ELSEVIER

NeuroImage

www.elsevier.com/locate/ynimg
NeuroImage 30 (2006) 794 – 812

A Bayesian approach to modeling dynamic effective connectivity with fMRI data ☆

Sourabh Bhattacharya,[a,1] Moon-Ho Ringo Ho,[b,c,*] and Sumitra Purkayastha [d,2]

[a] *Applied Statistics Unit, Indian Statistical Institute, 203 B.T. Road, Kolkata 700 108, India*
[b] *Division of Psychology, Nanyang Technological University, 639798, Singapore*
[c] *Department of Psychology, McGill University, Montreal, Canada, QC H3A 1B1*
[d] *Theoretical Statistics and Mathematics Unit, Indian Statistical Institute, 203 B.T. Road, Kolkata 700 108, India*

Received 2 June 2005; revised 6 October 2005; accepted 10 October 2005
Available online 20 December 2005

A state-space modeling approach for examining dynamic relationship between multiple brain regions was proposed in Ho, Ombao and Shumway (Ho, M.R., Ombao, H., Shumway, R., 2005. A State-Space Approach to Modelling Brain Dynamics to Appear in Statistica Sinica). Their approach assumed that the quantity representing the influence of one neuronal system over another, or effective *connectivity*, is time-invariant. However, more and more empirical evidence suggests that the connectivity between brain areas may be dynamic which calls for temporal modeling of effective connectivity. A Bayesian approach is proposed to solve this problem in this paper. Our approach first decomposes the observed time series into measurement error and the BOLD (blood oxygenation level-dependent) signals. To capture the complexities of the dynamic processes in the brain, region-specific activations are subsequently modeled, as a linear function of the BOLD signals history at other brain regions. The coefficients in these linear functions represent effective connectivity between the regions under consideration. They are further assumed to follow a random walk process so to characterize the dynamic nature of brain connectivity. We also consider the temporal dependence that may be present in the measurement errors. ML-II method (Berger, J.O., 1985. Statistical Decision Theory and Bayesian Analysis (2nd ed.). Springer, New York) was employed to estimate the hyperparameters in the model and Bayes factor was used to compare among competing models. Statistical inference of the effective connectivity coefficients was based on their posterior distributions and the corresponding Bayesian credible regions (Carlin, B.P., Louis, T.A., 2000. Bayes and Empirical Bayes Methods for Data Analysis (2nd ed.). Chapman and Hall, Boca Raton). The proposed method was applied to a functional magnetic resonance imaging data set and results support the theory of attentional control network and demonstrate that this network is dynamic in nature.

Keywords: Bayes factor; Bayesian inference; Effective connectivity; Functional magnetic resonance imaging; Gibbs sampling; Human brain mapping; MCMC; Model selection

Introduction

Functional magnetic resonance imaging (fMRI) is a technique to determine which parts of the brain are activated by different types of physical sensation or activity, such as sight, sound or the movement of a subject's fingers. In functional brain imaging studies, the signal changes of interest are caused by the neuronal activity, but such electrical activity is not directly detectable by fMRI. The signal being measured in fMRI experiments is called the blood oxygenation level-dependent (BOLD) response, which is a consequence of the hemodynamic changes (including local changes in blood flow, volume and oxygenation level) occurring within a few seconds of changes in neuronal activity induced by external stimuli. The BOLD signal is usually used as a proxy for the underlying neuronal activity. Most fMRI studies concern the detection of sites of activation ('hot-spots') in the brain and their relationship to the experimental stimulation and we refer to these studies as *activation* studies. Interested readers can refer to Clare (1997), Frackowiak et al. (2004), and Jezzard et al. (2001) for the details on the theory and application of fMRI in neurology, neuroscience, psychology and psychiatry. There are also a large amount of recent works on estimation of the hemodynamic response function (Glover, 1999; Burock and Dale, 2000; Genovese, 2000; Marrelect et al., 2003; Donnet et al., 2004; Woolrich et al., 2004a). These studies do not reveal a great deal about how different brain regions relate or 'communicate' to each other.

☆ Authors' names appear in alphabetical order. All authors equally contributed to the paper.

* Corresponding author. Department of Psychology, McGill University, Montreal, Canada, QC H3A 1B1. Fax: +65 6794 6303.
E-mail addresses: sourabh@isds.duke.edu (S. Bhattacharya), homh@ntu.edu.sg (M.-H. Ringo Ho), sumitra@isical.ac.in (S. Purkayastha).
[1] Current address: Institute of Statistics and Decision Sciences, Duke University Durham, NC 27708-0251, USA.
[2] Currently visiting: Department of Biostatistics, University of Michigan, School of Public Health 1420 Washington Heights, Ann Arbor, MI 48109-2029, USA.

Available online on ScienceDirect (www.sciencedirect.com).

1053-8119/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2005.10.019

Disparate brain regions do not operate in isolation. There is growing interest in studying the possible interactions between different brain regions to understand the functional organization of the brain. The study of the influence of one neuronal system over another is usually referred to as *effective connectivity* analysis (Friston, 1994; Nyberg and McIntosh, 2001) in brain imaging literature. Two common approaches, namely, structural equation modeling (see, for example, McIntosh and Gonzalez-Lima, 1994; Kirk et al., 2005; Penny et al., 2004b) and time-varying parameter regression (see, for example, Büchel and Friston, 1998), have been applied to fMRI data for studying effective connectivity. However, these approaches suffer from several limitations. For both approaches, fMRI researchers first select their regions of interest. Each region of interest usually consists of a few voxels. The first mode of the Singular Value Decomposition of the time series of these voxels is used to represent the temporal response for the selected region of interest. In the applications of structural equation modeling, a within-subject covariance matrix of the regions of interest is derived and a path model specifying the connections between brain regions is then fitted to this matrix. The strength of effective connectivity is measured by the path or structural coefficients in this method. This approach ignores the temporal correlation in the data which can lead to inaccurate standard errors and test statistics. Connectivity between brain regions is also assumed to be time-invariant.

Time-varying parameter regression, on the other hand, relaxes this assumption and allows time-varying connectivity. In the applications of time-varying parameter regression, one brain area's time series $y_i(t)$ is regressed on another brain area's time series $y_2(t)$ as (Büchel and Friston, 1998):

$$y_1(t) = \beta(t)y_2(t) + e(t),$$

$$\beta(t) = \beta(t-1) + w(t),$$

where $e(t)$ and $w(t)$ are white noises and are independent of each other. The regression coefficient, $\beta(t)$, following a random walk process, measures the dynamic effective connectivity in this approach. Applications of time-varying parameter regression, however, have been primarily limited to studying the relationship between two brain regions so far. A more general way to characterize the dynamics of $\beta(t)$ can be through functional coefficients model which was first explored in Chen and Tsay (1993) and later extended to nonlinear time series model by Cai et al. (2000). Harrison et al. (2003) proposed the use of multivariate/vector autoregressive (MAR) model for modeling effective connectivity. They included interaction terms between pairs of contemporaneous regional time series to account for the nonlinear inter-regional dependence in the model. Lagged regional time series were also included to account for the temporal autocorrelation. Their approach models the behavior of the brain system simply by quantifying the relationships within the measured data only. Our approach differs from theirs in a way that we model the brain as a dynamic system and attempts to account for correlations within the data by invoking state variables whose dynamics generate the fMRI data. The MAR approach used by Harrison et al. (2003) models temporal effects across different brain regions without using state variables and inter-regional dependencies within data are characterized in terms of the historical influence one region has on another. Ho et al. (2005) (henceforth abbreviated to HOS) proposed a state-space approach for studying effective connectivity, which overcomes some of the limitations encountered in the structural equation modeling and time-varying parameter regression methods.

Horwitz (1998) and Horwitz et al. (1999) refer to all the aforementioned approaches collectively as *system-level neural modeling*, which attempts to address the problem that large covariance in inter-regional activity can come about by direct and indirect effects. Recently, Horwitz and his colleagues proposed another approach, referred to as *large-scale neural modeling*, using neurobiologically realistic networks to simulate neural data at multiple spatial and temporal levels, such as single unit electrophysiological data (Deco et al., 2004) and Positron Emission Tomography (PET) data (Tagamets and Horwitz, 1998). Their approach is computationally intensive and depends primarily on emulating the *qualitative* patterns observed in the brain imaging experiments. Statistical estimation of unknown parameters is of secondary importance and these parameters are usually chosen or fixed a priori (usually based on animal studies). HOS and the method proposed in this paper can be classified as the system-level neural modeling paradigm in Horwitz's terminology. HOS differs from the aforementioned methods primarily in its emphasis which is to model the stochastic inter-relationship between the different components of the multiple signals. Their approach treats the brain as an input–output system, which is perturbed by known inputs (i.e., experimental stimulus) in the experiments. The measured responses (i.e., observed fMRI signal) are then used to estimate various parameters that govern the evolution of the activation. The HOS approach: (1) allows modeling relationships among multiple brain areas; (2) separates the signal-of-interest (BOLD) from the measurement noise; (3) models the temporal correlation explicitly by the recent history of the experimental inputs.

This is similar to the dynamic causal model (DCM) recently proposed by Friston et al. (2003). A major difference between the DCM and HOS approach is that a biophysical model called 'Balloon Model' (Buxton et al., 1998) is used to link the hemodynamic response to the underlying neuronal activation in DCM. However, one limitation of the DCM approach is that it assumes a deterministic relationship between brain regions and does not allow for noisy dynamics. HOS approach, however, does not make such restrictive assumption. DCM uses a bilinear expansion to approximate the time/task dependence of effective connectivity (Friston et al., 2003). In HOS, the quantities measuring effective connectivity are time-invariant. There is increasing empirical evidence suggesting that connectivity between different brain areas is dynamic and should be understood as experiment- and time-dependent (Aertsen and Preißl, 1991; Friston, 1994; McIntosh and Gonzalez-Lima, 1994; Büchel and Friston, 1998; McIntosh, 2000). This evidence calls for temporal modeling of effective connectivity, which is the goal of this paper. We propose a Bayesian approach to solve this problem. DCM also uses a Bayesian scheme for estimating model parameters and model selection, which has been implemented in SPM2 (Friston et al., 2003; Penny et al., 2004a). Following the HOS approach, we decompose the observed multiple time series into measurement error and the BOLD signals. The observed fMRI signals at each brain region are modeled as a function of the BOLD signal. To capture the complexities of the dynamic processes in the brain, the region-specific time-varying coefficients in the activation equation are subsequently modeled, as a linear function of the BOLD signals at other brain regions, combined with error. These coefficients measure the connectivity or coupling between the regions of interest. We extend the HOS model to allow the connectivity coefficients to vary in a stochastic manner. For simplicity, we assume that these coefficients change as a random walk process but other processes such as autoregressive and regime switching can be incorporated in a

similar manner. In HOS, the measurement noise is assumed to be independent but this assumption may not be appropriate, especially for fMRI experiments with fast brain scanning rates (but see Penny et al., 2003; Kiebel et al., 2003, for nonsphericity correction implemented in SPM2 to deal with serial correlations in noise through a variational Bayesian scheme). Therefore, we will consider an extension which incorporates an autoregressive structure in the measurement noise as well.

The organization of the paper is as follows. Our proposed model will be presented in Model formulation. Several alternative models are also discussed. Bayesian procedures for parameter estimation and associated computational details are described in Statistical analysis. We employ a variant of the ML-II method (Berger, 1985) for choosing values for the hyperparameters of the final-stage priors. Gibbs sampling is used for generating the posterior distribution for the parameters of interest. Model selection among the competing models considered in Model formulation is based on Bayes factor (see, e.g., Kass and Raftery, 1995; Penny et al., 2004a). This section also includes the results from a Monte Carlo simulation which demonstrates the validity of our proposed method. We further illustrate our technique on an fMRI data set so to investigate the attentional control network in the human brain. The details of the experiment and the data set are described in Application. We conclude our paper in Conclusions and future work.

Model formulation

A typical modeled BOLD response, $x(t)$, usually occurs between 3 and 10 s after the stimulus, $s(t)$, is presented and reaches the peak at about 6 s. The delay of the BOLD response is usually modeled by a hemodynamic response function (HRF), $h(t)$, which weighs the past stimulus values by a convolution:

$$x(t) = \int_0^t h(u)s(t-u)du, \tag{1}$$

where $s(t)$ takes the value of '1' when the stimulus is 'ON' and '0' when the stimulus is 'OFF'. The top panel in Fig. 1 shows a stimulus, $s(t)$, presented periodically in an fMRI experiment. The HRF is usually modeled by Poisson, Gaussian or Gamma density, or by the difference of two Gamma functions which was used in this paper. The second panel in Fig. 1 shows a typical hemodynamic response function and the bottom panel shows what the BOLD signal looks like after the convolution with the periodic stimulus function in the top panel. The magnitude of the BOLD signal, denoted as β, varies over brain regions and experimental conditions, and is usually estimated by the general linear model:

$$y_i(t) = \alpha_i + \beta_i x_i(t) + \varepsilon_i(t), \tag{2}$$

where $y_i(t)$ is the observed fMRI signal or the measured BOLD response (as opposed to the modeled BOLD, $x(t)$) recorded by the MRI scanner, $e_i(t)$ is measurement noise at voxel i (voxel is 3D generalization of pixel). The coefficient, β_i, measures the 'activation' at voxel i in fMRI studies and α_i represents the baseline. Without loss of generality, we assume that the fMRI data are detrended here. Detrending is a common preprocessing step in fMRI experiments and attempts to remove the drift (mainly caused by the MRI scanner) present in the fMRI data. A priori detrending is not necessary but can simplify the computation.

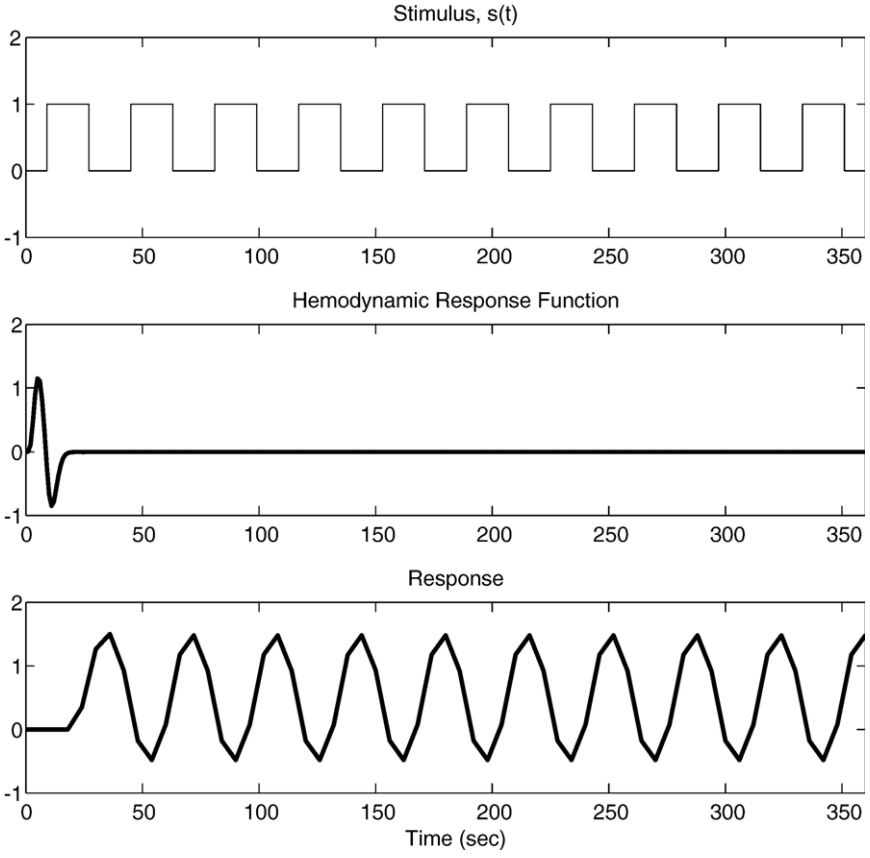

Fig. 1. Convolution of hemodynamic response function.

