

## Review Article

# Bayesian Inference for Functional Dynamics Exploring in fMRI Data

Xuan Guo,<sup>1</sup> Bing Liu,<sup>2</sup> Le Chen,<sup>2</sup> Guantao Chen,<sup>2</sup> Yi Pan,<sup>1,3</sup> and Jing Zhang<sup>2</sup>

<sup>1</sup>Department of Computer Science, Georgia State University, Atlanta, GA 30303, USA

<sup>2</sup>Department of Mathematics and Statistics, Georgia State University, Atlanta, GA 30303, USA

<sup>3</sup>Department of Biology, Georgia State University, Atlanta, GA 30303, USA

Correspondence should be addressed to Yi Pan; [yipan@gsu.edu](mailto:yipan@gsu.edu) and Jing Zhang; [jzhang47@gsu.edu](mailto:jzhang47@gsu.edu)

Received 6 December 2015; Accepted 1 February 2016

Academic Editor: Sung-Phil Kim

Copyright © 2016 Xuan Guo et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper aims to review state-of-the-art Bayesian-inference-based methods applied to functional magnetic resonance imaging (fMRI) data. Particularly, we focus on one specific long-standing challenge in the computational modeling of fMRI datasets: how to effectively explore typical functional interactions from fMRI time series and the corresponding boundaries of temporal segments. Bayesian inference is a method of statistical inference which has been shown to be a powerful tool to encode dependence relationships among the variables with uncertainty. Here we provide an introduction to a group of Bayesian-inference-based methods for fMRI data analysis, which were designed to detect magnitude or functional connectivity change points and to infer their functional interaction patterns based on corresponding temporal boundaries. We also provide a comparison of three popular Bayesian models, that is, Bayesian Magnitude Change Point Model (BMCPM), Bayesian Connectivity Change Point Model (BCCPM), and Dynamic Bayesian Variable Partition Model (DBVPM), and give a summary of their applications. We envision that more delicate Bayesian inference models will be emerging and play increasingly important roles in modeling brain functions in the years to come.

## 1. Introduction

An intriguing quest regarding the brain science is the following: what are the origin and the principles behind the functional architectures, which define who we are and what we are in a great extent? Compared to other methods, functional Magnetic Resonance Imaging (fMRI) has been well recognized as the most popular method that is able to explore the functional activities of the whole brain when participants are in resting states (have a rest) or performing a carefully designed task, due to its in vivo and noninvasive nature. After decades of active research, there have been numerous evidences [1–5] that the brain function is realized and emerges from the interaction of multiple concurrent neural processes or brain networks, each of which is spatially distributed across different neuroanatomical areas and exhibits statistically interdependent activity patterns. Recently, researches on brain functional connectivity or

interactions based on fMRI BOLD (fMRI blood-oxygen-level dependent) signals have received substantial interests. In previous studies, the functional connectivity was assumed to be temporal stationery [6–9]. However, the emerging evidence from neuroscience research indicates that different cortical regions are acting as adaptive processors, which involve moment-by-moment functional switching. Many studies implied that the function of different cortical regions is subject to top-down influences of brain cognitive process [10]. Task-based studies have also revealed that cognitive processes, like attention and learning, might significantly change in functional connectivity when performing tasks across the duration of a scan [11–13]. In addition, recent works suggest that fine-grained brain subnetworks, operating as communication hubs in graph theoretical terms, are equally involved in task performance [14, 15]. Besides functional connectivity, functional network connectivity (FNC), which studies the interactions in network level, requires estimating

the clusters of brain regions having similar functions and has been applied to many diseases to examine brain network differences between healthy and diseased brains [16–18]. For instance, by using a maximal lagged correlation approach, changes in FNC were found in schizophrenia, a brain disorder that is known as disrupted cognitive functions [19, 20].

In current stage, it is still challenging to infer the temporal dynamics of functional connectivity, which can be summarized as two points: identification of change points and estimation of dynamic interaction patterns. There are accumulating studies analyzing the brain functional dynamics, connectivity dynamics, and the state changes. Lindquist et al. proposed a statistical method, named as Hierarchical Exponentially Weighted Moving Average (HEWMA), on fMRI data to detect the state change of BOLD signals in response to stimulus [21]. Independent component analysis (ICA) method, including dynamic spatial ICA, was developed by Sakoglu et al. to investigate the connectivity dynamics [20]. Also, sliding-time-window-based approaches were designed to capture the dynamics of brain functional interactions across different time windows [22–27]. For example, Allen and colleagues described an approach to assess whole-brain FC dynamics based on spatial ICA, sliding-time-window correlation analysis, and  $k$ -means clustering of windowed correlation matrices [27], and it revealed unanticipated FC states which were strongly different from stationary connectivity patterns. Li and his colleagues derived functional connectomes (FCs) to characterize brain conditions from resting-state fMRI data and then FCs were divided into quasi-stable segments temporally via a sliding-time-window approach [22]. In [23], a novel framework was designed by the combination of sliding-window approach and multiview spectral clustering to extract temporally dynamic functional connectome patterns for resting-state networks, and the four detected clusters are believed to play critical roles in functional brain dynamics during resting states. In [24], a novel algorithmic framework based on hidden Markov models was presented to cluster and label the brain's functional states, represented by a large-scale functional connectivity matrix and derived via an overlapping sliding-time-window approach. The framework achieved decent classification performance on the data including 25 ADHD (attention-deficit/hyperactivity disorder) patients and 49 normal controls. Lv and his colleagues [25] also adopted the sliding-window-based method and employed a dynamic programming strategy to infer functional information transition routines on structural networks and identified the hub routers that participate in these routines most frequently.

Inspired by the multivariate graphical models based on Bayesian networks, which has been shown to be robust and reliable in estimating functional interactions and less sensitive to noise in the fMRI signals [28–30], recently, several Bayesian-inference-based methods were proposed to infer global functional interactions within brain networks and their temporal transition boundaries [31–36]. By evaluating and estimating both simulated and real data, Bayesian-inference-based methods are proved to be more powerful approaches for analysis fMRI data comparing to the methods mentioned above.

