., 2010. Toward discovery science of human brain function. Proc. Indian Natl. Sci. 107, 4734–4739.

Bowman, F.D., 2014. Brain imaging analysis. Annu. Rev. Stat. Appl. 1, 61–85.

Dawid, A.P., 1981. Some matrix-variate distribution theory: notational considerations and a Bayesian application. Biometrika 68, 265–274.

Eddelbuettel, D., Sanderson, C., 2014. Rcpparmadillo: accelerating R with high-performance C++ linear algebra. Comput. Stat. Data Anal. 71, 1054–1063. https://doi.org/10.1016/j.csda.2013.02.005.

Eklund, A., Andersson, M., Josephson, C., Johannesson, M., Knutsson, H., 2012. Does parametric fMRI analysis with spm yield valid results?—an empirical study of 1484 rest datasets. NeuroImage 61, 565–578.

Eklund, A., Lindquist, M.A., Villani, M., 2017. A Bayesian heteroscedastic glm with application to fMRI data with motion spikes. NeuroImage 155, 354–369.

Eklund, A., Nichols, T.E., Knutsson, H., 2016. Cluster failure: why fMRI inferences for spatial extent have inflated false-positive rates. Proc. Indian Natl. Sci. 113, 7900–7905.

Fei, X., Lu, C.-C., Liu, K., 2011. A Bayesian dynamic linear model approach for real-time short-term freeway travel time prediction. Transp. Res., Part C, Emerg. Technol. 19, 1306–1318.

Friston, K., Penny, W., 2003. Posterior probability maps and spms. NeuroImage 19, 1240–1249.

Friston, K.J., Holmes, A.P., Worsley, K.J., Poline, J.-P., Frith, C.D., Frackowiak, R.S., 1994. Statistical parametric maps in functional imaging: a general linear approach. Hum. Brain Mapp. 2, 189–210.

Frühwirth-Schnatter, S., 1994. Data augmentation and dynamic linear models. J. Time Ser. Anal. 15, 183–202.

Gamerman, D., 1998. Markov chain Monte Carlo for dynamic generalised linear models. Biometrika 85, 215–227.

Ho, M.-H.R., Ombao, H., Shumway, R., 2005. A state-space approach to modelling brain dynamics. Stat. Sin., 407–425.

Jenkinson, M., Beckmann, C.F., Behrens, T.E., Woolrich, M.W., Smith, S.M., 2012. FSL. NeuroImage 62, 782–790.

Kashou, N.H., 2014. A practical guide to an fMRI experiment. In: Advanced Brain Neuroimaging Topics in Health and Disease-Methods and Applications, vol. 4.

Muschelli, J., 2018. Neurobase: 'neuroconductor' base package with helper functions for 'nifti' objects. https://CRAN.R-project.org/package=neurobase. r package version 1.27.6.

Nichols, T.E., Eklund, A., Knutsson, H., 2017. A defense of using resting-state fMRI as null data for estimating false positive rates. Cogn. Neurosci. 8, 144–149.

Penny, W.D., Friston, K.J., Ashburner, J.T., Kiebel, S.J., Nichols, T.E., 2011. Statistical Parametric Mapping: The Analysis of Functional Brain Images. Elsevier.

Pernet, C.R., McAleer, P., Latinus, M., Gorgolewski, K.J., Charest, I., Bestelmeyer, P.E., Watson, R.H., Fleming, D., Crabbe, F., Valdes-Sosa, M., et al., 2015. The human voice areas: spatial organization and inter-individual variability in temporal and extra-temporal cortices. NeuroImage 119, 164–174.

Poldrack, R.A., Mumford, J.A., Nichols, T.E., 2011. Handbook of Functional MRI Data Analysis. Cambridge University Press.

Quintana, J.M., 1985. A Dynamic Linear Matrix–Variate Regression Model.

Quintana, J.M., 1987. Multivariate Bayesian forecasting models. Ph.D. thesis. University of Warwick.

Quintana, J.M., West, M., 1987. An analysis of international exchange rates using multivariate dlm's. Statistician, 275–281.

R Core Team, 2018. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Shen, X., Papademetris, X., Constable, R.T., 2010. Graph-theory based parcellation of functional subunits in the brain from resting-state fMRI data. NeuroImage 50, 1027–1035.

Slotnick, S.D., 2017. Resting-state fMRI data reflects default network activity rather than null data: a defense of commonly employed methods to correct for multiple comparisons. Cogn. Neurosci. 8, 141–143.
Wegmann, B., Eklund, A., Villani, M., 2017. Bayesian Rician regression for neuroimaging. Front. Neurosci. 11, 586.
Welvaert, M., Durnez, J., Moerkerke, B., Verdoolaege, G., Rosseel, Y., 2011. neuRosim: an R package for generating fMRI data. J. Stat. Softw. 44, 1–18.
West, M., Harrison, J., 1997. Bayesian Forecasting and Dynamic Models, 2nd ed. Springer-Verlag New York, Inc., New York, NY, USA.
West, M., Harrison, P.J., Migon, H.S., 1985. Dynamic generalized linear models and Bayesian forecasting. J. Am. Stat. Assoc. 80, 73–83.
Whitcher, B., Schmid, V.J., Thornton, A., 2011. Working with the DICOM and NIfTI data standards in R. J. Stat. Softw. 44, 1–28. http://www.jstatsoft.org/v44/i06/.
Yu, C.-H., Prado, R., Ombao, H., Rowe, D., 2018. A Bayesian variable selection approach yields improved detection of brain activation from complex-valued fMRI. J. Am. Stat. Assoc. 113, 1395–1410.
Zhang, L., Guindani, M., Vannucci, M., 2015. Bayesian models for functional magnetic resonance imaging data analysis. Wiley Interdiscip. Rev.: Comput. Stat. 7, 21–41.
Zhang, L., Guindani, M., Versace, F., Engelmann, J.M., Vannucci, M., et al., 2016. A spatiotemporal nonparametric Bayesian model of multi-subject fMRI data. Ann. Appl. Stat. 10, 638–666.