In model (2), the 'activation', β, is assumed to be time-invariant, which may not be realistic. Many studies report 'learning' effect in fMRI experiments, where strong fMRI activation is mainly detected in the beginning of the experiment but becomes weaker later on (see Milham et al., 2002, 2003a,b, for example). Therefore, it is reasonable to consider the activation to be time-varying as:

$$y_i(t) = \alpha_i + \beta_i(t)x_i(t) + \varepsilon_i(t). \quad (3)$$

The time-varying activation idea was also exploited in Gössl et al. (2001).

There is an alternative formulation where the activation can be expanded in terms of time-varying basis functions.[3] The parameters then become time-invariant coefficients of these basis functions. Mathematically, this is equivalent to the implementation of time-varying activations in SPM2, where the stimulus functions $x_i(t)$ are modulated with one or more time-varying basis functions to give an expanded set of regressors. These regressors are then convolved and form explanatory variables in the general linear model. The most usual forms for these basis functions are mono-exponential decays to model adaptation and learning effects. The key difference between the SPM2 approach and the notion of a time-varying activation that we proposed here is that the former is deterministic whereas, in our model, the activation can change stochastically.

Functionally specialized brain areas do not operate on their own but interact with one another depending on the context. HOS therefore augmented the above model to model the influence of one brain region on another. The model assumes the time-varying 'activation' in Eq. (3) in terms of the history of the error-free fMRI signal from itself and another region as:

$$\beta_i(t) = \gamma_{i,i}Z_i(t-1) + \gamma_{i,j}Z_j(t-1) + w_i(t), \quad (4)$$

where $Z_i(t-1) = x_i(t-1)\beta_i(t-1)$ represents the error-free fMRI signal for the i-th region and at time point t. Hence, the right-hand side of Eq. (4) involves this error-free fMRI signal for the previous time-point $t-1$. Higher order effects of the BOLD history (from $t-2$, $t-3$ and so on) may also be considered (see also Lahaye et al., 2003 for a similar idea in using the signal history in the quantification of large-scale connectivity). In this model, the coefficients $\gamma_{i,i}$ and $\gamma_{i,j}$ measure the influence on region i from itself and region j, respectively. The first coefficient reflects the self-feedback and second coefficient characterizes the coupling relationship between two regions i and j.

In HOS, these connectivity coefficients are assumed to be time-invariant. In this paper, we relax this assumption to accommodate time-dependent connectivity between brain regions and focus on the connectivity varying in a random walk manner (see Eq. (5.3)). Without loss of generality, we discuss our approach for examining effective connectivity among three regions. It will be applied to study attentional control network from an fMRI experiment, described in Application. Our approach is very general and is readily generalized to an arbitrary number of regions and other types of time-varying processes for characterizing connectivity.

In summary, our models have the following three components, which are connected in a hierarchical manner.

A. The first component connects the observed fMRI signals to the (convolved) stimulus at every brain region (see Eq. (3)). The strength of connection is characterized by the time-varying 'activation' parameter, $\beta(t)$.
B. The second component connects the activation parameter, $\beta(t)$, from one region at any time point to the noise-free BOLD signal corresponding to all three regions at the previous time-point (see Eq. (4)). The strength of connection between brain regions is characterized by the effective connectivity parameter, $\gamma_{i,j}$.
C. The effective connectivity parameter, $\gamma_{i,j}$, is modeled as time-varying, which allows a dynamic relation between effective connectivity parameter at any time point and the previous time point. Without loss of generality, we illustrate our method based on three regions of interest below. Connectivity between more regions of interest can be examined in a similar manner.

We introduce the following notation. For $i = 1, 2, 3$; $t = 1,\ldots,T$, let

$y_i(t)$ = observed fMRI signal corresponding to the i-th region at time-point t,
$x_i(t)$ = stimulus, convolved with the hemodynamic response function, for the i-th region and time-point t,
α_i = baseline trend corresponding to the i-th region,
$\beta_i(t)$ = activation coefficient corresponding to the i-th region at time point t,
$\gamma_{i,j}(t)$ = influence of j-th region on the i-th region at time-point t.

In fMRI studies, one often uses the same hemodynamic response function h (see Eq. (1) above) for all voxels and the hemodynamic response to a particular event is taken to be the time-invariant voxel-specific activation times the convolution of h and the stimulus function, which results in voxel-specific (or region-specific) hemodynamic response. However, it is possible to relax this assumption and allow 'heterogeneity' in hemodynamic response function (see, for example, Birn et al., 2001; Friston et al., 1998; Liao et al., 2002). One way is to expand the hemodynamic response function with a small number of temporal basis functions (Friston et al., 1998). This provides a multivariate characterization of the hemodynamic response to a particular event or brain state that is unique to each voxel. Our proposed approach, on the other hand, accommodates voxel-specific hemodynamic response via voxel-specific time-varying activation and it would be redundant to further assume different hemodynamic response function for each voxel, thus we consider only homogenous response function $x_1(t) = x_2(t) = x_3(t)$ in this paper. We denote the common value of $x_1(t)$, $x_2(t)$, and $x_3(t)$, by $x(t)$. The generalization to heterogenous hemodynamic response function is straightforward. Our proposed model is given by

Model M_1

$$\boldsymbol{y}_i(t) = \alpha_i + x(t)\beta_i(t) + \varepsilon_i(t), \quad t = 1,\ldots,T, \quad i = 1,2,3, \quad (5.1)$$

$$\beta_i(t) = x(t-1)\left[\sum_{k=1}^{3}\gamma_{i,k}(t)\beta_k(t-1)\right] + w_i(t), \quad t = 2,\ldots,T,\ i = 1,2,3, \quad (5.2)$$

$$\gamma_{i,j}(t) = \gamma_{i,j}(t-1) + \delta_{i,j}(t), \quad t = 2,\ldots,T,\ i,j = 1,2,3. \quad (5.3)$$

[3] We thank Professor Karl J. Friston for bringing our attention to this alternative formulation.

Notice that Eqs. (5.1)–(5.3) correspond to A–C above. The quantities $\varepsilon_i(t)$, $w_i(t)$, $\delta_{ij}(t)$ are random errors. We assume that

(E1) $\varepsilon_i(t)$, s are independent and identically distributed (i.i.d.) $N(0, \sigma_i^2)$;
(E2) $w_i(t)$, s are i.i.d. $N(0, \sigma_w^2)$;
(E3) $\delta_{i,j}(t)$, s are i.i.d. $N(0, \sigma_\delta^2)$;
(E4) all the errors in (E1–E3) are independent.

In matrix notation, the relations given by Eqs. (5.1)–(5.3) can be expressed as

$$\boldsymbol{y}_t = \boldsymbol{\alpha} + \boldsymbol{A}_t\boldsymbol{\beta}_t + \varepsilon_t, \qquad t = 1,\ldots,T, \tag{6.1}$$

$$\boldsymbol{\beta}_t = \boldsymbol{\Gamma}_t\boldsymbol{A}_{t-1}\boldsymbol{\beta}_{t-1} + \boldsymbol{w}_t, \qquad t = 2,\ldots,T, \tag{6.2}$$

$$\boldsymbol{\Gamma}_t = \boldsymbol{\Gamma}_{t-1} + \boldsymbol{\Delta}_t, \qquad t = 2,\ldots,T, \tag{6.3}$$

where

$$\boldsymbol{y}_t = \begin{pmatrix} y_1(t) \\ y_2(t) \\ y_3(t) \end{pmatrix}, \boldsymbol{A}_t = \begin{pmatrix} x_1(t) & 0 & 0 \\ 0 & x_2(t) & 0 \\ 0 & 0 & x_3(t) \end{pmatrix} = x(t)\boldsymbol{I}_3, \boldsymbol{\alpha} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix},$$
$$\boldsymbol{\beta}_t = \begin{pmatrix} \beta_1(t) \\ \beta_2(t) \\ \beta_3(t) \end{pmatrix}, \boldsymbol{\Gamma}_t = \begin{pmatrix} \gamma_{1,1}(t) & \gamma_{1,2}(t) & \gamma_{1,3}(t) \\ \gamma_{2,1}(t) & \gamma_{2,2}(t) & \gamma_{2,3}(t) \\ \gamma_{3,1}(t) & \gamma_{3,2}(t) & \gamma_{3,3}(t) \end{pmatrix},$$
$$\varepsilon_t = \begin{pmatrix} \varepsilon_1(t) \\ \varepsilon_2(t) \\ \varepsilon_3(t) \end{pmatrix}, \boldsymbol{w}_t = \begin{pmatrix} w_1(t) \\ w_2(t) \\ w_3(t) \end{pmatrix}, \boldsymbol{\Delta}_t = \begin{pmatrix} \delta_{1,1}(t) & \delta_{1,2}(t) & \delta_{1,3}(t) \\ \delta_{2,1}(t) & \delta_{2,2}(t) & \delta_{2,3}(t) \\ \delta_{3,1}(t) & \delta_{3,2}(t) & \delta_{3,3}(t) \end{pmatrix}. \tag{7}$$

Moreover, we make the following assumptions.

(P1) The prior for $\boldsymbol{\beta}_1$ is given by a trivariate normal distribution with mean $\boldsymbol{\beta}_0$ and dispersion matrix $\sigma_\beta^2\boldsymbol{I}_3$, and $\boldsymbol{\beta}_0$ and σ_β^2 are assumed to be known. Write $\boldsymbol{\beta}_0 = (\beta_1(0), \beta_2(0), \beta_3(0))^{\mathrm{T}}$.
(P2) The prior for $\gamma_{i,j}(1)$ is given by a normal distribution with mean $\gamma_{i,j}(0)$ and variance σ_γ^2. These nine normal distributions are assumed to be independent and the associated parameters (10 in number) are also assumed to be known.
(P3) The prior $\boldsymbol{\alpha}$ is given by a trivariate normal with mean $\boldsymbol{\mu} = (\mu_1, \mu_2, \mu_3)^{\mathrm{T}}$, and dispersion matrix $\sigma_\beta^2\boldsymbol{I}_3$, and $\boldsymbol{\mu}$ and σ_α^2 are assumed to be known.
(P4) The priors for σ_1^2, σ_2^2, σ_3^2, σ_w^2, σ_δ^2 are assumed to be independent and follow Inverse Gamma distribution (denoted as IG(a, d)). The IG distribution means that the inverse of each of the variance terms is distributed as two-parameter Gamma distribution with parameters a and d, which control the shape and scale of the distribution.[4] In our analyses, we set $a = d = 0$ which results in non-informative prior. Given that we are ignorant of these parameters, noninformative prior is a natural choice. Interested readers can refer to O'Hagan and Forster (2004, p. 306) for more details on the definition of IG distribution.
(P5) All the distributions given by (P1)–(P4) above are independent.

The set of relations given by Eqs. (5.1)–(5.3), and the assumptions (E1)–(E4), (P1)–(P5) constitute M_1.

[4] The probability density function for an IG random variable, x, has the form, $f(x) \propto (1/x)^{((d+2)/2)} \exp(-a/2x)$ where a is the scale parameter and d is the shape parameter.

Model M_2

In model M_1, for every region i, the measurement errors $\varepsilon_i(1),\ldots,\varepsilon_i(T)$, are assumed to be independent. This assumption may not be realistic especially for fMRI data collected at a fast image acquisition rate. In typical fMRI activation analysis, the measurement noise is usually assumed to follow an autoregressive process (see, for example, Penny et al., 2003; Woolrich et al., 2001, 2004b; Worsley, 2003). That is, there may still be temporal dependence in the background process which cannot be accounted for by models (5.1)–(5.3). To address this issue, we consider another extension and introduce the temporal dependence in $\varepsilon_i(t)$. For illustration, we assume that for every region i, the $\varepsilon_i(t)$ follows an autoregressive process of order 1 with an *unknown* parameter $\rho_i, |\rho_i| < 1$. Higher order autoregressive process can be incorporated in a similar manner. In other words, we replace the assumption (E1) by the following:

(E5) for every i, the following hold: $\varepsilon_i(1)$, $\varepsilon_i(2) - \rho_i\varepsilon_i(1),\ldots,\varepsilon_i(T) - \rho_i\varepsilon_i(T-1)$ are independent, $\varepsilon_i(1) \sim N(0, \sigma_i^2/(1-\rho_i^2))$, $\varepsilon_i(2) - \rho_i\varepsilon_i(1),\ldots,\varepsilon_i(T) - \rho_i\varepsilon_i(T-1)$ are i.i.d. $N(0, \sigma_i^2)$, and moreover, $\varepsilon \stackrel{\text{def}}{=} (\varepsilon_i(1), \varepsilon_i(2) - \rho_i\varepsilon_i(1),\ldots,\varepsilon_i(T) - \rho_i\varepsilon_i(T-1))^{\mathrm{T}}, i = 1, 2, 3$ are assumed to be independent,
(E6) all the distributions in (E2), (E3) given (E5) are independent.

In the absence of prior knowledge on how ρ_i varies over i, we decide to choose noninformative prior for ρ_i in our analysis. The assumption (P5) of model M_1 is, therefore, replaced by the following:

(P6) the (noninformative) prior for ρ_i is given by a uniform distribution over $(-1, 1)$,
(P7) all the distributions given by P1–P4 and (P6) above are independent.

The set of relations given by Eqs. (5.1)–(5.3), and the assumptions , (E5) and (E6), and (P1) (P2) (P3) (P4), (P6) and (P7) constitute M_2.

Model M_3

Notice that in Eq. (5.2), we assume that the influence of the j-th region on the i-th region to be given not by $\beta_j(t-1)$ but rather by $x(t-1)\beta_j(t-1)$: An alternative way to explore the time-varying effective connectivity, as in time-varying parameter regression models, is to use $\beta_i(t-1)$ only on the right hand side of Eq. (4) which leads to:

$$y_i(t) = \alpha_i + x(t)\beta_i(t) + \varepsilon_i(t), \qquad t = 1,\ldots,T,\ i = 1, 2, 3, \tag{8.1}$$

$$\beta_i(t) = \sum_{k=1}^{3} \gamma_{i,k}(t)\beta_k(t-1) + w_i(t), \quad t = 2,\ldots,T,\ i = 1, 2, 3, \tag{8.2}$$

$$\gamma_{i,j}(t) = \gamma_{i,j}(t-1) + \delta_{i,j}(t), \qquad t = 2,\ldots,T,\ i,j = 1, 2, 3. \tag{8.3}$$

The set of relations given by Eqs. (8.1)–(8.3), and the assumptions (E1)–(E4), (P1)–(P5) constitute M_3.

This model can indeed be thought of a dynamic version of multivariate time-varying parameter regression, which is related to dynamic hierarchical models considered in Gamerman and Migon (1993) and in West and Harrison (1999).

Model M_4

This model is similar to M_3 but autoregressive error structure is incorporated in the measurement noise (as in M_2). The set of relations given by Eqs. (8.1)–(8.3), and the assumptions (E2), (E3), (E5) and (E6), and (P1)–(P4), (P6) and (P7) constitute M_4.