In this paper, we are aiming to pinpoint the function dynamic problems, including detecting magnitude change points, functional connectivity change points, and functional interaction patterns, which have been addressed using Bayesian inference. The organization of this paper is as follows. In Section 2, we introduce some necessary basic concepts of Bayesian inference, explaining the fMRI data with necessary preprocessing before applying Bayesian methods. Section 3 describes and compares three Bayesian models for exploring dynamic functions and their case studies. Section 4 rounds the paper off with a discussion.

## 2. Bayesian Inference

*2.1. Concept of Bayesian Inference.* Bayesian inference is a method of statistical inference in which Bayes' Rule is applied to update the probability estimate for a hypothesis as evidence is acquired, and it is the formal method for combining prior beliefs with observed (quantitative) information to answer the questions that researchers are usually interested in like “what is the probability of getting lung cancer for certain patients who smoke one pack per day?” [37]. It is a natural way to combine multiple experiments' information and it can fit realistic but complicated models.

On the other hand, Bayesian inference often costs more computationally and requires at least one of elicitation of real subjective probability distributions of prior beliefs. The good thing is that sensitivity analysis to show that the choice of prior does not strongly affect inference.

*2.1.1. Bayes' Rule [38].* Consider

$$p(\theta | y) = \frac{p(y | \theta) p(\theta)}{\sum_i p(y | \theta_i) p(\theta_i)} \quad (1)$$

if  $\theta$  is a discrete random variable with a pmf.

When  $\theta$  is continuous, Bayes' Rule becomes

$$\begin{aligned} p(\theta | y) &= \frac{p(\theta, y)}{p(y)} = \frac{p(y | \theta) p(\theta)}{p(y)} \\ &= \frac{p(y | \theta) p(\theta)}{\int_{\theta} p(y | \theta) p(\theta) d\theta}. \end{aligned} \quad (2)$$

Bayes' Rule is often written as  $p(\theta | y) \propto p(\theta)p(y | \theta)$ , when treated as a function of  $\theta$  for a fixed  $y$ , and  $p(y | \theta)$  is the likelihood  $L(y | \theta)$ . So Bayes' Rule can be thought of as

$$\text{Posterior} \propto \text{Prior} \times \text{Likelihood} \quad (3)$$

$$p(\theta | y) \propto p(\theta) p(y | \theta).$$

Note that this is expressed in words like “the posterior is proportional to the product of prior and likelihood.”

*2.1.2. Basics of the Bayesian Inference [38].* Consider the following:

- (1) Setting up a probability model.
- (2) Using the probability theory and the Bayes Rule.

We take binomial model as an example. If the sample size is  $n$  and the probability of success is  $\pi$ , we have  $y | \pi \sim \text{Bin}(n, \pi)$ . The likelihood is

$$p(y | \pi) = \binom{n}{y} \pi^y (1 - \pi)^{n-y}; \quad \pi \in [0, 1]. \quad (4)$$

If we want to make inference on  $\pi$ , given  $y$  and  $n$ , we need a prior distribution  $p(\pi)$  for  $\pi$ . We can choose a uniform distribution:

$$\begin{aligned} \pi &\sim U(0, 1) \\ p(\pi) &= \begin{cases} 1 & 0 \leq \pi \leq 1 \\ 0 & \text{Otherwise.} \end{cases} \end{aligned} \quad (5)$$

Then apply Bayes' Rule, and we get

$$\begin{aligned} p(y, \pi) &= \binom{n}{y} \pi^y (1 - \pi)^{n-y}, \\ p(y) &= \int_0^1 \binom{n}{y} \pi^y (1 - \pi)^{n-y} d\pi = \frac{1}{n+1}, \\ p(\pi | y) &= \frac{p(y, \pi)}{p(y)} = (n+1) \binom{n}{y} \pi^y (1 - \pi)^{n-y}. \end{aligned} \quad (6)$$

**2.1.3. Conjugate Prior [38].** A prior probability distribution is said to be conjugate to the sampling density if the resulting posterior distribution is a member of the same parametric family as the prior. For example, binomial likelihood  $\times$  beta prior = beta posterior.

**2.2. Bayesian Analysis Applied to fMRI Data.** An fMRI data series consists of values recorded at a specific voxel of the image at some time point  $t$ . These series are collected into a  $T$ -dimensional vector  $\vec{y} = (y_1, y_2, y_3, \dots, y_T)$ . If the data we collect has  $m$  regions of interest or  $m$  neurons based on different experiments, generally our dataset is  $m \times T$  matrix  $Y$ . A simple matrix example is shown in Figure 1.

One of our purposes is to find the dynamic functional connectivity or interaction, in which we need to establish the change point from fMRI data first. In order to apply Bayesian analysis to fMRI data and make inference to the parameters that we are interested in, like finding the change points in the dataset, we need to set up a probability model for the data and find a prior to apply Bayes' Rule.

For example, we want to find the change points in fMRI data and define a block indicator  $\vec{I} = (I_1, I_2, I_3, \dots, I_T)$  to indicate the possible locations of change points in the  $m \times T$  matrix. Now that the change point indicator is the parameter of interest and we want to make inference to it. We can assume the prior of  $\vec{I}$  is Bernoulli (0.5) so that we have  $p(\vec{I}) = \prod_{t=1}^T p(I_t)$ .

If  $\vec{I}$  is given, we can calculate the likelihood of the fMRI data matrix as  $p(Y | \vec{I})$ . Sometimes, we need to integrate some nuisance parameter to get this probability.

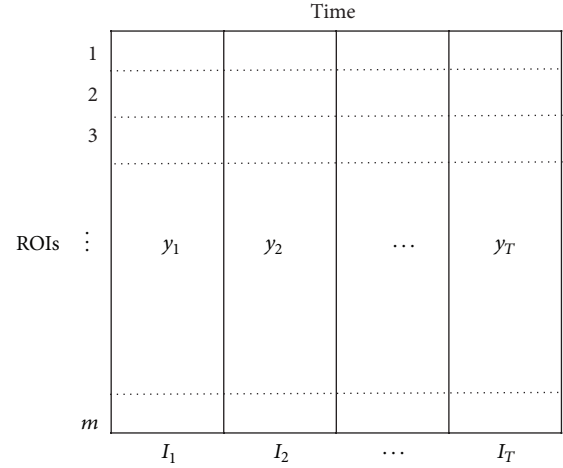


FIGURE 1: Illustration of data matrix of  $Y$ ,  $m$  ROIs, and a block indicator vector  $\vec{I}$ , where  $y_i$  is the values of all  $m$  ROIs at the time point  $i$ .

Thus, the posterior distribution of  $p(\vec{I} | Y)$  can be obtained by

$$p(\vec{I} | Y) \propto p(\vec{I}) p(Y | \vec{I}). \quad (7)$$

After the application of Bayesian analysis theoretically, we design a Markov Chain Monte Carlo (MCMC) scheme to sample the posterior with a random initial block indicator because that is the information we do not know ahead and we want to make inference. So the Bayesian MCMC is applied to the fMRI data. (The MCMC scheme is introduced in details in Section 3.)

Following these procedures, there are several models established and used. In our review, three Bayesian models are elaborated: Bayesian magnitude change point model, Bayesian connectivity change point model, and Bayesian variable partition model for detecting functional interaction and transition patterns, which contains Dynamic Bayesian variable partition model with a two-level MCMC scheme.

In real experiments, before applying the Bayesian methods reviewed here, 358 DICCCOL (Dense Individualized and Common Connectivity-based Cortical Landmarks) [39, 40] ROIs (Regions of Interest) of each subject's brain were first obtained via the publicly available open-source tools in [40] and extracted the fMRI signals. The preprocessing steps of the DTI/R-fMRI (diffusion tensor imaging/resting-state fMRI) images were similar to those used in previous publications [39–43].

### 3. Bayesian-Inference-Based Functional Dynamic Methods

In this section, we first describe two change point detecting models with their one-level MCMC scheme. Next, we introduce another powerful Bayesian model, which is able to infer functional interaction and transition patterns with temporal boundaries identified simultaneously using two-level MCMC

scheme. At the end, we summarize these three models and their applications of fMRI data analyses.

**3.1. Bayesian Magnitude Change Point Model.** Lian et al. proposed a Bayesian Magnitude Change Point Model (BMCPM) to detect group-wise consistent magnitude change points on which further pattern recognition of temporal and spatial activations was applied based [35]. A key feature of BMCPM is the capability to consider the group-wise fMRI signals of corresponding cortical landmarks across a population of subjects and optimally determines the change boundaries. In this section, we will elaborate this model in the Bayesian context.

Given a vector  $\vec{a} = (a_1, a_2, \dots, a_t)$  i.i.d. (independent and identically distributed) from a normal distribution,  $a \sim N(\mu, \sigma^2)$ , where  $t$  denotes the dimension of vector  $\vec{a}$ ,  $\mu$  denotes the mean, and  $\sigma^2$  denotes the variance. The common way to obtain the posterior distribution of a one-dimensional normal model with unknown mean and variance is to use the conjugate prior Normal-Inverse-Chi-square ( $N\text{-Inv-}\chi^2$ ) [44]. In the Bayesian theory, we know that posterior probability  $\propto$  likelihood  $\times$  prior probability. Since the conjugate prior is used for  $(\mu, \sigma^2)$ , the posterior distributions are then in the same family as the prior probability distribution. Therefore, assuming a conjugate prior  $N\text{-Inv-}\chi^2(\mu_0, \sigma_0^2/\kappa_0, v_0, \sigma_0^2)$  for  $(\mu, \sigma^2)$ , then the posterior distribution of  $(\mu, \sigma^2)$  is the  $N\text{-Inv-}\chi^2(\mu_t, \sigma_t^2/\kappa_t, v_t, \sigma_t^2)$ . So the probability of  $a_1, a_2, \dots, a_t$  can be calculated as follows:

$$p(a_1, a_2, \dots, a_t) = \left(\frac{1}{2\pi}\right)^{t/2} \frac{\sqrt{\kappa_0} \Gamma(v_t/2) (v_0 \sigma_0^2/2)^{v_0/2}}{\sqrt{\kappa_t} \Gamma(v_0/2) (v_t \sigma_t^2/2)^{v_t/2}}. \quad (8)$$

Based on (8), given a data matrix  $A = (\vec{a}_1, \vec{a}_2, \dots, \vec{a}_m)$ , where each  $\vec{a}_i$  is a vector with the data i.i.d. from the normal distribution as described above and  $\vec{a}_i$  is a vector independent from  $\vec{a}_j$  ( $i \neq j$ ), the probability of  $A$  is calculated as

$$p(A) = \prod_{i=1}^m p(\vec{a}_i), \quad (9)$$

where  $p(\vec{a}_i)$  is computed according to (8).

Magnitude change points are defined as the temporal points dividing ROI data matrix into blocks which exhibit substantial differences in brain states from each other. Figure 2 demonstrates the basic idea in BMCPM. One temporal change point located at time point  $T_{100}$  partitions the ROI data matrix into two time blocks with different distributions.

To infer the magnitude temporal change points, we can define a block indicator vector  $\vec{I} = (I_1, I_2, \dots, I_T)$ , where  $I_i = 1$  if  $i$ th observation is the beginning of a block, and otherwise  $I_i = 0$ . Therefore,  $T$  temporal observations were segmented into total  $\sum_{i=1}^T I_i$  blocks, because  $I_1 = 1$  is always considered as a change point.  $Y_i$  is denoted as the observation data of  $m$  ROIs inside the  $i$ th. For the fMRI data  $Y$  defined in

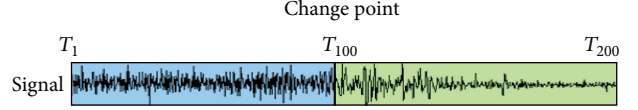


FIGURE 2: One ROI signal with one magnitude change point at time point  $T_{100}$ .

Section 2.2, the likelihood of a block indicator vector can be written as follows:

$$p(Y | \vec{I}) = \prod_{i=1}^{\sum I_j} p(Y_i), \quad (10)$$

where  $p(Y_i)$  can be calculated according to (9). Therefore, the posterior distribution of  $\vec{I}$  can be obtained as

$$p(\vec{I} | Y) \propto p(\vec{I}) p(Y | \vec{I}). \quad (11)$$

We let  $p(\vec{I}) \propto \theta^{\sum I_j} (1 - \theta)^{(T - \sum I_j)}$  (Bernoulli distribution with parameter  $\theta$ ) which may also be modified to reflect our prior knowledge of the estimated number of temporal blocks.