Statistical analysis

We propose a Bayesian approach to estimate the four models described in the previous section. Recall that the (measurement) errors in Eq. (5.1) are assumed to have different variances. In our preliminary empirical analysis, we did not find compelling evidences supporting the fact that these variances are different in our data.[5] Therefore, we present our method under the additional assumption that $\sigma_1 = \sigma_2 = \sigma_3 = \sigma_\varepsilon$ in Eq. (5.1) here. Under this additional condition, assumptions (E1), (P4) and (E5) (and consequently (E4), (P5) and (P7)) are then modified accordingly. To save space, we do not mention these modifications separately. It must, however, be stressed that our model is flexible enough to accommodate heterogenous σ_is.

As we demonstrate in Application section, model M_1 with homogenous measurement error variance assumption fits the data the best among the four candidate models. All our results and computational details given in the present section refer to this model. Our interest focuses on the connectivity parameters $\{\gamma_{i,j}(t),\ i,\ j\ =\ 1,\ 2,\ 3\}$. The steps to construct the posterior distribution for this set of parameters using a Gibbs sampler are described below. To summarize the posterior inference, posterior mean and posterior standard deviation of the $\gamma_{i,j}(t)$s, and the corresponding highest posterior density Bayesian credible region (also known as HPD BCR) were obtained. These regions were constructed by solving a nonlinear equation (see, e.g., Carlin and Louis, 2000, pp. 35–38) and they contain, by definition, the posterior modes with a chosen posterior probability $(1\ -\ \alpha)$. We use $\alpha\ =\ 0.05$ for the present work.

Notice that the HPD BCR is not the same as the confidence interval and should not be interpreted or used in the same way as in frequentist paradigm for hypothesis testing. The HPD BCR tells us the probability of a parameter falling in a specific range given the observed data. Even though the HPD BCR of $\gamma_{i,j}(t)$ includes zero, this may not imply that the corresponding $\gamma_{i,j}(t)$ is statistically insignificant. This can be seen clearly in our simulation study reported later in Simulation.

It should be emphasized that the Gibbs sampler scheme (Gelfand and Smith, 1990) we used assumes that the following final-stage hyperparameters are known:

$$\{\beta_0(i){:}i\ =\ 1{,}2{,}3\},\{\mu_i{:}i\ =\ 1{,}2{,}3\},\{\gamma_{i,j}(0)\ {:}i{,}j\ =\ 1{,}2{,}3\},$$
$$\{\sigma_\beta{,}\sigma_\gamma{,}\sigma_\alpha\},a{,}d.$$

[5] Before fitting our fMRI data to the four models, we checked the homogeneity of variance assumption on the time series from these three regions by Barlett's test (Bartlett, 1937).

These hyperparameters are usually unknown in practice as in our case. We applied a method similar to the empirical Bayes technique (Berger, 1985) for selecting appropriate values of these parameters.

Posterior distributions

In this section, we describe the computational details for obtaining the required posterior distributions and associated statistics (posterior mean, posterior variance and Bayesian credible regions), and the initial-stage hyperparameters empirically. The notation '$Y\,|\,\cdot$' below refers to the conditional distribution of a random variable Y given another variables and/or parameters. Also, $\boldsymbol{X} = ((x_{ij})) \sim N_9(\boldsymbol{A},\ \tau^2\boldsymbol{I}_{81})$, means that x_{ij}s are i.i.d. $N(a_{ij},\ \tau^2)$, where $\boldsymbol{X} = ((x_{ij}))$ is a 9×9 random matrix and $\boldsymbol{A} = ((a_{ij}))$ is a 9×9 matrix of real entries. Introducing the notation $\boldsymbol{\sigma} = (\sigma_\varepsilon^2,\ \sigma_w^2,\ \sigma_\delta^2)$, our model can now be rewritten in the form of probability distributions:

$$\begin{aligned}
&\boldsymbol{y}_T\,|\,\boldsymbol{y}_1,\ldots,\boldsymbol{y}_{T-1},\boldsymbol{\alpha},\boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_T,\boldsymbol{\Gamma}_1,\ldots,\boldsymbol{\Gamma}_T,\sigma &&\sim N_3\left(\boldsymbol{\alpha}+x_T\boldsymbol{T},\sigma_\varepsilon^2 I_3\right),\\
&\boldsymbol{\beta}_T\,|\,\boldsymbol{y}_1,\ldots,\boldsymbol{y}_{T-1},\boldsymbol{\alpha},\boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_{T-1},\boldsymbol{\Gamma}_1,\ldots,\boldsymbol{\Gamma}_T,\sigma &&\sim N_3\left(x_{T-1}\boldsymbol{\Gamma}_T\boldsymbol{\beta}_{T-1},\sigma_w^2 I_3\right),\\
&\boldsymbol{\Gamma}_T\,|\,\boldsymbol{y}_1,\ldots,\boldsymbol{y}_{T-1},\boldsymbol{\alpha},\boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_{T-1},\boldsymbol{\Gamma}_1,\ldots,\boldsymbol{\Gamma}_{T-1},\sigma &&\sim N_9\left(\boldsymbol{\Gamma}_{T-1},\sigma_\delta^2 I_{81}\right),\\
&\qquad\vdots &&\qquad\vdots\\
&\boldsymbol{y}_2\,|\,\boldsymbol{y}_1,\boldsymbol{\alpha},\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\boldsymbol{\Gamma}_1,\boldsymbol{\Gamma}_2,\sigma &&\sim N_3\left(\boldsymbol{\alpha}+x_2\boldsymbol{\beta}_2,\sigma_\varepsilon^2 I_3\right),\\
&\boldsymbol{\beta}_2\,|\,\boldsymbol{y}_1,\boldsymbol{\alpha},\boldsymbol{\beta}_1,\boldsymbol{\Gamma}_1,\boldsymbol{\Gamma}_2,\sigma &&\sim N_3\left(x_1\boldsymbol{\Gamma}_2\boldsymbol{\beta}_1,\sigma_w^2 I_3\right),\\
&\boldsymbol{\Gamma}_2\,|\,\boldsymbol{y}_1,\boldsymbol{\alpha},\boldsymbol{\beta}_1,\boldsymbol{\Gamma}_1,\sigma &&\sim N_9\left(\boldsymbol{\Gamma}_1,\sigma_\delta^2,I_{81}\right),\\
&\boldsymbol{y}_1\,|\,\boldsymbol{\alpha},\boldsymbol{\beta}_1,\boldsymbol{\Gamma}_1,\sigma &&\sim N_3\left(\boldsymbol{\alpha}+x_1\boldsymbol{\beta}_1,\sigma_\varepsilon^2 I_3\right),\\
&\boldsymbol{\beta}_1\,|\,\boldsymbol{\alpha},\boldsymbol{\Gamma}_1,\sigma &&\sim N_3\left(\boldsymbol{\beta}_0,\sigma_\beta^2,I_3\right),\\
&\boldsymbol{\alpha}\,|\,\boldsymbol{\Gamma}_1,\sigma &&\sim N_3\left(\boldsymbol{\mu}_0,\sigma_\alpha^2,I_3\right),\\
&\boldsymbol{\Gamma}_1\,|\,\sigma &&\sim N_9\left(\boldsymbol{\beta}_0,\sigma_\gamma^2,I_3\right),\\
&\sigma_\varepsilon^2,\sigma_w^2,\sigma_\delta^2 &&\sim \text{ i.i.d. } IG(a,d).
\end{aligned} \tag{9}$$

To save space, we further denote

$$\psi = (\boldsymbol{\alpha},\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\ldots,\boldsymbol{\beta}_T,\boldsymbol{\Gamma}_1,\boldsymbol{\Gamma}_2,\ldots,\boldsymbol{\Gamma}_T) \equiv (\alpha_1,\alpha_2,\alpha_3,\beta_1(t),\beta_2(t),\beta_3(t),t = 1{,}2,\ldots,T;\gamma_{i,j}(t),i{,}j\ =\ 1{,}2{,}3{,}t\ =\ 1,\ldots,T),$$

and

$$\boldsymbol{\theta}\ =\ (\psi,\sigma).$$

Notice now that the joint posterior distribution of

$$\boldsymbol{\theta} \equiv \left(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\ldots,\boldsymbol{\beta}_T,\boldsymbol{\Gamma}_1,\boldsymbol{\Gamma}_2,\ldots,\boldsymbol{\Gamma}_T,\boldsymbol{\alpha},\sigma_\varepsilon^2,\sigma_w^2,\sigma_\delta^2\right)$$

is *proportional* to the following:

$$\prod = \exp\left(-\frac{Q_1+a}{2\sigma_\varepsilon^2}-\frac{Q_2+a}{2\sigma_w^2}-\frac{Q_3+a}{2\sigma_\delta^2}-\frac{Q_4}{2\sigma_\beta^2}-\frac{Q_5}{2\sigma_\beta^2}-\frac{Q_6}{2\sigma_\gamma^2}\right) \times\left[\left(\frac{1}{\sigma_\varepsilon^2}\right)^{\frac{d+2}{1}+\frac{3T}{2}}\cdot\left(\frac{1}{\sigma_w^2}\right)^{\frac{d+2}{2}+\frac{3(T-1)}{2}}\cdot\left(\frac{1}{\sigma_\delta^2}\right)^{\frac{d+2}{2}+\frac{9(T-1)}{2}}\right], \tag{10}$$

where (writing x_t for $x(t)$)

$$\begin{aligned}
&Q_1 \overset{\text{def}}{=} \sum_{i=1}^{3}\sum_{t=1}^{T}(\boldsymbol{y}_i(t)-\alpha_i-x_t\beta_i(t))^2,\\
&Q_2 \overset{\text{def}}{=} \sum_{i=1}^{3}\sum_{t=2}^{T}\left(\beta_i(t)-x_{t-1}\left[\sum_{k=1}^{3}\gamma_{ik}\beta_k(t-1)\right]\right)^2,\\
&Q_3 \overset{\text{def}}{=} \sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{t=2}^{T}\left(\gamma_{i,j}(t)-\gamma_{i,j}(t-1)\right)^2,\ Q_4 \overset{\text{def}}{=} \sum_{i=1}^{3}(\beta_i(1)-\beta_i(0))^2,\\
&Q_5 \overset{\text{def}}{=} \sum_{i=1}^{3}(\alpha_i-\mu_i)^2, \qquad Q_6 \overset{\text{def}}{=} \sum_{i=1}^{3}\sum_{i=1}^{3}\left(\gamma_{i,j}(1)-\gamma_{i,j}(0)\right)^2.
\end{aligned} \tag{11}$$

The details leading to the expression in Eq. (10) are provided in the Appendix A.1.

Our interests focus on the posterior distribution of the connectivity parameters, $\gamma_{i,j}(t)$s, which were obtained by the Gibbs sampling scheme. In the Appendix A.2, we list all the full conditionals necessary for implementation of the Gibbs sampling scheme.

Choice of hyperparameters

Our computation above assumes known values of the final-stage hyperparameters. In practice, these quantities are usually unknown and can be obtained by a slight modification of the ML-II method proposed by Berger (1985, pp. 99–101). The details of the computation are presented below.

We denote the hyperparameters collectively as:

$$\boldsymbol{\xi} \stackrel{\text{def}}{=} (\boldsymbol{\xi}_{\psi},\boldsymbol{\xi}_{\sigma}),$$

where

$$\boldsymbol{\xi}_{\psi} \stackrel{\text{def}}{=} (\beta_0(i), i = 1, 2, 3;\ \mu_i, i = 1, 2, 3;\ \gamma_{i,j}(0), i, j = 1, 2, 3;\ \sigma_{\beta},\sigma_{\gamma},\sigma_{\alpha}) \text{ and } \boldsymbol{\xi}_{\boldsymbol{\sigma}} \stackrel{\text{def}}{=} (a,d).$$

To estimate these hyperparameters, one can maximize the marginal density of observed $\boldsymbol{y} \stackrel{\text{def}}{=} (\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_T)$, given by

$$p(\boldsymbol{Y} \mid \boldsymbol{\xi}) = \int p(\boldsymbol{Y} \mid \boldsymbol{\theta}) \pi(\boldsymbol{\theta} \mid \boldsymbol{\xi}) d\boldsymbol{\theta}, \tag{12}$$

with respect to $\boldsymbol{\xi}$ (this expression can also be regarded as the likelihood function for the hyperparameters given the data). This approach to setting hyperparameters is called ML-II (Berger, 1985). Note that the marginal density $p(\cdot \mid)$ will not be proper if the prior distribution $\pi(\boldsymbol{\theta} \mid \boldsymbol{\xi})$ is improper. Despite this, maximization of Eq. (l2) is possible. Thus

$$\hat{\boldsymbol{\xi}} = \arg\max_{\xi} p(\boldsymbol{Y} \mid \boldsymbol{\xi})$$

may be chosen as the set of appropriate hyperparameters. The approach we employ for maximization is new from a computational point of view. The rationales of our method are provided in the Appendix C. The details of our method are now described.

Notice that we can re-write Eq. (12) as

$$p(\boldsymbol{Y} \mid \boldsymbol{\xi}) = \int \left[\int p(\boldsymbol{Y} \mid \psi, \boldsymbol{\sigma}) \pi(\psi \mid \boldsymbol{\xi}_{\psi}) d\psi \right] \pi(\boldsymbol{\sigma} \mid \boldsymbol{\xi}_{\boldsymbol{r}}) d\boldsymbol{\sigma} \tag{13}$$

Instead of working with Eq. (13) (or with Eq. (12)), we work with an analogue of it. The analogue of Eq.(13) (or of Eq. (12)) we propose is given by

$$\hat{p}(\boldsymbol{Y} \mid \boldsymbol{\xi}_{\psi}) = \int p(\boldsymbol{Y} \mid \psi, \hat{\boldsymbol{\sigma}}) \pi(\psi \mid \boldsymbol{\xi}_{\boldsymbol{w}}) d\psi, \tag{14}$$

where $\boldsymbol{\sigma}$ is the posterior mode of $\boldsymbol{\sigma}$, i.e., $\boldsymbol{\sigma}$ maximizes the function

$$\int p(\boldsymbol{Y} \mid \psi, \boldsymbol{\sigma}) \pi(\psi \mid \boldsymbol{\xi}_{\psi}) d\psi, \tag{15}$$

over $\boldsymbol{\sigma}$. A heuristic justification for this step can be offered by noting that the integral in Eq. (13) is bounded above by that in Eq. (14) if $\pi(\boldsymbol{\sigma} \mid \boldsymbol{\xi}_{\sigma})$ is proper. A more detailed explanation of choosing $\boldsymbol{\sigma}$ as an estimate of σ can be found in the Appendix C.

The next step involves the maximization of the integrated likelihood (14) with respect to ψ. We denote the maxima of this likelihood function as $\hat{\mathbf{p}}(\boldsymbol{Y} \mid \hat{\boldsymbol{\xi}}_{\psi})$ which is attained at $\hat{\boldsymbol{\xi}}_{\psi}$.

Model selection

We now discuss how to pick the best fit model out of a set of competing models via Bayes factor (Clyde and George, 2004; Kass and Raftery, 1995; Penny et al., 2004a). It must, however, be stressed that what we computed and employed for the purpose of model selection is not exactly Bayes factor but a variant of it. This is because Bayes factor is a ratio of integrated likelihoods but we worked with the ratio of some analogues of these likelihoods. In what follows, the expression 'Bayes factor' refers to this modified ratio.