The BMCPM has been successfully applied to the operational span (OSPAN) data, a working memory task-based fMRI dataset with 10 participants, which were acquired on a 3T GE Signa scanner. Totally 81 ROIs were shown with posterior probability larger than 0. Clustering analysis on these 81 ROIs found that some clusters are highly related to specific brain functions. For example, one of the clusters only locates at areas of V1, V2, and motor cortex with well response to low level task-design, while another cluster locates on the frontal or parietal lobes with respect to information processing and responses. More results can be found in [35].

**3.2. Bayesian Functional Connectivity Change Point Model.** In order to analyze the joint probabilities among the nodes of brain networks between different time periods, a Bayesian Connectivity Change Point Model (BCCPM) [34] was proposed to determine the temporal boundary where there is an abrupt change of multivariate functional interactions in the brain networks. Different from BMCPM, which considers ROIs independent of each other, BCCPM infers the boundaries of temporal blocks via a unified Bayesian framework by analyzing the dynamics of multivariate functional interactions.

Given a vector  $\{b_1, b_2, \dots, b_t\}$  i.i.d. from  $m$ -dimensional multivariate normal distribution  $b_i \sim N(\vec{\mu}, \Sigma)$   $i = 1, 2, \dots, t$ , where  $t$  denotes the number of vectors,  $m$  denotes the dimension of vector  $b_i$ ,  $\vec{\mu}$  denotes the  $m$ -dimensional mean vector, and  $\Sigma$  denotes  $m \times m$  the covariance matrix. The common way to obtain the posterior distribution of a multidimensional normal model with unknown mean and covariance matrix is to use the conjugate prior Normal-Inverse-Wishart ( $N\text{-Inv-Wishart}$ ) [44]. Therefore, assuming a conjugate prior  $N\text{-Inv-Wishart}(\mu_0, \Lambda_0/\kappa_0, v_0, \Lambda_0)$  for  $(\vec{\mu}, \Sigma)$ , the posterior distribution of  $(\vec{\mu}, \Sigma)$  is  $N\text{-Inv-Wishart}(\mu_t, \Lambda_t/\kappa_t, v_t, \Lambda_t)$ .

Since we are interested in the posterior distribution of the configuration, the joint probability of  $b_1, b_2, \dots, b_t$  is calculated as follows:

$$p(b_1, b_2, \dots, b_t) = \left(\frac{1}{2\pi}\right)^{mt/2} \left(\frac{\kappa_0}{\kappa_T}\right)^{m/2} \frac{\Gamma_m(\gamma_t/2) (\det(\Lambda_0))^{\gamma_0/2}}{\Gamma_m(\gamma_0/2) (\det(\Lambda_t))^{\gamma_t/2}} 2^{mt/2}, \quad (12)$$

where  $\Gamma_m$  is the multivariate gamma function.

BCCPM are interested in the connectivity change points, which define the temporal segments where there are underlying differences in the joint probabilities (defining functional interactions) among  $m$  ROIs between different time-periods. Figure 3 demonstrates the basic idea in BCCPM. Similar to BMCPM, a block indicator vector is introduced as  $\vec{I} = (I_1, I_2, \dots, I_T)$ . The marginal likelihood of the data matrix  $Y = (y_1, y_2, \dots, y_T)$  can be computed as follows:

$$p(Y | \vec{I}) = \prod_{i=1}^{\sum I_j} p(Y_i), \quad (13)$$

where  $p(Y_i)$  is calculated according to (12). Please note that one important assumption in the BCCPM is the statistical independence among the temporal segments (blocks). Similar to BMCPM, the posterior distribution of the configuration  $p(\vec{I} | Y)$  can be obtained by (11). Note that blocks indicated by  $\vec{I}$  are mutually independent to each other across  $T$ .

BCCPM has been applied to analyze the ADHD data [31], including 25 ADHD-c patients and 49 normal development children as NCs, coming from the Imaging Center for Brain Research, Beijing Normal University. All the ROI time series are partitioned into three blocks with two change points detected by BCCPM. Further analysis based on dictionary learning algorithm identified two pairs of atomic functional interaction patterns. The first pair shared 310 common connections, while the second pair only shared 27 common connections. The matrixes recovered from these two pairs gave 100% discriminations between ADHD patients and NC subjects. More results can be found in [31].

**3.3. Bayesian Change Point Model Using One-Level MCMC Scheme.** Metropolis-Hastings (MH) scheme is widely used for calculating the Bayesian inference. A one-level MH (MCMC) scheme for calculating the posterior distribution under BMCPM or BCCPM is provided as follows with a randomly initialized block indicator vector  $\vec{I}^0$  and user defined iteration number  $N$ :

- (1) Generate a new block indicator vector  $\vec{I}^*$  by randomly switching the value of an element in  $\vec{I}^{n-1}$ . And calculate  $p(\vec{I}^* | Y)$  according to (11).
- (2) Generate a random number  $u$  from uniform (0, 1) and set

$$\vec{I}^n = \begin{cases} \vec{I}^* & \text{if } u \leq \min \left[ 1, \frac{p(\vec{I}^* | Y)}{p(\vec{I}^{n-1} | Y)} \right] \\ \vec{I}^{n-1} & \text{otherwise.} \end{cases} \quad (14)$$

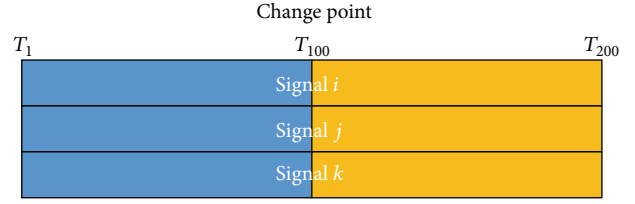


FIGURE 3: Three ROI signals with one connectivity change point at time point  $T_{100}$  where the multivariate normal distribution inside the block  $T_1-T_{100}$  of color blue is different from the distribution of signal in the rest block of color orange.

- (3) Iterate step (1) and step (2) until  $n$  reaches the given number  $N$ .
- (4) Finally, the posterior probability for each time point  $(1, 2, \dots, T)$  being a change point is calculated from MCMC samples without burn-in samples.

By default, the control parameters, that is,  $\mu_0, \kappa_0, \nu_0, \sigma_0^2, \Lambda_0$ , are fixed as constants. To determine the number of iterations to ensure Markov chains converged with contain configurations, the trace plot of posterior probability and Gelman and Rubin scale reduction factor [45] can be employed.

**3.4. Dynamic Bayesian Variable Partition Model.** Recently, there are several studies which utilized the sliding time window based framework to model multivariate functional connectivity interaction, as well as other works that used graphical modeling methods to detect temporal brain dynamics as mentioned in Section 1. However, it is much needed for an integrated framework to infer the representative signature patterns of the multivariate functional interactions and to simultaneously characterize the temporal transitions of these signature patterns. For this purpose, Zhang et al. [36] proposed a Dynamic Bayesian Variable Partition Model (DBVPM) to simultaneously infer global functional interactions within brain networks and their temporal transition boundaries. To capture all the conditional independence global structure, two dependence structures, that is, chain- and  $V$ -dependence structures, are designed, which will be first elaborated in the following paragraphs. The DBVPM is trying to simultaneously infer the temporal change points and corresponding global structures inside these temporal blocks.

**3.4.1. Chain-Dependence Model.** A group of variables  $C_G$  follows a chain-dependence model if the index set  $G$  can be partitioned into three subsets  $U, V$ , and  $W$  such that  $C_U$  and  $C_W$  are independent given  $C_V$ , such as  $C_U \rightarrow C_V \rightarrow C_W$ . The joint distribution of a chain-dependence model is

$$p(C_G) = p(C_U) p(C_V | C_U) p(C_W | C_V) = \frac{F(C_V, C_U) F(C_W | C_V)}{F(C_V)}, \quad (15)$$

where  $F(C_V, C_U, \dots)$  is the joint probability of  $(C_V, C_U, \dots)$ .

**3.4.2. V-Dependence Model.** A group of variables  $C_G$  follows V-dependence model if the index set  $G$  can be partitioned into three subsets  $U$ ,  $V$ , and  $W$  such that  $C_U$  and  $C_W$  are mutually independent, that is  $C_U \rightarrow C_V \leftarrow C_W$ . The joint distribution of a chain-dependence model is

$$\begin{aligned} p(C_G) &= p(C_U) p(C_W) p(C_V | C_U, C_W) \\ &= F(C_U) F(C_W) \frac{F(C_U, C_V, C_W)}{F(C_U, C_W)}, \end{aligned} \quad (16)$$

where  $F(C_V, C_U, \dots)$  is the joint probability of  $(C_V, C_U, \dots)$ . The calculation of this joint probability will be discussed in next section.

**3.4.3. Dynamic Bayesian Variable Partition Model.** DBVPM uses the similar Bayesian inference as in BCCPM except that the numbers of dimension in the multivariate normal distribution are not fixed across temporal order: specifically, given  $c_1, c_2, \dots, c_t$  i.i.d. observations from the  $r$ -dimensional multivariate normal distribution,  $c_i \sim N(\vec{\mu}, \Sigma)$   $i = 1, 2, \dots, t$ . To calculate the marginal distribution of the data  $c_1, c_2, \dots, c_t$ , we only need to change the dimension from  $m$  to  $r$  in (12), which is the same calculation for the joint probability in (15) and (16), and  $r$  is determined by the joint of variables. Figure 4 demonstrates the basic idea in DBVPM.

In DBVPM, to infer the dependence structure (chain or V structure) among  $m$  ROIs, another indicator vector  $\vec{Y} = (Y_1, Y_2, \dots, Y_m)$  is used to denote the grouping of the index  $G$  of ROIs to subgroups  $U$ ,  $V$ , and  $W$ , where  $Y_i = j$  means  $i$ th ROI is grouped in subgroup  $j$  ( $j = 0$  means  $U$ ,  $j = 1$  means  $V$ , and  $j = 2$  means  $W$ ), and a binary indicator  $S$  is used to denote the dependence structures, that is, chain- or V-dependence structure. The posterior distribution for observations in  $i$ th temporal block can be calculated as

$$p(\vec{Y}, S | Y_i) \propto p(Y_i | \vec{Y}, S) p(\vec{Y}) p(S), \quad (17)$$

where  $Y_i$  is the same definition in previous sections and  $p(Y_i | \vec{Y}, S)$  is calculated as (15) when  $S = 0$  and (16) when  $S = 1$ .

To incorporate the multivariate functional interactions and temporal dynamics in DBVPM, the same block indicator  $\vec{I} = (I_1, I_2, \dots, I_T)$  is used. Let  $\vec{S} = (S_1, S_2, \dots, S_{\sum I_i})$  be the structure indicator vector, and let  $\vec{Y} = (\vec{Y}_1, \vec{Y}_2, \dots, \vec{Y}_{\sum I_i})$ , where  $\vec{Y}_i = (Y_1, Y_2, \dots, Y_m)$  with  $Y_j = 0, 1, 2$ , be the partition indicator vector in the  $i$ th block. The posterior distribution of the data matrix  $Y = (y_1, y_2, \dots, y_T)$  can be represented as follows:

$$p(\vec{I}, \vec{Y}, \vec{S} | Y) \propto p(Y | \vec{I}, \vec{Y}, \vec{S}) p(\vec{I}, \vec{Y}, \vec{S}), \quad (18)$$

where  $p(Y | \vec{I}, \vec{Y}, \vec{S}) = \prod p(Y_i | \vec{Y}_i, S_i)$  and  $p(\vec{I}, \vec{Y}, \vec{S}) = p(\vec{I}) \prod p(\vec{Y}_i, S_i | I_i)$ . By default, the uniform prior can be used for  $p(\vec{I})$  and  $p(\vec{Y}_i, S_i | I_i)$ .