Given the values of the hyperparameters from model j and model k, calculation of Bayes factor requires only the computation of the marginals, $\hat{\mathbf{p}}_{M_j}(\boldsymbol{Y} \mid \hat{\boldsymbol{\xi}}_{\psi})$, or equivalently the logarithms of the marginals, to obtain the ratios

$$\mathrm{BF}(j,k) = \frac{\hat{p}_{M_j}(\boldsymbol{Y} \mid \hat{\boldsymbol{\xi}}_{\psi})}{\hat{p}_{M_k}(\boldsymbol{Y} \mid \hat{\boldsymbol{\xi}}_{\psi})},$$

for $j \neq k$. Note that estimates of the hyperparameters are obtained separately and independently for each model M_j. Following Jeffreys (1961), a model M_j is said to be significantly better than model M_k if $\mathrm{BF}(j,k) > 3$. Thus, by a series of pairwise comparisons, we can pick the 'best model' from a set of competing models.

Model implementation by Gibbs sampler

Bayesian inference of the time-varying connectivity parameters requires the computation of the posterior distribution $\pi_{M_1}(\boldsymbol{\theta} \mid \boldsymbol{Y}, \hat{\boldsymbol{\xi}}_{\psi}, \hat{\boldsymbol{\xi}}_{\boldsymbol{\sigma}})$, which is analytically nontractable but can be approximated numerically by Gibbs sampling. The algorithm proceeds iteratively by first starting from an arbitrary choice of initial values from the parameter space and then simulating realizations from the 'full' conditional distribution of each parameter given the data and current values of other parameters. The full conditional distribution of each parameter is given in Appendix A.2. After sufficiently large number of iterations, also known as the burn-in time, the algorithm is said to be 'converged' to the true posterior distribution of the parameters.

In our analysis, arbitrary initial values were used for a trial run of the Gibbs sampler but were then modified. In particular, we used posterior estimates of the parameters obtained from the trial run as initial values for the final run of the Gibbs sampler. In the final run, we discarded the realizations obtained in the first 5000 burn-in iterations and retained the next 5000 realizations as samples from the posterior. Indeed, informal convergence checks indicated that convergence reached in much fewer iterations (see Gelman and Rubin, 1992; Raftery and Lewis, 1992, for more discussion on convergence diagnostics). We experimented a few different initial values of the parameters but results stayed the same.

Essentially the same Gibbs sampling scheme was used for implementing all four models. Because of the assumption on temporal dependence between the $\varepsilon_i(t)$s in models M_2 and M_4, minor modification is needed when implementing these two models. The full conditional distribution for each ρ_i within posterior is needed. The corresponding density function can be shown to be the product of a normal density function, truncated to $(-1, 1)$ and another function which is bounded by one. The details are described in the Appendix D.

All of our computations were carried out on a 650 MHz Pentium III machine with 64 MB memory and 20 GB diskspace.

Computations for each model run required about 3 min. Our programs were written in C language.

Simulation

To verify our proposed Bayesian procedure, we generated time series data according to M_1 (see Eqs. (5.1)–(5.3)). The length of each time series was 285 and the values of the hemodynamic response function, $x(t)$ were the same as those used in the Application section (Application). Models M_1 to M_4 were then fitted to the generated data. A total of 50 Monte Carlo simulations were run. The goal is to examine if our proposed method can correctly identify the true model and can capture the true parameter values.

To assess the performance of Bayes factor in identifying the correct model (i.e., M_1), it is required to compute the logarithm of the marginal of the simulated data $\mathbf{Y}$ under all four models $M_1 - M_4$ and examine which of the four models yields the maximum marginal. The logarithm of marginals under both models M_3 and M_4 turned out to be $-\infty$ for all the simulated data sets, correctly indicating their inappropriateness. Over 60% cases of our simulated data sets, the marginal under M_1 turned out to be higher than the marginal under M_2 (i.e., M_1 fits the data better than M_2). This demonstrates the capability of Bayes factor in selecting the true model correctly.

Given that the correct model can be selected using Bayes factor, we now turn to assess the performance of our proposed parameter estimation procedure. Because the results across 50 simulations are very similar, we chose the results from one simulation for the discussion below. We focus on the results for the connectivity parameters $\gamma_{i,j}(t)$. Fig. 2 shows the true value (solid line) and the 95% HPD BCR (dashed lines) for these connectivity parameters. Nearly all true values of $\gamma_{i,j}(t)$ are included within the respective 95% HPD BCR. Notice that the true values were chosen to vary over time and fluctuate around zero. As mentioned above, HPD BCR cannot be used in the same way as confidence interval in frequentist paradigm for hypothesis testing. In other words, the HPD BCR includes zeroes in it at all time points but it may not imply that the $\gamma_{i,j}(t)$ is insignificant.

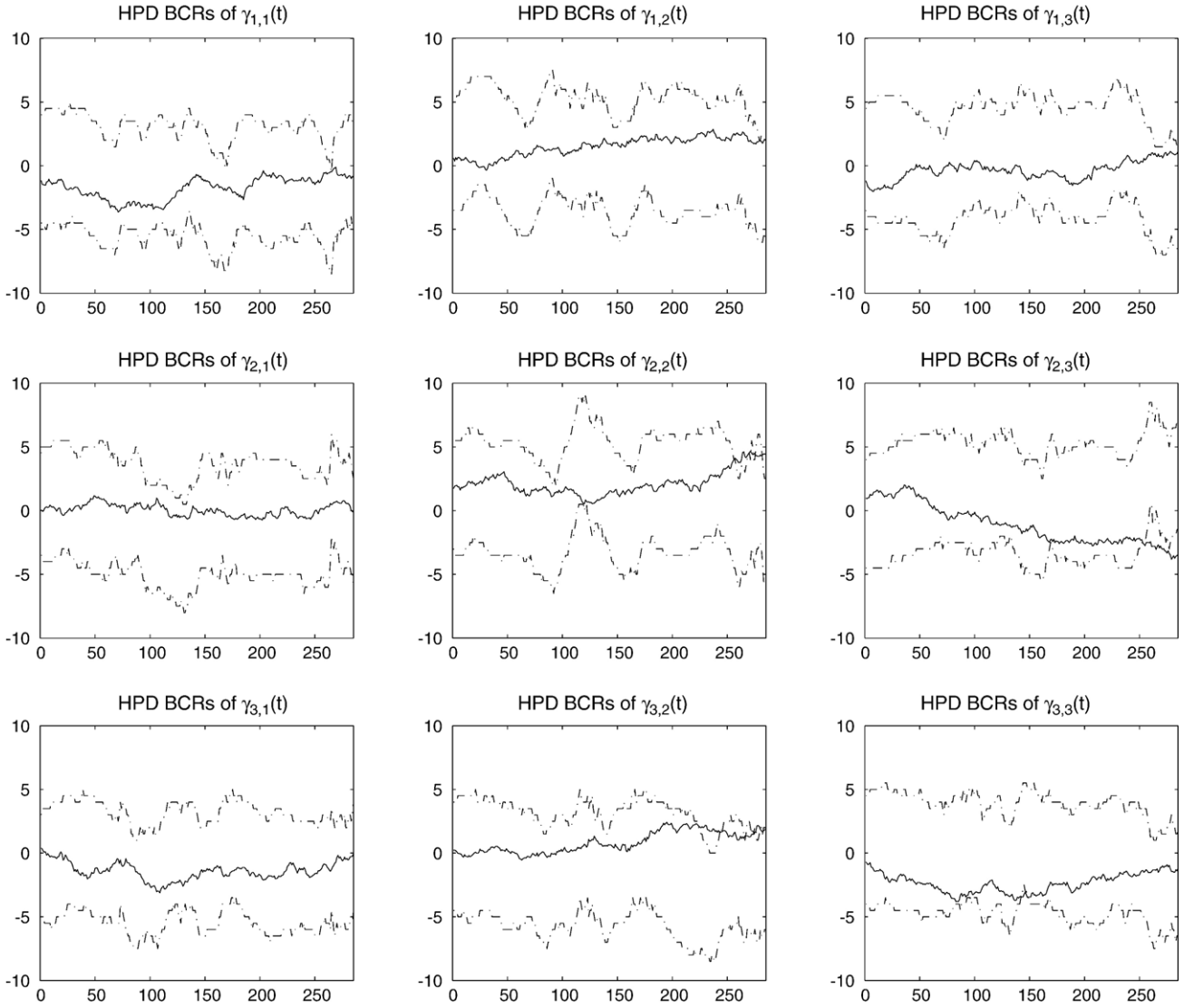

Fig. 2. Highest posterior density Bayesian credible region's of the $\gamma_{i,j}(t)$, s for model Msub 1 from Monte Carlo simulation.

In sum, the above results demonstrate the validity of our proposed Bayesian procedure for modeling dynamic effective connectivity.

Application

In this section, we present an empirical application of the proposed approach for investigating the mechanism of attentional control using fMRI data from a single subject.

Attentional control network

The human brain has limited processing capacity and it is important to have mechanisms to filter out task-irrelevant information and select task-relevant information. Attention is the cognitive function underlying the human brain that can discriminate between relevant and irrelevant information. In a recent review, Frith (2001) argued that there are two types of selection processes in the brain, bottom-up and top-down. Bottom-up selection is driven by the intrinsic properties of a stimulus. Conversely, top-down selection favors the task-relevant feature(s) of the stimulus, independent of its intrinsic properties. Such top-down selection bias requires coordination of neural activity within the attentional network and is usually referred to as the attentional control. Implementation of the attentional control has been suggested to involve (at least) three systems (e.g., Banich et al., 2000): (1) a system processing task-relevant stimulus dimensions (task-relevant processing system); (2) a system processing task-irrelevant stimulus dimensions (task-irrelevant processing system); and (3) a higher order executive control system (source of control) performing the top-down selection bias that may increase the neural activity within the task-relevant processing system and/or may suppress the neural activity within the task-irrelevant processing system. Many studies have found the dorsal lateral prefrontal cortex to be a main source of the attentional control. Depending on the types of the task (visual, auditory, etc.), the sites of the attentional control (task-relevant and task-irrelevant processing systems) might vary.

Experimental design

There were two phases in the experiment. In the learning phase, the subject learned to associate each of three unfamiliar shapes with a unique color word ('Blue', 'Yellow' and 'Green'), until they were able to name the three shapes with 100% accuracy before the test phase started. In the test phase, two types of trials were presented. In the *interference* trials, the shape was printed in an ink color incongruent with the color used to name the shape whereas, in the *neutral* trials, the shape was printed in white, which was not a color name for any of the shapes. A block design was used, in which a block of neutral trials was alternated with a block of interference trials. A total of 6 interference and 6 neutral blocks were presented, with each block consisting of 18 trials, presented at a rate of one trial every 2 s. Each trial consisted of a 300 ms fixation cross by a 1200 ms presentation of the stimulus (shape) and 500 ms inter-trial interval. The subject was instructed to subvocally name each shape with the corresponding color from the learning phase, while ignoring the ink color in which the shape was presented.

Data acquisition and preprocessing

A GE Signa (1.5 T) magnetic resonance imaging system equipped for echo-planar imaging (EPI) was used for data acquisition. For each run, a total of 300 EPI images were acquired (TR = 1517 ms, TE = 40 ms, flip angle = 90°), each consisting of 15 contiguous slices (thickness = 7 mm, in-plan resolution = 3.75 mm, parallel to the AC–PC line). A high-resolution 3D anatomical set (T1-weighted 3-dimensional spoiled gradient echo images) was also collected. The head coil was fitted with a bite bar to minimize head motion during the session. Stimuli were presented on a goggle system. Interested readers can find more details about the experiment in Milham et al. (2003a). The first seven volumes of the images were discarded to allow the MR signal to reach steady state.

Identification of region of interests

For illustration, three regions were selected to investigate the attentional control network in the Stroop task. They were the lingual gyrus, the middle occipital gyrus, and the dorsolateral prefrontal cortex. The lingual gyrus (LG) is a visual area sensitive to color information (Corbetta et al., 1991) representing a *site for processing task-irrelevant information* (i.e., the ink color) in the present experiment (Kelley et al., 1998). The middle occipital gyrus (MOG) is another visual area sensitive to shape information and represents *a site for processing task-relevant information* (the shape's form). The dorsolateral prefrontal cortex (DLPFC) is selected to represent the *source of attentional control*. These areas were also found to be significantly activated in the interference trials comparing to neutral trials in this experiment (see Milham et al., 2003a for more details). For each of these regions, the location of peak activation was identified. A sphere (radius 2 voxels or 4 mm; total number of voxels in sphere is 33) with the peak at its center was defined. The time series of these selected voxels were then subjected to Singular Value Decomposition. The first mode was used to represent the time series response for the selected region. There were also other regions (e.g., anterior cingulate gyrus) but with much weaker activation (P = 0.01), we therefore focus only on the three chosen areas in this paper to demonstrate our approach. We will explore the role of other activated regions in the attentional control network in the future.

Fig. 3 shows the time series of these three selected regions after being detrended by running-line smoother (Marchini and Ripley, 2000). A running-line smoother is a linear regression fitted to the k-nearest neighbors of a given point and used to predict the response at that point. For an fMRI experiment with periodic design (e.g., a block of experimental stimuli and a block of control stimuli are alternatively presented to the subjects), Marchini and Ripley (2000) suggested setting k equal to at least twice the cycle length (1 cycle = 1 block of experimental stimuli + 1 block of control stimuli as in the previous example).

Statistical analysis

We denote the detrended fMRI time series corresponding to LG, MOG and DLPFC by $y_1(t)$, $y_2(t)$ and $y_3(t)$, respectively. Our approach is very flexible and allows us to explore many

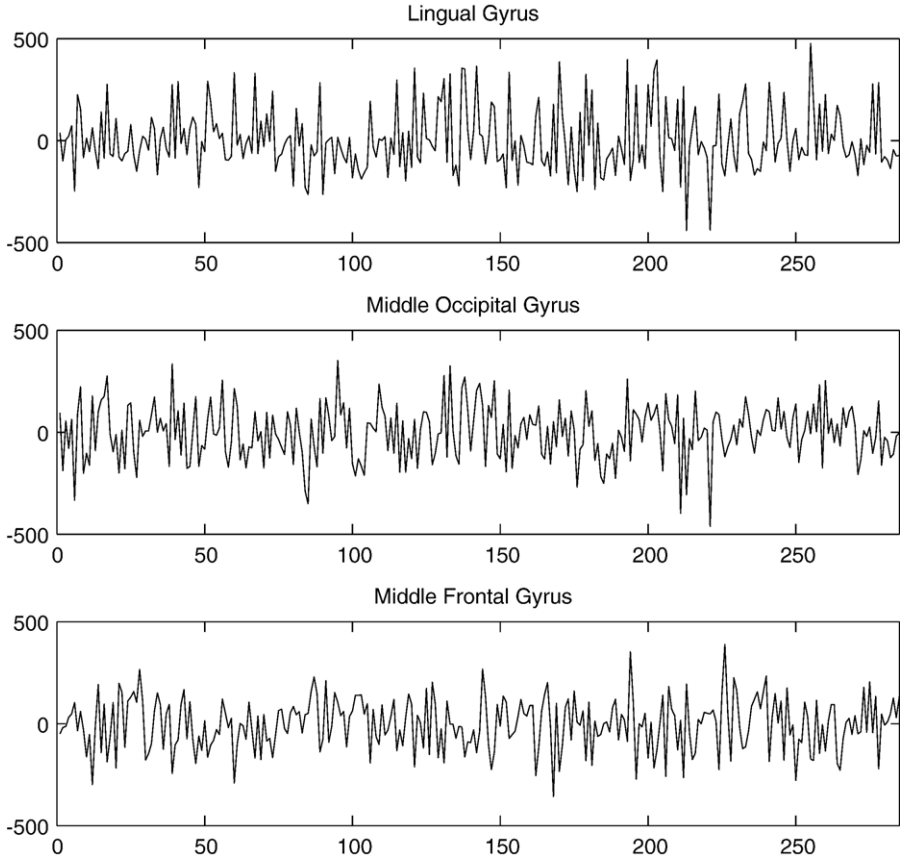

Fig. 3. Three detrended fMRI time series.

possible alternative mechanisms of attentional control, for example, if the DLPFC suppresses the activation of the LG ($\gamma_{1,3}(t)$), and facilitates the activation of the MOG ($\gamma_{2,3}(t)$); if there is reciprocal suppression between LG and MOG, ($\gamma_{1,2}(t)$ and $\gamma_{2,1}(t)$), (as our brain's processing capacity is limited, the LG and the MOG may need to compete for the limited 'resources' in the brain); or if there is feedback from the LG and the MOG on the DLPFC ($\gamma_{3,1}(t)$ and $\gamma_{3,2}$ (t)). These mechanisms can be explored separately or simultaneously. We examined whether the LG, the MOG and the DLPFC reciprocally influence each other simultaneously in models M_1 to M_4.