DBVPM has been applied to analyze the posttraumatic stress disorder (PTSD) data with 45 patients and 53 healthy controls coming from the Second Xiangya Hospital and the Central South of University on a 3T MRI (magnetic

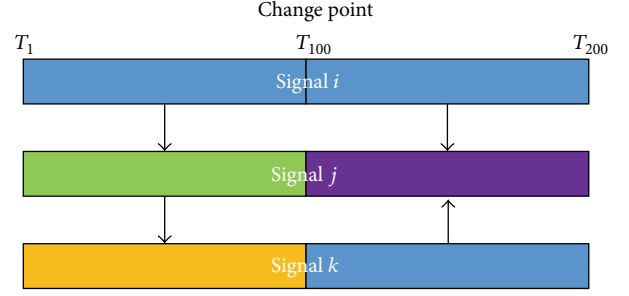


FIGURE 4: Three ROI signals with one temporal change point at time point  $T_{100}$  where a chain dependence structure is in the left block and V dependence structure is in the right block.

resonance imaging) scanner. From the difference between the manually labeled change points and the DBVPM-derived time for 98 subjects, it showed that most change points only have 3.5 time-point distance. Further analysis found some substantial state distribution differences based on the detected functional interactions between PTSD patients and healthy controls, which may need additional structural or neural basis investigation in the future. More results can be found in [36].

**3.5. DBVPM Using Two-Level MCMC Scheme.** A two-level MH (MCMC) scheme can be applied to sample from the posterior distribution of the block boundaries and dependency structures within each block: Given the block boundaries, the lower level MCMC samples from the posterior distribution of dependency structures within a block and the higher level MCMC samples from the posterior distribution of block boundaries. Specifically, the lower level MCMC involves alternating between the chain and V structures and changing the group labels of each variable of ROI. The likelihood can be calculated using (15), (16), and (18). The higher level MCMC involves segmenting one block into two, merging two neighboring blocks, and shifting a block boundary. In each higher level step, every block runs through a lower level MCMC. A dependency structure is sampled for each block as the dependency structure for that block in the higher level proposal. Then the log likelihood of the proposal can be calculated by summing up the log likelihood of each block. More details can be found in [36].

**3.6. Bayesian Model Comparison.** We conducted a systematical comparison of the above three Bayesian inference models and summarized it in Table 1. The major differences of these models come from their assumptions. In BMCPM, no explicit connection is assumed between ROIs, so the one-dimensional normal model is employed to capture this condition. In BMCPM, all the ROIs are linked together to follow a multi-dimensional normal distribution. In DBVPM, it tries to capture more complex connections between ROIs, so the chain- and V-dependence structures are used. As mentioned in Section 1, Bayesian inference methods have been applied to solve three issues in the analysis of fMRI data: detecting magnitude change points; detecting functional connectivities change points; and identifying functional interaction

TABLE 1: Summary of BMCPM, BCCPM, and DBVPM.

	BMCPM	BCCPM	DBVPM
MCMC scheme	One-level	One-level	Two-level
Ability to infer magnitude change points	Yes	Yes	Yes
Ability to infer functional connectivity change points	No	Yes	Yes
Ability to infer functional interaction patterns	No	No	Yes
Running time			
50 ROIs, 200 time points	2 seconds	49 seconds	948 minutes
300 ROIs, 200 time points	11 seconds	41 minutes	NA
1000 ROIs, 200 time points	36 seconds	483 minutes	NA

patterns. Since DBVPM uses the chain- and  $V$ -dependence structures to capture global functional interactions, it owns the power to identify not only all possible change points but also functional interaction patterns, comparing to BMCPM and BCCPM. However, an explicit drawback of DBVPM is the convergence speed that it takes a tremendous time (several weeks on an average power server) to converge when dealing with more than 50 ROIs. This is the reason that selected ROIs from certain brain networks were used in the experiments of real data. We did a simulation test to measure the convergence speed and robust of these three methods. We generated three datasets with 200 time points fixed and the number of ROIs set to 50, 300, and 1000, respectively. The data are generated from the normal distributions, and one change point is embedded at time point 101. Each dataset has five replicas. The iterations are set to 2000 for one-level MCMC and 2000 for both levels for two-level MCMC as suggested by the original studies. We used the machine with the Linux OS, Intel Xeon E5-2620, and 64 GB memory. The average running times (wall clock time) are shown at the end of Table 1. Note that NA means that the method did not finish the processing in three days. Three indicator vectors were used in DBVPM that cause one-level MCMC scheme not suitable for DBVPM. Although BMCPM is only capable of finding magnitude change points, its execution is very fast. Typically, for 358 ROIs, it only needs no more than 2000 iterations and a couple of hours to converge, so it is efficient for handling large fMRI time series data. For BCCPM of multivariate normal distribution assumption, it is good at inference of the boundaries of temporal blocks based on dynamics of multivariate functional interactions. It also converges fast when analyzing data with 358 ROIs. Therefore, if there are some better approaches, like dictionary learning, to further infer functional interactions between ROIs, BCCPM should be the best choice as the first step to establish temporal change points for functional interactions and dynamics.

#### 4. Summary and Discussion

In this review paper, we presented and summarized three important applications of the Bayesian inference paradigm to fMRI data. By assuming that normal distribution for each ROI inside temporal block and ROIs are independent of each other, the BMCPM calculates the posterior probability of the

temporal block indicator vectors using the conjugate prior of  $N\text{-Inv-}\chi^2$ . The essential feature of BMCPM is the capability of considering the group-wise ROIs in the fMRI signals across a population of subjects and optimally determining the change boundaries. BMCPM converges very fast, and it is able to analyze more than 1000 ROIs, but it only detects magnitude change points. Different from BMCPM considering ROIs independent of each other, the BCCPM assumes that all  $m$  ROIs are an integrated variable that follows an  $m$ -dimensional multivariate normal distribution, and thus it can detect function connectivity change points. BCCPM, like BMCPM, only needs one-level MCMC scheme to sample the posterior distribution and converges relatively fast when there are no more than 1000 ROIs. To derive the representative signature patterns of the multivariate functional interactions within temporal blocks, DBVPM was proposed to incorporate two structures, chain- and  $V$ -dependence structures. Comparing to BMCPM and BCCPM, DBVPM is able to identify the functional interaction patterns in addition to the boundaries of temporal blocks. The cost of DBVPM is that it spends unacceptable time to analyze fMRI data with more than 50 ROIs. The performance of all three Bayesian models were illustrated on three datasets, including task-based OSPAN dataset, attention-deficit/hyperactivity disorder dataset, and posttraumatic stress disorder dataset, which confirmed the superiority of Bayesian inference methods for functional dynamic analysis on fMRI data.

There are still some challenging issues for using Bayesian inference to analyze functional dynamics of fMRI data. For example, from the evaluation of BCCPM on simulation for local dynamics, it showed that BCCPM is not sensitive in detecting small local changes of structure. Also, although global structure inside the temporal blocks can be inferred from DBVPM, the detailed interaction patterns and network structure between ROIs are unknown. Functional connectivity network is also one of the critical applications by evaluating regional interactions using fMRI. One of the possible extensions of BCCPM and DBVPM is to infer the detailed functional connectivity network by estimating the covariance matrix  $\Sigma$ , which could be a valuable future direction for the Bayesian models. As we know, most of the task-based fMRI experiments fall into two categories: (1) block design and (2) event-related design. Currently, the most Bayesian methods are specified and verified for tackling problems based on the data from block design

experiments. Comparing to block design, MR (magnetic resonance) signal of efMRI (event-related fMRI) is smaller, and statistical power is relatively lower due to the complexity and the lower SNR (Signal-to-Noise Ratio). The Bayesian model could be a possible tool to analyze event-related design data as well because no explicit assumption is made about what caused the difference in brain connectivity among ROIs. The difference can be originated from either tasks or events. In the future, systematical analyses on simulated and real data are needed to verify the effectiveness and accuracy of Bayesian models for event-based fMRI data, which is an open question and one of the future extensions for Bayesian-inference-based approach.

### Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

### Acknowledgments

The authors are grateful for support from Georgia State University Brains-Behavior Seed grant. This research was supported by the Molecular Basis of Disease (MBD) at Georgia State University.

### References

- [1] E. Fedorenko, J. Duncan, and N. Kanwisher, "Broad domain generality in focal regions of frontal and parietal cortex," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, no. 41, pp. 16616–16621, 2013.
- [2] M. D. Fox, A. Z. Snyder, J. L. Vincent, M. Corbetta, D. C. V. Essen, and M. E. Raichle, "The human brain is intrinsically organized into dynamic, anticorrelated functional networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 27, pp. 9673–9678, 2005.
- [3] N. U. F. Dosenbach, K. M. Visscher, E. D. Palmer et al., "A core system for the implementation of task sets," *Neuron*, vol. 50, no. 5, pp. 799–812, 2006.
- [4] J. Duncan, "The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour," *Trends in Cognitive Sciences*, vol. 14, no. 4, pp. 172–179, 2010.
- [5] L. Pessoa, "Beyond brain regions: network perspective of cognition-emotion interactions," *Behavioral and Brain Sciences*, vol. 35, no. 3, pp. 158–159, 2012.
- [6] K. Li, D. Zhu, L. Guo et al., "Connectomics signatures of prenatal cocaine exposure affected adolescent brains," *Human Brain Mapping*, vol. 34, no. 10, pp. 2494–2510, 2013.
- [7] B. C. Dickerson and R. A. Sperling, "Large-scale functional brain network abnormalities in alzheimer's disease: insights from functional neuroimaging," *Behavioural Neurology*, vol. 21, no. 1-2, pp. 63–75, 2009.
- [8] M. D. Fox and M. E. Raichle, "Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging," *Nature Reviews Neuroscience*, vol. 8, no. 9, pp. 700–711, 2007.
- [9] M.-E. Lynall, D. S. Bassett, R. Kerwin et al., "Functional connectivity and brain networks in schizophrenia," *The Journal of Neuroscience*, vol. 30, no. 28, pp. 9477–9487, 2010.
- [10] C. D. Gilbert and M. Sigman, "Brain states top-down influences in sensory processing," *Neuron*, vol. 54, no. 5, pp. 677–696, 2007.
- [11] P. Fransson, "How default is the default mode of brain function? Further evidence from intrinsic BOLD signal fluctuations," *Neuropsychologia*, vol. 44, no. 14, pp. 2836–2845, 2006.
- [12] F. T. Sun, L. M. Miller, A. A. Rao, and M. D'Esposito, "Functional connectivity of cortical networks involved in bimanual motor sequence learning," *Cerebral Cortex*, vol. 17, no. 5, pp. 1227–1234, 2007.
- [13] C. Chang and G. H. Glover, "Time-frequency dynamics of resting-state brain connectivity measured with fMRI," *NeuroImage*, vol. 50, no. 1, pp. 81–98, 2010.
- [14] P. Hagmann, L. Cammoun, X. Gigandet et al., "Mapping the structural core of human cerebral cortex," *PLoS Biology*, vol. 6, no. 7, article e159, pp. 1479–1493, 2008.
- [15] R. L. Buckner, J. Sepulcre, T. Talukdar et al., "Cortical hubs revealed by intrinsic functional connectivity: mapping, assessment of stability, and relation to Alzheimer's disease," *The Journal of Neuroscience*, vol. 29, no. 6, pp. 1860–1873, 2009.
- [16] M. Assaf, K. Jagannathan, V. Calhoun, M. Kraut, J. Hart Jr., and G. Pearlson, "Temporal sequence of hemispheric network activation during semantic processing: a functional network connectivity analysis," *Brain and Cognition*, vol. 70, no. 2, pp. 238–246, 2009.
- [17] V. D. Calhoun, K. A. Kiehl, and G. D. Pearlson, "Modulation of temporally coherent brain networks estimated using ICA at rest and during cognitive tasks," *Human Brain Mapping*, vol. 29, no. 7, pp. 828–838, 2008.
- [18] D. Zhu, K. Li, D. P. Terry et al., "Connectome-scale assessments of structural and functional connectivity in MCI," *Human Brain Mapping*, vol. 35, no. 7, pp. 2911–2923, 2014.
- [19] M. J. Jafri, G. D. Pearlson, M. Stevens, and V. D. Calhoun, "A method for functional network connectivity among spatially independent resting-state components in schizophrenia," *NeuroImage*, vol. 39, no. 4, pp. 1666–1681, 2008.
- [20] Ü. Sakoğlu, G. D. Pearlson, K. A. Kiehl, Y. M. Wang, A. M. Michael, and V. D. Calhoun, "A method for evaluating dynamic functional network connectivity and task-modulation: application to schizophrenia," *Magnetic Resonance Materials in Physics, Biology and Medicine*, vol. 23, no. 5-6, pp. 351–366, 2010.
- [21] M. A. Lindquist, C. Waugh, and T. D. Wager, "Modeling state-related fMRI activity using change-point theory," *NeuroImage*, vol. 35, no. 3, pp. 1125–1141, 2007.
- [22] X. Li, D. Zhu, X. Jiang et al., "Dynamic functional connectomics signatures for characterization and differentiation of PTSD patients," *Human Brain Mapping*, vol. 35, no. 4, pp. 1761–1778, 2014.
- [23] X. Zhang, X. Li, C. Jin et al., "Identifying and characterizing resting state networks in temporally dynamic functional connectomes," *Brain Topography*, vol. 27, no. 6, pp. 747–765, 2014.
- [24] J. Ou, L. Xie, P. Wang et al., "Modeling brain functional dynamics via hidden Markov models," in *Proceedings of the 6th International IEEE/EMBS Conference on Neural Engineering (NER '13)*, pp. 569–572, IEEE, San Diego, Calif, USA, November 2013.
- [25] P. Lv, L. Guo, X. Hu et al., "Modeling dynamic functional information flows on large-scale brain networks," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2013*, Springer, 2013.
- [26] X. Zhang, L. Guo, X. Li et al., "Characterization of task-free and task-performance brain states via functional connectome