The values of BF(j, k) were calculated between each pair of candidate models based on the fMRI data. We have calculated (logarithm of) marginal density, denoted $\hat{\mathbf{p}}_{M_j}(\boldsymbol{Y}\,|\,\hat{\boldsymbol{\xi}}_{\psi})$ for each model M_j. The Bayes factors can be obtained in two steps: (i) by computing appropriate differences between these numbers and (ii) by exponentiating these differences (cf. Eq. (15) above). Table 1 shows the values of logarithm of $\hat{\mathbf{p}}_{M_j}(\boldsymbol{Y}\,|\,\hat{\boldsymbol{\xi}}_{\psi})$. As in Eq. (14), the value of $\boldsymbol{\sigma}$ for each of the four models was fixed in the calculation of BF.

Table 1
Logarithm of the approximate marginal density $\hat{\mathbf{p}}_{M_j}(\boldsymbol{Y}\,|\,\hat{\boldsymbol{\xi}})$

Model	Logarithm of $\hat{\mathbf{p}}_{M_j}(\boldsymbol{Y}\,|\,\hat{\boldsymbol{\xi}}_{\psi})$
M_1	-5.0784×10^{199}
M_2	-7.7936×10^{269}
M_3	$-\infty$
M_4	$-\infty$
M_5	-4.1350×10^{177}

Notice that the marginal density of the observed data is very small for all four models. It is not surprising since the data are of very high dimensionality (285 time points × 3 regions). The above table clearly shows that M_1 is the best fit model among the four. The posterior mode (solid line) for $\gamma_{i,j}(t)$, $i, j = 1, 2, 3$ with the corresponding 95% HPD BCR (dash lines) for M_1 is shown in Fig. 4. The mean and variance of the posterior modes over time are shown in Table 2.

Notice that, in Table 2, the variances of the posterior modes for $\gamma_{3,1}(t)$ and $\gamma_{3,2}(t)$ are very small which suggest that the corresponding connectivity may not vary much over time and may regard as constant. Moreover, the averages of the posterior modes for these two coefficients are almost zero throughout the experiment. Together, it suggests we may consider to simplify M_1 by constraining these two coefficients to be zero.

This 'constrained' model (model M_5) was fitted and the Bayes factor between M_5 and M_1, BF(5,1), is much larger than 3 (see Table 1). This means M_5 fits the data significantly better than model M_1.[6]

[6] We did not perform an exhaustive research for all possible models (more than 500) which is computationally very intensive. Instead, we rely on the mean and variance of posterior modes to guide us for model modification. We ran another simulation to verify this procedure and the results supported this approach for model modification.

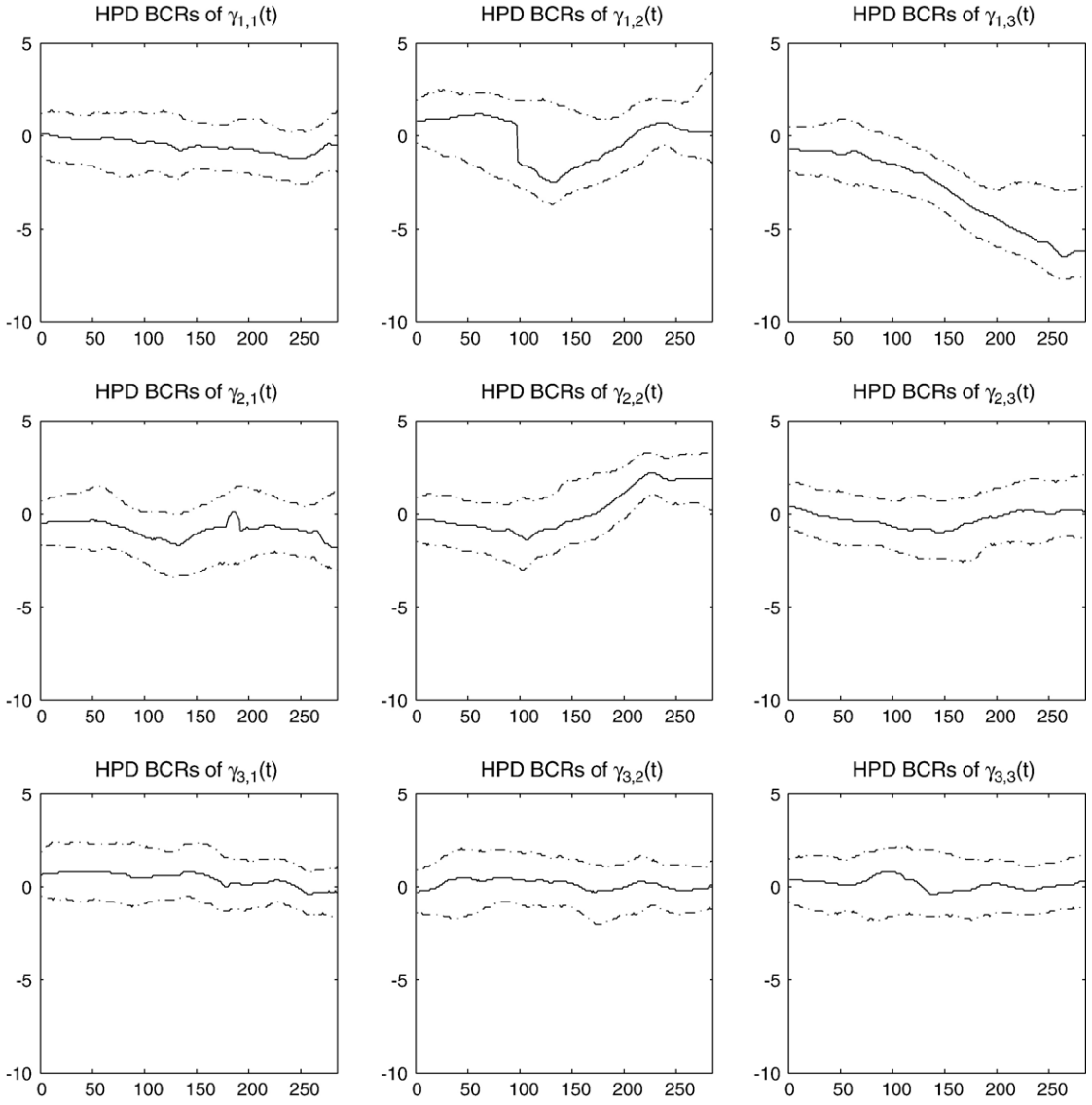

Fig. 4. Highest posterior density Bayesian credible region's of the $\gamma_{i,j}(t)$, s for model Msub 1 from the analysis of the empirical fMRI data set.

We have considered some other simplifications in combination with the above constraints (i.e., $\gamma_{3,1}(t) = \gamma_{3,2}(t) = 0$) such as $\gamma_{2,3}(t) = 0$ for all t, by the same token. However, fitting such model did not further improve the results in the sense of increasing (logarithm of) the marginal density. This may be due to the fact that the posterior variability of $\gamma_{2,3}(t)$ is nonnegligible (i.e., the connectivity between MOG and LG is time-varying), and thus replacing $\gamma_{2,3}(t)$ by zero may not be appropriate.

Table 2
Summary statistics for the connectivity parameter estimates in M_1

$\Gamma(t)$	Mean of posterior mode	Variance of posterior mode
$\gamma_{12}(t)$	−0.1495	1.3942
$\gamma_{13}(t)$	−3.0382	3.9779
$\gamma_{21}(t)$	−0.8365	0.1838
$\gamma_{23}(t)$	−0.2667	0.1577
$\gamma_{31}(t)$	0.4179	0.1330
$\gamma_{32}(t)$	0.1365	0.0634

Results and interpretations

We focus our discussion on the results of the connectivity coefficients in model M_5. The posterior mode for $\gamma_{1,2}(t)$, $\gamma_{2,1}(t)$, $\gamma_{1,3}(t)$ and $\gamma_{2,3}(t)$ with the corresponding 95% HPD BCRs (intervals) are shown in Fig. 5. It can be seen that the coefficient $\gamma_{1,3}(t)$ became stronger in a negative direction over time. This result suggests that there was a substantial suppression from DLPFC on LG ($\gamma_{1,3}(t)$) and the strength of suppression grew steadily stronger over time. Fig. 6 shows the snapshot for the smoothed posterior distribution of this coefficient at various time points in the experiment. The coefficient of $\gamma_{2,3}(t)$ was also negative throughout the experiment but was much weaker than $\gamma_{1,3}(t)$.

The LG and MOG processed conflicting information (ink color vs. shape) in the experiment and it has been argued that competitive inhibition may exist between these two sites of control (see Herd et al., in press, for example). Our results showed that there was substantial suppression from MOG on LG

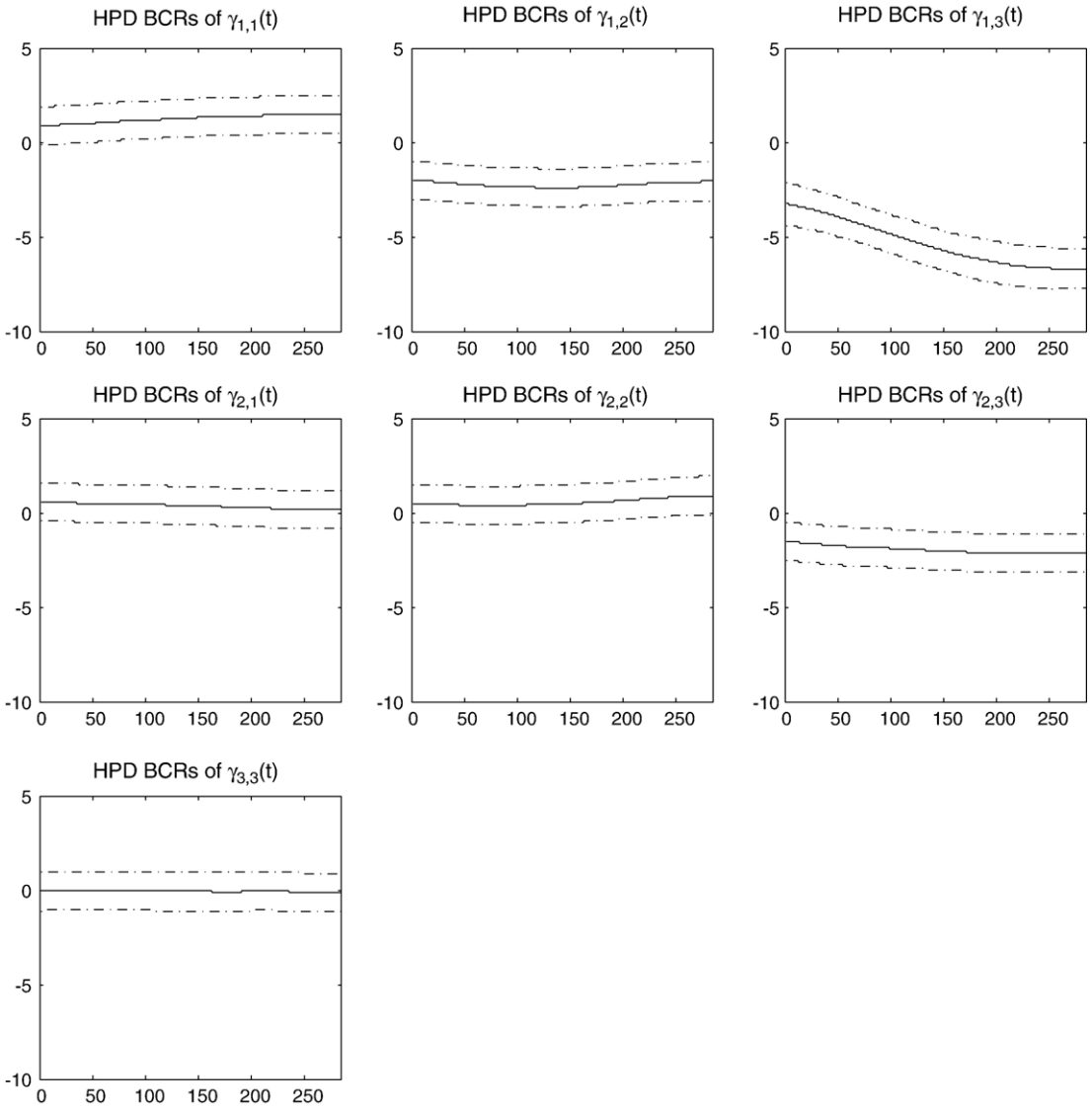

Fig. 5. Highest posterior density Bayesian credible region's of the $\gamma_{i,j}(t)$s for model M_5 from the analysis of the empirical fMRI data set.

($\gamma_{1,2}$ (t)) but no clear indication of suppression from LG to MOG ($\gamma_{2,1}(t)$).

Since model M_5 was found to fit better than model M_1, it implies that the assumption that the coefficients $\gamma_{3,1}(t)$ and $\gamma_{3,2}(t)$ were zero in the whole experiment was tenable. We may conclude that there was no feedback from the two sites of control (LG and MOG) to the source of control (DLPFC) directly. Furthermore, there is no strong evidence suggesting self-feedback at the DLPFC ($\gamma_{3,3}$ (t). However, positive self-feedback at LG ($\gamma_{1,1}(t)$) and MOG ($\gamma_{2,2}(t)$) can be seen throughout the experiment.

Overall speaking, our results are in accord with the theory of attentional control that DLPFC serves as a top-down control to decrease the neural activity within processing systems containing irrelevant and potential interfering information (Banich et al., 2000; Milham et al., 2002, 2003a,b). Our analyses suggest that the connectivity was dynamic (between DLPFC and LG) in this network. The present findings were based on a single subject's data. More fMRI data will be analyzed in the future to cross-validate the present findings.

Conclusions and future work

In this paper, we have extended the HOS approach to examine time-varying effective connectivity. Bayesian procedure for parameter estimation via Gibbs sampling has been presented. Issues pertaining to model selection and model simplification have also been addressed. Our results support the attentional control network theory and provide evidence that effective connectivity was dynamic in this network. For convenience, we detrended our fMRI data first before the Bayesian analysis. A priori detrending is not a necessary step and the drift component in the form of polynomial, random walk or cubic spline can be added to our model in a straightforward manner (see HOS, Discussion section for an illustration).