- patterns,” *Medical Image Analysis*, vol. 17, no. 8, pp. 1106–1122, 2013.
- [27] E. A. Allen, E. Damaraju, S. M. Plis, E. B. Erhardt, T. Eichele, and V. D. Calhoun, “Tracking whole-brain connectivity dynamics in the resting state,” *Cerebral Cortex*, vol. 24, no. 3, pp. 663–676, 2014.
- [28] J. D. Ramsey, S. J. Hanson, and C. Glymour, “Multi-subject search correctly identifies causal connections and most causal directions in the DCM models of the Smith et al. simulation study,” *NeuroImage*, vol. 58, no. 3, pp. 838–848, 2011.
- [29] J. Sun, X. Hu, X. Huang et al., “Inferring consistent functional interaction patterns from natural stimulus fMRI data,” *NeuroImage*, vol. 61, no. 4, pp. 987–999, 2012.
- [30] S. M. Smith, K. L. Miller, G. Salimi-Khorshidi et al., “Network modelling methods for fMRI,” *NeuroImage*, vol. 54, no. 2, pp. 875–891, 2011.
- [31] J. Ou, Z. Lian, L. Xie et al., “Atomic dynamic functional interaction patterns for characterization of ADHD,” *Human Brain Mapping*, vol. 35, no. 10, pp. 5262–5278, 2014.
- [32] J. Ou, L. Xie, C. Jin et al., “Characterizing and differentiating brain state dynamics via hidden Markov Models,” *Brain Topography*, vol. 28, no. 5, pp. 666–679, 2015.
- [33] Z. Lian, X. Li, H. Zhang et al., “Detecting cell assembly interaction patterns via Bayesian based change-point detection and graph inference model,” in *Proceedings of the IEEE 11th International Symposium Biomedical Imaging (ISBI '14)*, Beijing, China, April 2014.
- [34] Z. Lian, X. Li, J. Xing et al., “Exploring functional brain dynamics via a Bayesian connectivity change point model,” in *Proceedings of the 11th International Symposium on Biomedical Imaging (ISBI '14)*, pp. 600–603, IEEE, Beijing, China, April 2014.
- [35] Z. Lian, J. Lv, J. Xing et al., “Generalized fMRI activation detection via Bayesian magnitude change point model,” in *Proceedings of the IEEE 11th International Symposium on Biomedical Imaging (ISBI '14)*, pp. 21–24, IEEE, Beijing, China, April 2014.
- [36] J. Zhang, X. Li, C. Li et al., “Inferring functional interaction and transition patterns via dynamic bayesian variable partition models,” *Human Brain Mapping*, vol. 35, no. 7, pp. 3314–3331, 2014.
- [37] J. Kershaw, B. A. Ardekani, and I. Kanno, “Application of Bayesian inference to fMRI data analysis,” *IEEE Transactions on Medical Imaging*, vol. 18, no. 12, pp. 1138–1153, 1999.
- [38] Bayesian Inference on Wikipedia, [https://en.wikipedia.org/wiki/Bayesian\\_inference](https://en.wikipedia.org/wiki/Bayesian_inference).
- [39] D. Zhu, K. Li, C. C. Faraco et al., “Optimization of functional brain ROIs via maximization of consistency of structural connectivity profiles,” *NeuroImage*, vol. 59, no. 2, pp. 1382–1393, 2012.
- [40] D. Zhu, K. Li, L. Guo et al., “DICCCOL: dense individualized and common connectivity-based cortical landmarks,” *Cerebral Cortex*, vol. 23, no. 4, pp. 786–800, 2013.
- [41] D. Zhu, D. Zhang, C. Faraco et al., “Discovering dense and consistent landmarks in the brain,” in *Information Processing in Medical Imaging*, G. Székely and H. K. Hahn, Eds., vol. 6801 of *Lecture Notes in Computer Science*, pp. 97–110, 2011.
- [42] T. Zhang, L. Guo, K. Li et al., “Predicting functional cortical ROIs via DTI-derived fiber shape models,” *Cerebral Cortex*, vol. 22, no. 4, Article ID bhr152, pp. 854–864, 2012.
- [43] K. Li, L. Guo, C. Faraco et al., “Individualized ROI optimization via maximization of group-wise consistency of structural and functional profiles,” in *Advances in Neural Information Processing Systems*, pp. 1369–1377, MIT Press, 2010.
- [44] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis*, vol. 2, Chapman & Hall, London, UK, 2014.
- [45] A. Gelman and D. B. Rubin, “Inference from iterative simulation using multiple sequences,” *Statistical Science*, vol. 7, no. 4, pp. 457–472, 1992.