We focus the analysis on fMRI data from a single-subject in this paper. In fMRI experiments, data from multiple subjects are usually collected. Estimates for the connectivity can vary greatly across subjects when analyzing the connectivity separately for each subject. This raises the question of whether these differences

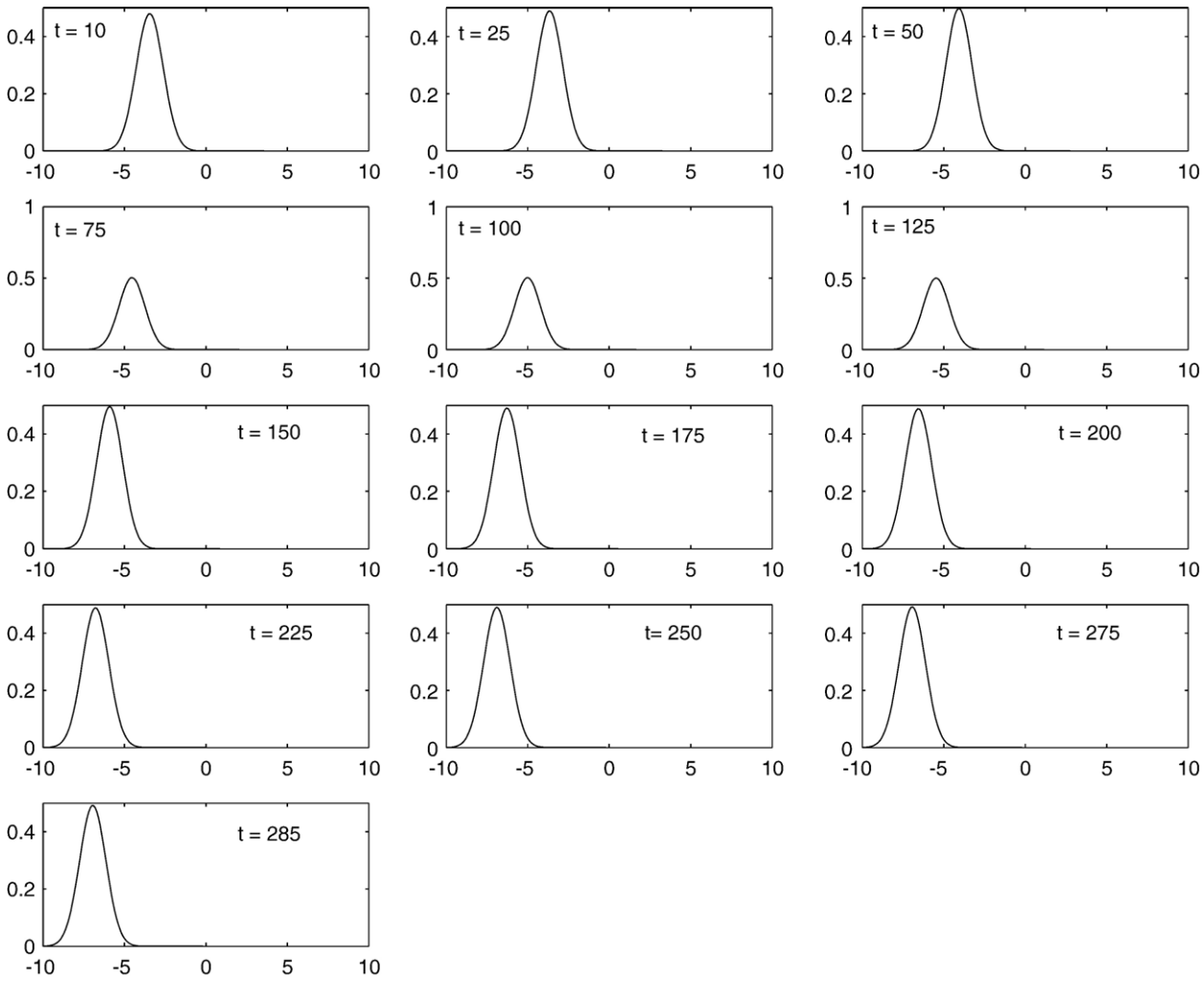

Fig. 6. Smoothed posterior distribution of $\gamma_{1,3}(t)$ for model M_5 from the analysis of the empirical fMRI data set.

in connectivity are significant or simply due to chance (Mechelli et al., 2002). Therefore, it is important to develop a method to characterize subject-specific variability in connectivity that permits evaluation of between-subject or between-group differences. One future direction is to extend the current work to modeling individual differences in dynamic connectivity patterns across the subjects.

A random walk process was postulated for characterizing the dynamics of the connectivity in this paper. Other time-varying mechanisms for the effective connectivity parameters and possibly in the error variances are possible and will be explored in the future. In particular, we plan to investigate the behavior of time-varying effective connectivity parameters assuming (i) very little or no change in them within every epoch of *rest* and *activation*, but (ii) substantial change from one epoch to another (*rest* to *activation*, and *activation* to *rest*).

To end this paper, we would like to offer a few comments about the differences between our approach and DCM (Friston et al., 2003) for effective connectivity analysis. DCM is a generative or forward model that explains hemodynamic responses as the product of neuronal activity. Neuronal activity in one area has an influence over neuronal activity in another area. In DCM, the coupling is at the level of neuronal activity as encoded by some hidden variables. In Eq. (5), we formulate the coupling among regions in terms of modeled hemodynamic responses namely $x_i(t)\beta_i(t)$. The key difference is that our approach is not based on an explicit physiologically motivated forward model but, unlike DCM, can accommodate stochastic variations in coupling. In our model, the influences are exerted through stimulus-dependent activation that is mediated by the hemodynamic response. Our approach is more agnostic to the underlying physiological mechanisms and, as such, represents a more parsimonious characterization of dependencies among different regions. In our future work, we plan to modify our approach to couple the underlying neuronal activities. Besides DCM, partial least squares approach (Lobaugh et al., 2001; McIntosh and Lobaugh, 2004) has been proposed recently for studying effective connectivity in brain imaging literature. A systematic comparison between our approach and these other techniques will be pursued in the future.

Acknowledgments

The authors thank Mike Milham for the permission and the help in preparation of the fMRI data used in this paper. The

authors are also grateful to Professor J.K. Ghosh for helpful discussions, Professor Karl J. Friston for insightful comments, Professor Nicole A. Lazar for careful reading on an earlier draft of our paper and the two referees for going through an earlier version of the paper very critically and offering many helpful suggestions. These have led to improvement in presentation and have also helped the authors in clarifying many issues. Sourabh Bhattacharya thanks the Institute of Statistics and Decision Sciences, Duke University for providing necessary facilities for carrying out part of this work. Moon-Ho R. Ho acknowledges the financial support of the National Sciences and Engineering Council of Canada. Research of Sumitra Purkayastha has been supported partially by fund available through a project titled "Analysis of fMRI Data and Human Brain Mapping" of Division of Theoretical Statistics and Mathematics, Indian Statistical Institute. Sumitra Purkayastha also expresses his sincere gratitude to Professor Keith J. Worsley for supporting from an NSERC Discovery Grant his visits to the Department of Mathematics and Statistics at McGill University in Montreal, Canada during 2003 and 2004 when this work began and continued, and thanks the Department of Mathematics and Statistics at McGill University for providing necessary facilities for carrying out part of this work.

Appendix A. Derivation for the joint posterior distribution (10)

Recall the definitions of $Q_1,\ldots,Q_6$ given in Eq. (11). Notice that the likelihood, from model (5.1), is given by:

$$\frac{\exp\left(-Q_1/2\sigma_\varepsilon^2\right)}{\left(\sqrt{2\pi}\right)^{3T}\sigma_\varepsilon^{3T}}.$$

Moreover, the contributions of Eqs. (5.2) and (5.3) to the posterior are given respectively by

$$\frac{\exp\left(-Q_2/2\sigma_w^2\right)}{\left(\sqrt{2\pi}\right)^{3(T-1)}\sigma_w^{3(T-1)}} \text{ and } \frac{\exp\left(-Q_3/2\sigma_\delta^2\right)}{\left(\sqrt{2\pi}\right)^{9(T-1)}\sigma_\delta^{9(T-1)}}.$$

The priors for $\boldsymbol{\beta}_1$, $\boldsymbol{\alpha}$ and $\{\gamma_{ij}(1), i,j = 1,2,3\}$ are given respectively by

$$\frac{\exp\left(-Q_4/2\sigma_\beta^2\right)}{\left(\sqrt{2\pi}\right)^3\sigma_\beta^3}, \frac{\exp\left(-Q_5/2\sigma_\alpha^2\right)}{\left(\sqrt{2\pi}\right)^3\sigma_\alpha^3} \text{ and } \frac{\exp\left(-Q_6/2\sigma_\gamma^2\right)}{\left(\sqrt{2\pi}\right)^9\sigma_\gamma^9}.$$

The prior for σ_ε^2 is proportional to

$$\left(\frac{1}{\sigma_\varepsilon^2}\right)^{\frac{d+2}{2}}\exp\left(-\frac{a}{2\sigma_\varepsilon^2}\right). \tag{A.1}$$

The priors for σ_w^2 and σ_δ^2 are obtained by replacing σ_ε^2 by σ_w^2 and σ_δ^2, respectively, in Eq. (A.1).

Therefore, the joint posterior distribution of $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2,\ldots,\boldsymbol{\beta}_T\boldsymbol{\Gamma}_1, \boldsymbol{\Gamma}_2,\ldots,\boldsymbol{\Gamma}_T, \boldsymbol{\alpha}, \sigma_\varepsilon^2, \sigma_w^2, \sigma_\delta^2$ is proportional to the following:

$$\prod = \exp\left(-\frac{Q_1+a}{2\sigma_\varepsilon^2}-\frac{Q_2+a}{2\sigma_w^2}-\frac{Q_3+a}{2\sigma_\delta^2}-\frac{Q_4}{2\sigma_\beta^2}-\frac{Q_5}{2\sigma_\alpha^2}-\frac{Q_6}{2\sigma_\gamma^2}\right)\times\left(\frac{1}{\sigma_\varepsilon^2}\right)^{\frac{d+2}{2}+\frac{3T}{2}}\cdot\left(\frac{1}{\sigma_w^2}\right)^{\frac{d+2}{2}+\frac{3(T-1)}{2}}\cdot\left(\frac{1}{\sigma_\delta^2}\right)^{\frac{d+2}{2}+\frac{9(T-1)}{2}}.$$

This establishes Eq. (10).

Appendix B. Derivation for the full conditional distribution of the parameters in model M_1

The propositions stated below enable us to implement Gibbs sampling scheme for obtaining the posterior distribution of the $\gamma_{i,j}(t)$s. The proof is simple and hence omitted.

For every parameter λ, the distribution of the parameter λ given all other parameters and the observations $\boldsymbol{y}_1, \boldsymbol{y}_2,\ldots,\boldsymbol{y}_T$ is written as $p(\lambda \mid \theta_{-\lambda}, \boldsymbol{y}_1, \boldsymbol{y}_2,\ldots,\boldsymbol{y}_T)$:

1. $p(\alpha_i \mid \boldsymbol{\theta}_{-\alpha_i}, \boldsymbol{y}_1, \boldsymbol{y}_2,\ldots,\boldsymbol{y}_T)$ is normal with mean and variance

$$\frac{\sum_{t=1}^{T}\{\boldsymbol{y}_i(t) - x_t\beta_i(t)\}}{\sigma_\varepsilon^2}+\frac{\mu_i}{\sigma_\alpha^2}}{\frac{T}{\sigma_\varepsilon^2}+\frac{1}{\sigma_\alpha^2}} \text{ and } \left(\frac{T}{\sigma_\varepsilon^2}+\frac{1}{\sigma_\alpha^2}\right)^{-1}, i = 1, 2, 3.$$

2. $p(\beta_i(1)\,|\,\boldsymbol{\theta}_{-\beta_i(1)}, \boldsymbol{y}_1, \boldsymbol{y}_2,\ldots,\boldsymbol{y}_T)$ is normal with mean and variance

$$\frac{\frac{x_1\{\boldsymbol{y}_i(1)-\alpha_i\}}{\sigma_\varepsilon^2}+\frac{x_i\sum_{l=1}^{3}\gamma_{l,i}(2)\left\{\beta_l(2)-x_i\sum_{k=1,\,k\neq i}^{3}\gamma_{l,k}(2)\beta_k(1)\right\}}{\sigma_w^2}+\frac{\beta_1(0)}{\sigma_\beta^2}}{\frac{x_1^2}{\sigma_\varepsilon^2}+\frac{x_1^2\left(\gamma_{1,i}^2(2)+\gamma_{2,i}^2(2)+\gamma_{3,i}^2(2)\right)}{\sigma_w^2}+\frac{1}{\sigma_\beta^2}}$$
and $$\left(\frac{x_1^2}{\sigma_\varepsilon^2}+\frac{x_1^2\left(\gamma_{1,1}^2(2)+\gamma_{2,1}^2(2)+\gamma_{3,i}^2(2)\right)}{\sigma_w^2}+\frac{1}{\sigma_\beta^2}\right)^{-1}, i=1,2,3.$$

3. $p(\beta_i(t)\,|\,\boldsymbol{\theta}_{-\beta_i(t)}, \boldsymbol{y}_1, \boldsymbol{y}_2,\ldots,\boldsymbol{y}_T)$ $(t=2,\ldots,T-1)$ is normal with mean and variance

$$\frac{\frac{x_t\{\boldsymbol{y}_i(t)-\alpha_i\}}{\sigma_\varepsilon^2}+\frac{x_{t-1}\sum_{j=1}^{3}\gamma_{i,j}(t)\beta_j(t-1)}{\sigma_w^2}+\frac{x_t\sum_{l=1}^{3}\gamma_{l,i}(t+1)\left[\beta_l(t+1)-x_t\sum_{k=1,\,k\neq i}^{3}\gamma_{l,k}(t+1)\beta_k(t)\right]}{\sigma_w^2}}{\frac{x_t^2}{\sigma_\varepsilon^2}+\frac{1+x_t^2\left(\gamma_{1,i}^2(t+1)+\gamma_{2,i}^2(t+1)+\gamma_{3,i}^2(t+1)\right)}{\sigma_w^2}}$$
$$\text{and}\left(\frac{x_t^2}{\sigma_\varepsilon^2}+\frac{1+x_t^2\left(\gamma_{1,i}^2(t+1)+\gamma_{2,i}^2(t+1)+\gamma_{3,i}^2(t+1)\right)}{\sigma_w^2}\right)^{-1}, i=1,2,3.$$

4. $p(\beta_i(T)\,|\,\boldsymbol{\theta}_{-\beta_i(T)}, \boldsymbol{y}_1, \boldsymbol{y}_2,\ldots,\boldsymbol{y}_T)$ is normal with mean and variance

$$\frac{\frac{x_T\{\boldsymbol{y}_i(T)-\alpha_i\}}{\sigma_\varepsilon^2}+\frac{x_{T-1}\sum_{j=1}^{3}\gamma_{i,j}(T)\beta_j(T-1)}{\sigma_w^2}}{\frac{x_T^2}{\sigma_\varepsilon^2}+\frac{1}{\sigma_w^2}}\ \text{and}\ \left(\frac{x_T^2}{\sigma_\varepsilon^2}+\frac{1}{\sigma_w^2}\right)^{-1}, i=1,2,3.$$

5. $p(\gamma_{i,j}(1)\,|\,\boldsymbol{\theta}_{-\gamma_{i,j}(1)}, \boldsymbol{y}_1, \boldsymbol{y}_2,\ldots,\boldsymbol{y}_T)$ is normal with mean and variance

$$\frac{\frac{\gamma_{i,j}(2)}{\sigma_\delta^2}+\frac{\gamma_{i,j}(0)}{\sigma_\gamma^2}}{\frac{1}{\sigma_\delta^2}+\frac{1}{\sigma_\gamma^2}}\ \text{and}\ \left(\frac{1}{\sigma_\delta^2}+\frac{1}{\sigma_\gamma^2}\right)^{-1}, i,j=1,2,3.$$

6. $p(\gamma_{i,j}(T)\,|\,\boldsymbol{\theta}_{-\gamma_{i,j}(T)}, \boldsymbol{y}_1, \boldsymbol{y}_2,\ldots,\boldsymbol{y}_T)$ is normal with mean and variance

$$\frac{\frac{\gamma_{i,j}(T-1)}{\sigma_\delta^2}+\frac{x_{T-1}\beta_j(T-1)\left\{\beta_i(T)-x_{T-1}\sum_{k=1,\,k\neq j}^{3}\gamma_{i,k}(T)\beta_k(T-1)\right\}}{\sigma_w^2}}{\frac{1}{\sigma_\delta^2}+\frac{x_{T-1}^2\beta_j^2(T-1)}{\sigma_w^2}}\ \text{and}\ \left(\frac{1}{\sigma_\delta^2}+\frac{x_{T-1}^2\beta_j^2(T-1)}{\sigma_w^2}\right)^{-1}, i,j=1,2,3.$$

7. $p(\gamma_{i,j}(t)\,|\,\boldsymbol{\theta}_{-\gamma_{i,j}(t)}, \boldsymbol{y}_1, \boldsymbol{y}_2,\ldots,\boldsymbol{y}_T)$ $(t=2,\ldots,T-1)$ is normal with mean and variance

$$\frac{\frac{\gamma_{i,j}(t-1)+\gamma_{i,j}(t+1)}{\sigma_\delta^2}+\frac{x_{t-1}\beta_j(t-1)\left\{\beta_i(t)-x_{t-1}\sum_{k=1,k\neq j}^{3}\gamma_{i,k}(t)\beta_k(t-1)\right\}}{\sigma_w^2}}{\frac{2}{\sigma_\delta^2}+\frac{x_{t-1}^2\beta_j^2(t-1)}{\sigma_w^2}}\ \text{and}\ \left(\frac{2}{\sigma_\delta^2}+\frac{x_{t-1}^2\beta_j^2(t-1)}{\sigma_w^2}\right)^{-1}, i,j=1,2,3.$$

8. (a) $p(\sigma_\varepsilon^2\,|\,\boldsymbol{\theta}-\sigma_\varepsilon^2, y_1, \boldsymbol{y}_2,\ldots,\boldsymbol{y}_T)$ is IG(Q_1+a, $d+3T$). [For definition of Q_1, refer to Eq. (11).]
8. (b) $p(\sigma_w^2\,|\,\boldsymbol{\theta}-\sigma_\varepsilon^2, \boldsymbol{y_1}, \boldsymbol{y_2},\ldots,\boldsymbol{y}_T)$ is IG(Q_2+a, $d+3(T-1)$). [For definition of Q_2, refer to Eq. (11).]
8. (c) $p(\sigma_\delta^2\,|\,\boldsymbol{\theta}-\sigma_\varepsilon^2, \boldsymbol{y_1}, \boldsymbol{y_2},\ldots,\boldsymbol{y}_T)$ is IG(Q_3+a, $d+9(T-1)$). [For definition of Q_3, refer to Eq. (11).]

Appendix C. Computational details and rationale for the choice of hyperparameters

Since the integral (12) is not, in general, analytically tractable, an alternative to analytical integration is Monte Carlo integration. In other words, provided that the prior $\pi(\boldsymbol{\theta}\,|\,\boldsymbol{\xi})$ is proper for all plausible $\boldsymbol{\xi}$, one can draw samples of size, say, N of $\boldsymbol{\theta}$, denoted by $\boldsymbol{\theta}^{(k)}$, $k = 1,\ldots,N$. Given the samples, Eq. (12) may be estimated as

$$\hat{p}(\boldsymbol{Y}\,|\,\boldsymbol{\xi}) = \frac{1}{N}\sum_{k=1}^{N} p\left(\boldsymbol{Y}\,|\,\boldsymbol{\theta}^{(k)}\right) \tag{A.2}$$

Note that it is necessary to compute the estimate (A.2) for each of many possible choices of $\boldsymbol{\xi}$. In practice, hyperparameters can be chosen by simulation from a given plausible range. Thus, it is very important that a fast efficient Monte Carlo simulation scheme is adopted to reduce the computational burden. Markov Chain Monte Carlo (MCMC) is a general-purpose algorithm but is computationally very expensive for the current purpose.

The discussion above motivates us to see if we can employ direct Monte Carlo integration. However, in order to employ this, it is necessary that the prior on $\boldsymbol{\theta}$ given $\boldsymbol{\xi}$ ($\pi(\boldsymbol{\theta}\,|\,\boldsymbol{\xi})$) is proper; otherwise, samples cannot be drawn from it. For our analysis, we chose independent Inverse Gamma IG(a,d) priors on components of $\boldsymbol{\sigma}$, which means the inverse of the component in $\boldsymbol{\sigma}$ follows independent Gamma distribution (see also footnote 4 for the mathematical definition of IG(a,d)). We set the two parameters $a = d = 0$ in our application which implies impropriety of the prior for $\boldsymbol{\sigma}$ and consequently for all the parameters.

This, in turn, suggests that without some modification Monte Carlo integration cannot be employed in this context. Notice now that the problem arising out of impropriety of the prior for $\boldsymbol{\sigma}$ can be avoided by integrating out $\boldsymbol{\sigma}$ in Eq. (12). This implies that the marginal density under consideration is given by the following:

$$p(\boldsymbol{Y}\,|\,\boldsymbol{\xi}_{\psi}) \propto \int \exp\left(-\frac{Q_4}{2\sigma_{\beta}^2} - \frac{Q_5}{2\sigma_{\alpha}^2} - \frac{Q_6}{2\sigma_{\gamma}^2}\right) \times \frac{1}{Q_1^{(3T-1)/2} Q_2^{(3T-4)/2} Q_3^{(9T-10)/2}}\, d\psi. \tag{A.3}$$

However, there does not appear to be any obvious Monte Carlo based method for evaluating this integral numerically.

One way to avoid numerical evaluation of Eq. (12) or of Eq. (A.3) is to plug in a reasonable *estimate* (typically, the posterior mode) of $\boldsymbol{\sigma}$ the integrand in Eq. (12) and evaluate the integral of the resulting integrand with respect only to ψ by Monte Carlo method. The quantity obtained in this manner is readily seen to be a function of $\boldsymbol{\xi}_{\psi}$ alone. This function is subsequently maximized with respect to $\boldsymbol{\xi}_{\psi}$. Details have been provided in Choice of hyperparameters.

Notice that in Eq. (12), or in Eq. (13), the set of parameters, denoted by $\boldsymbol{\theta}$, is sought to be eliminated by integration for the purpose of obtaining the hyperparameters. Therefore, $\boldsymbol{\theta}$ may be considered as a set of "nuisance parameters". A relevant reference on elimination of nuisance parameters can be found in Berger et al. (1999). In their paper, these authors discussed primarily elimination of the nuisance parameters by integrating them out, but also discussed maximization with respect to the nuisance parameters for the same purpose. In particular, they refer to the functional form that results after integrating out nuisance parameters from the likelihood as "uniform-integrated likelihood". Maximization with respect to the nuisance parameters, fixing those of interest, produces what is known as "profile likelihood". For implementation of the ML-II method, integration alone has been recommended for elimination of nuisance parameters (see Eq. (12)). However, as discussed earlier, eliminating nuisance parameters in our problem with a combination of integration and maximization is computationally more convenient than integration alone. In particular, eliminating $\boldsymbol{\sigma}$ by fixing it to its posterior mode takes care of the problem caused by impropriety of the prior for $\boldsymbol{\sigma}$. This then allows for Monte Carlo integration which is needed for obtaining an estimate of $p(\boldsymbol{Y}\,|\,\boldsymbol{\xi})$, denoted as $\mathrm{p}(\boldsymbol{Y}\,|\,\boldsymbol{\xi}_{\psi})$.

Given improper priors on all the components of $\boldsymbol{\sigma}$ as mentioned before, we computed their posterior distributions under different choices of the hyperparameters of ψ. The posteriors remained almost unchanged which suggests considerable robustness of our approach. Therefore, in replacing Eq. (12) by Eq. (14), we do not introduce any additional dependence among the quantity to be maximized over $\boldsymbol{\xi}_{\psi}$. This certainly is a computational advantage (see also Example 3 of Berger et al., 1999). Therefore, for the purpose of finding a meaningful analogue of Eq. (12), we fixed $\boldsymbol{\sigma}$ to its posterior mode and avoided consequently the integration with respect to $\boldsymbol{\sigma}$ in Eq.(12).

Appendix D. Derivation for the full conditional distribution of the parameters in model M_2

The propositions stated below enable us to implement Gibbs sampling scheme for obtaining the posterior distribution of the $\gamma_{i,j}(t)$s in M_2. The prior for ρ is assumed to be the i.i.d. uniform priors over $(-1, 1)$. Notice also that the definition of Q_1 (p. 11) needs to be modified to:

$$Q_1^{*} \stackrel{\text{def}}{=} \sum_{i=1}^{3} Q_{1,i}, \tag{A.2.1}$$

where

$$Q_{1,i} \stackrel{\text{def}}{=} \sum_{t=2}^{T} \left[\{\boldsymbol{y}_i(t) - \alpha_i - x_t\beta_i(t)\} - \rho\{\boldsymbol{y}_i(t-1) - \alpha_i - x_{t-1}\beta_i(t-1)\}\right]^2 + \left(1-\rho_i^2\right)\{\boldsymbol{y}_i(1) - \alpha_i - x_1\beta_i(1)\}^2. \tag{A.2.2}$$

Notice now that the joint posterior distribution of

$$\boldsymbol{\theta}\equiv\left(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\ldots\boldsymbol{\beta}_T,\boldsymbol{\Gamma}_1,\boldsymbol{\Gamma}_2,\ldots,\boldsymbol{\Gamma}_T,\boldsymbol{\alpha},\sigma_\varepsilon^2,\sigma_w^2,\sigma_\delta^2,\rho_1,\rho_2,\rho_3\right)$$

is *proportional* to the following:

$$\prod=\exp\left(-\frac{Q_1^*+a}{2\sigma_\varepsilon^2}-\frac{Q_2+a}{2\sigma_w^2}-\frac{Q_3+a}{2\sigma_\delta^2}-\frac{Q_4}{2\sigma_\beta^2}-\frac{Q_5}{2\sigma_\alpha^2}-\frac{Q_6}{2\sigma_\gamma^2}\right)\times\left[\left(\frac{1}{\sigma_\varepsilon^2}\right)^{\frac{d+2}{2}+\frac{3T}{2}}\cdot\left(\frac{1}{\sigma_w^2}\right)^{\frac{d+2}{2}+\frac{3(T-1)}{2}}\cdot\left(\frac{1}{\sigma_\delta^2}\right)^{\frac{d+2}{2}+\frac{9(T-1)}{2}}\right]\times\left[\prod_{i=1}^{3}\left(1-\rho_i^2\right)^{\frac{1}{2}}\right].$$

As in Appendix A.2, for every parameter λ, the distribution of the parameter λ given all other parameters and the observations $\boldsymbol{y}_1$, $\boldsymbol{y}_2,\ldots\boldsymbol{y}_T$ is written as $p(\lambda\,|\,\boldsymbol{\theta}_{-\lambda},\boldsymbol{y}_1,\boldsymbol{y}_2,\ldots\boldsymbol{y}_T)$:

1. $p(\alpha_i\,|\,\boldsymbol{\theta}_{-\alpha_i},\boldsymbol{y}_1,\boldsymbol{y}_2,\ldots\boldsymbol{y}_T)$ is normal with mean and variance

$$\frac{(1-\rho_i)\sum_{t=2}^{T}\left[\{\boldsymbol{y}_i(t)-x_t\beta_i(t)\}-\rho_i\{\boldsymbol{y}_i(t-1)-x_{t-1}\beta_i(t-1)\}\right]}{\sigma_\varepsilon^2}+\frac{\left(1-\rho_i^2\right)\{\boldsymbol{y}_i(1)-x_1\beta_i(1)\}}{\sigma_\varepsilon^2}+\frac{\mu_i}{\sigma_\alpha^2}}{\frac{(T-1)(1-\rho)^2+\left(1-\rho_i^2\right)}{\sigma_\varepsilon^2}+\frac{1}{\sigma_\alpha^2}}\text{ and }\left(\frac{(T-1)(1-\rho)^2+\left(1-\rho_i^2\right)}{\sigma_\varepsilon^2}+\frac{1}{\sigma_\alpha^2}\right)^{-1},$$

$i=1,2,3.$

2. $p(\beta_i(1)\,|\,\boldsymbol{\theta}_{-\beta_i(1)},\boldsymbol{y}_1,\boldsymbol{y}_2,\ldots\boldsymbol{y}_T)$ is normal with mean and variance

$$\frac{\frac{x_1(y_i(1)-\alpha_i)-\rho_i(y_i(2)-\alpha_i-x_2\beta_i(2))}{\sigma_\varepsilon^2}+\frac{x_1\sum_{l=i}^{3}\gamma_{l,i}(2)\{\beta_l(2)-x_1\sum_{k=1,k\neq i}^{3}\gamma_{l,k}(2)\beta_k(1)\}}{\sigma_w^2}+\frac{\beta_l(0)}{\sigma_\beta^2}}{\frac{x_1^2}{\sigma_\varepsilon^2}+\frac{x_1^2\left(\gamma_{1,i}^2(2)+\gamma_{2,i}^2(2)+\gamma_{3,i}^2(2)\right)}{\sigma_w^2}+\frac{1}{\sigma_\beta^2}}\text{ and}$$

$$\left(\frac{x_1^2}{\sigma_\varepsilon^2}+\frac{x_1^2\left(\gamma_{1,1}^2(2)+\gamma_{2,1}^2(2)+\gamma_{3,i}^2(2)\right)}{\sigma_w^2}+\frac{1}{\sigma_\beta^2}\right)^{-1},\ i=1,2,3.$$

3. $p(\beta_i(t)\,|\,\boldsymbol{\theta}_{-\beta_i}(t),\boldsymbol{y}_1,\boldsymbol{y}_2,\ldots\boldsymbol{y}_T)$ $(t=2,\ldots,T-1)$ is normal with mean and variance

$$\frac{\frac{x_t\{(1+\rho_i^2)(\boldsymbol{y}_i(t)-\alpha_i)\}}{\sigma_\varepsilon^2}-\frac{x_t\{\rho_i(\boldsymbol{y}_i(t+1)-\alpha_i-x_{t-1}\beta_i(t+1))\}}{\sigma_\varepsilon^2}\frac{x_t\{\rho_i(\boldsymbol{y}_i(t-1)-\alpha_i-x_{t-1}\beta_i(t-1))\}}{\sigma_\varepsilon^2}+\frac{x_{t-1}\sum_{j=1}^{3}\gamma_{i,j}(t)\beta_j(t-1)}{\sigma_w^2}+\frac{x_t\sum_{l=1}^{3}\gamma_{l,i}(t+1)\left[\beta_l(t+1)-x_t\sum_{k=1,\,k\neq i}^{3}\gamma_{l,k}(t+1)\beta_j(t-1)\right]}{\sigma_w^2}}{\frac{x_t^2(1+\rho_i^2)}{\sigma_\varepsilon^2}+\frac{1+x_t^2\left(\gamma_{1,i}^2(t+1)+\gamma_{2,i}^2(t+1)+\gamma_{3,i}^2(t+1)\right)}{\sigma_w^2}}$$

4. $p(\beta_i(T)\,|\,\boldsymbol{\theta}_{-\beta_i(T)},\boldsymbol{y}_1,\boldsymbol{y}_2,\ldots\boldsymbol{y}_T)$ is normal with mean and variance

$$\frac{\frac{x_T\{(\boldsymbol{y}_i(T)-\alpha_i)-\rho_i(\boldsymbol{y}_i(T-1)-\alpha_i-x_{T-1}\beta_1(T-1))\}}{\sigma_\varepsilon^2}+\frac{x_{T-1}\sum_{j=1}^{3}\gamma_{i,j}(T)\beta_j(T-1)}{\sigma_w^2}}{\frac{x_T^2}{\sigma_\varepsilon^2}+\frac{1}{\sigma_w^2}}\text{ and }\left(\frac{x_T^2}{\sigma_\varepsilon^2}+\frac{1}{\sigma_w^2}\right)^{-1},\ i=1,2,3.$$

5. $p(\gamma_{i,j}(1)\,|\,\boldsymbol{\theta}_{-\gamma_{i,j}(1)},\boldsymbol{y}_1,\boldsymbol{y}_2,\ldots\boldsymbol{y}_T)$ is normal with mean and variance

$$\frac{\frac{\gamma_{i,j}(2)}{\sigma_\delta^2}+\frac{\gamma_{i,j}(0)}{\sigma_\gamma^2}}{\frac{1}{\sigma_\delta^2}+\frac{1}{\sigma_\gamma^2}}\text{ and }\left(\frac{1}{\sigma_\delta^2}+\frac{1}{\sigma_\gamma^2}\right)^{-1},\ i,j=1,2,3.$$

6. $p(\gamma_{i,j}(T)\,|\,\boldsymbol{\theta}_{-\gamma_{i,j}(T)},\boldsymbol{y}_1,\boldsymbol{y}_2,\ldots\boldsymbol{y}_T)$ is normal with mean and variance

$$\frac{\frac{\gamma_{i,j}(T-1)}{\sigma_\delta^2}+\frac{x_{T-1}\beta_j(T-1)\left\{\beta_i(T)-x_{T-1}\sum_{k=1,\,k\neq k}^{3}\gamma_{i,k}(T)\beta_k(T-1)\right\}}{\sigma_w^2}}{\frac{1}{\sigma_\delta^2}+\frac{x_{T-1}^2\beta_j^2(T-1)}{\sigma_w^2}}\text{ and }\left(\frac{1}{\sigma_\delta^2}+\frac{x_{T-1}^2\beta_j^2(T-1)}{\sigma_w^2}\right)^{-1},\ \text{i , j}-1,2,3.$$

7. $p(\gamma_{i,j}(t)\,|\,\boldsymbol{\theta}_{-\gamma_{i,j}(t)}, \boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_T)$ $(t = 2, \ldots, T-1)$ is normal with mean and variance

$$\frac{\frac{\gamma_{i,j}(t-1)+\gamma_{i,j}(t+1)}{\sigma_\delta^2} + \frac{x_{t-1}\beta_j(t-1)\left\{\beta_i(t)-x_{t-1}\sum_{k=1,k\neq j}^{3}\gamma_{i,k}(t)\beta_k(t-1)\right\}}{\sigma_w^2}}{\frac{2}{\sigma_\delta^2} + \frac{x_{t-1}^2\beta_j^2(t-1)}{\sigma_w^2}} \text{ and } \left(\frac{2}{\sigma_\delta^2} + \frac{x_{t-1}^2\beta_j^2(t-1)}{\sigma_w^2}\right)^{-1}, i,j = 1,2,3.$$

8. (a) $p(\sigma_\varepsilon^2\,|\,\boldsymbol{\theta}-\sigma_\varepsilon^2, \boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_T)$ is IG$(Q_1^* + a, d + 3T)$. [For definition of Q_1^*, refer to Eq. (A.2.1).]
8. (b) $p(\sigma_w^2\,|\,\boldsymbol{\theta}-\sigma_w^2, \boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_T)$ is IG$(Q_2 + a, d + 3(T-1))$. [For definition of Q_2, refer to Eq. (11).]
8. (c) $p(\sigma_\delta^2\,|\,\boldsymbol{\theta}-\sigma_\delta^2, \boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_T)$ is IG$(Q_3 + a, d + 9(T-1))$. [For definition of Q_3, refer to Eq. (11).]
9. $p(\rho_i\,|\,\boldsymbol{\theta}-\rho_i, \boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_T)$ is a distribution having probability density function proportional to

$$\left(1-\rho_i^2\right)^{\frac{1}{2}} \exp\left(-\frac{Q_{1,i}}{2\sigma_e^2}\right),\ \rho_i \in (-1,1),\ i = 1, 2, 3.$$

References

Aertsen, A., Preißl, H., 1991. Dynamics of activity and connectivity in physiological neuronal networks. In: Schuster, H.G. (Ed.), Non-Linear Dynamics and Neuronal Networks. VCH Publishers, New York, pp. 281–302.

Banich, M.T., Milham, M.P., Atchley, R., Cohen, N.J., Webb, A., Wszalek, T., Kramer, A.F., Liang, Z.P., Wright, A., Shenker, J., Magin, R., 2000. fMRI studies of stroop tasks reveal unique roles of anterior and posterior brain systems in attentional selection. J. Cogn. Neurosci. 12, 988–1000.

Bartlett, M.S., 1937. Properties of sufficiency and statistical tests. Proc. R. Soc. Lond., Series A 160, 268–282.

Berger, J.O., 1985. Statistical Decision Theory and Bayesian Analysis. (2nd ed.). Springer, New York.

Berger, J.O., Liseo, B., Wolpert, R.L., 1999. Integrated likelihood methods for eliminating nuisance parameters. Stat. Sci. 14, 1–28.

Birn, R.M., Saad, Z.S., Bandetti, P.A., 2001. Spatial heterogeneity of the non-linear dynamics in the fMRI BOLD response. NeuroImage 14, 817–826.

Büchel, C., Friston, K., 1998. Dynamic changes in effective connectivity characterized by variable parameter regression and Kalman filtering. Hum. Brain Mapp. 6, 403–408.

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

Buxton, R.B., Wong, E.C., Frank, L.R., 1998. Dynamics of blood flow and oxygenation changes during brain activation: the balloon model. Magn. Reson. Med. 39, 855–864.

Cai, Z., Fan, J., Yao, Q., 2000. Functional coefficient regression models for non-linear time series. J. Am. Stat. Assoc. 95, 941–956.

Carlin, B.P., Louis, T.A., 2000. Bayes and Empirical Bayes Methods for Data Analysis (2nd ed.). Chapman and Hall, Boca Raton.

Chen, R., Tsay, R., 1993. Functional coefficient autoregressive models. J. Am. Stat. Assoc. 88, 298–308.

Clare, S. (1997). Functional Magnetic Resonance Imaging: Methods and Applications, Unpublished PhD thesis. [http://www.fmrib.ox.ac.uk/~stuart].

Clyde, M., George, E., 2004. Model uncertainty. Stat. Sci. 19, 81–94.

Corbetta, M., Miezin, F.M., Dobmeyer, S., Shulman, G.L., Petersen, S.E., 1991. Selective and divided attention during visual discriminations of shape, color, and speed: functional anatomy by positron emission tomography. J. Neurosci. 8, 2383–2402.

Deco, G., Rolls, E.T., Horwitz, B., 2004. Integrating fMRI and single-cell data of visual working memory. Neurocomputing 58–60, 729–737.

Donnet, S., Lavielle, M., Ciuciu, P., Poline, J.-B., 2004. Selection of temporal models for event-related fMRI. Proceedings of IEEE International Symposium on Biomedical Imaging: From Nano to Macro. IEEE, Arlington, VA.

Frackowiak, R.S.J., Friston, K.J., Frith, C.D., Dolan, R.J., Price, C.J., Zeki, S., Ashburner, J., and Penny, W. (2004). Human brain function (2nd ed.). Elsevier, Academic Press: Amsterdam.

Friston, K., 1994. Functional and effective connectivity in neuroimaging: a synthesis. Hum. Brain Mapp. 2, 56–78.

Friston, K., Fletcher, P., Josephs, O., Holmes, A., Rugg, M.D., Turner, R., 1998. Event-related fMRI: characterizing differential responses. NeuroImage 7, 30–40.

Friston, K.J., Harrison, L., Penny, W., 2003. Dynamic causal modelling. NeuroImage 19, 1273–1302.

Frith, C., 2001. A framework for studying the neural basis of attention. Neuropsychologia 39, 1367–1371.

Gamerman, D., Migon, H.S., 1993. Dynamic hierarchical models. J. R. Stat. Soc., Ser. B 55, 642–649.

Gelfand, A.E., Smith, A.F.M., 1990. Sampling-based approaches to calculating marginal densities. J. Am. Stat. Assoc. 85, 398–409.

Gelman, A., Rubin, D.B., 1992. Inference from iterative simulation using multiple sequences. Stat. Sci. 7, 457–472.

Genovese, C., 2000. A Bayesian time-course model for functional magnetic resonance imaging data (with discussion). J. Am. Stat. Assoc. 95, 691–719.

Glover, G.H., 1999. Deconvolution of impulse function response in event-related BOLD fMRI. NeuroImage 9, 416–429.

Gössl, C., Auer, D.P., Fahrmeir, F., 2001. Bayesian spatiotemporal inference in functional magnetic resonance imaging. Biometrics 57, 554–562.

Harrison, L., Penny, W.D., Friston, K., 2003. Multivariate autoregressive modelling of fMRI time series. NeuroImage 19, 1477–1491.

Herd, S.A., Banich, M.T., O'Reilly, R.C., in press. Neural mechanisms of cognitive control: an integrative model of stroop task performance and fMRI data. J. Cogn. Neurosci.

Ho, M.R., Ombao, H., Shumway, R., 2005. A State-Space Approach to Modelling Brain Dynamics to Appear in Statistica Sinica.

Horwitz, B., 1998. Using functional brain imaging to understand human cognition. Complexity 3, 39–52.

Horwitz, B., Tagamets, M., McIntosh, A.R., 1999. Neural modelling, functional, brain imaging, and cognition. Trends Cogn. Sci. 3, 91–98.

Jeffreys, H., 1961. Theory of Probability (Third ed.). Oxford Univ. Press, London.

Jezzard, P., Matthews, P.M., Smith, S.M. (Eds.), Functional MRI: An Introduction to Methods. Oxford Univ. Press, New York.

Kass, R.E., Raftery, A.E., 1995. Bayes factors. J. Am. Stat. Assoc. 90, 773–795.

Kelley, W.M., Miezin, F.M., McDermott, K.B., Buckner, R.L., Raichle, M.E., Cohen, N.J., Ollinger, J.M., Akbudak, E., Conturo, T.E., Snyder, A.Z., Petersen, S.E., 1998. Hemispheric specialization in human dorsal frontal cortex and medial temporal lobe for verbal and nonverbal memory encoding. Neuron 20, 927–936.

Kiebel, S.J., Glaser, D.E., Friston, K.J., 2003. A heuristic for the degrees of freedom of statistics based on multiple variance parameters. NeuroImage 20, 466–478.

Kirk, E., Ho, M.R., Colcombe, S.J., Kramer, A.F., 2005. A structural

equation modelling analysis of attentional control: an event-related fMRI study. Cogn. Brain Res. 22, 349–357.
Lahaye, P.-J., Poline, J.-B., Flandin, G., Dodel, S., Garnero, L., 2003. Functional connectivity: studying nonlinear, delayed interactions between BOLD signal. NeuroImage 20, 962–974.
Liao, C., Worsley, K.J., Poline, J.-B., Duncan, G.H., Evans, A.C., 2002. Estimating the delay of the response in fMRI data. NeuroImage 16, 593–606.
Lobaugh, N.J., West, R., McIntosh, A.R., 2001. Spatiotemporal analysis of experimental differences in event-related potential data with partial least squares. Psychophysiology 38, 517–530.
Marchini, J.L., Ripley, B.D., 2000. A new statistical approach to detecting significant activation in functional MRI. NeuroImage 12, 366–380.
Marrelect, G., Benali, H., Ciuciu, P., Pelegrini-Issac, M., Poline, J.-B., 2003. Robust Bayesian estimation of the hemodynamic response function in event-related BOLD MRI using basic physiological information. Hum. Brain Mapp. 19, 1–17.
McIntosh, A.R., 2000. Towards a network theory of cognition. Neural Netw. 13, 861–870.
McIntosh, A.R., Gonzalez-Lima, F., 1994. Structural equation modelling and its application to network analysis of functional brain imaging. Hum. Brain Mapp. 2, 2–22.
McIntosh, A.R., Lobaugh, N.J., 2004. Partial least squares analysis of neuroimaging data: applications and advances. NeuroImage 23, S250–S263.
Mechelli, A., Penny, W.D., Price, C., Gitelman, D., Friston, K.J., 2002. Effective connectivity and inter-subject variability: using a multi-subject network to test differences and commonalities. NeuroImage 17, 1459–1469.
Milham, M.P., Erickson, K.I., Banich, M.T., Kramer, A.F., Webb, A., Wszalek, T., Cohen, N.J., 2002. Attentional control in the aging brain: insights from an fMRI study of the stroop task. Brain Cogn. 49, 277–296.
Milham, M.P., Banich, M., Claus, E., Cohen, N., 2003a. Practice-related effects demonstrate complementary role of anterior cingulate and prefrontal cortices in attentional control. NeuroImage 18, 483–493.
Milham, M.P., Banich, M., Barad, V., 2003b. Competition for priority in processing increases prefrontal cortex's involvement in top-down control: an event-related fMRI study of the stroop task. Cogn. Brain Res. 17, 212–222.
Nyberg, L., McIntosh, A.R., 2001. Functional neuroimaging: network analysis. In: Cabeza, R., Kingstone, A. (Eds.), Handbook of Functional Neuroimaging of Cognition. The MIT Press, Cambridge, MA, pp. 49–72.
O'Hagan, A., Forster, J., 2004. Bayesian inference (2nd ed.). Kendall's Advanced Theory of Statistics, vol. 2B. Arnold, London.
Penny, W.D., Kiebel, S., Friston, K., 2003. Variational Bayesian inference for fMRI time series. NeuroImage 19, 727–741.
Penny, W.D., Stephan, K.E., Mechelli, A., Friston, K.J., 2004a. Comparing dynamic causal models. NeuroImage 22, 1157–1172.
Penny, W.D., Stephan, K.E., Mechelli, A., Friston, K.J., 2004b. Modelling functional integration: a comparison of structural equation and dynamic causal and models. NeuroImage 23 (Suppl. 1), 264–274.
Raftery, A.E., Lewis, S.M., 1992. How many iterations in the Gibbs sampler? In: Bernardo, J.M., et al., (Eds.), Bayesian Statistics vol. 4. Oxford Univ. Press, Oxford, pp. 763–773.
Tagamets, M.A., Horwitz, B., 1998. Integrating electrophysiological and anatomical experimental data to create a large-scale model that simulates a delayed match-to-sample human brain imaging study. Cereb. Cortex 8, 310–320.
West, M., Harrison, J., 1999. Bayesian Forecasting and Dynamic Models (2nd ed.). Springer, New York.
Woolrich, M.W., Ripley, B.D., Brady, M., Smith, S.M., 2001. Temporal autocorrelation in univariate linear modeling of fMRI Data. NeuroImage 14, 1370–1386.
Woolrich, M., Jenkinson, M., Brady, J.M., Smith, S., 2004a. Constrained linear basis set for HRF modelling using variational Bayes. Neuroimage 21, 1748–1761.
Woolrich, M., Jenkinson, M., Brady, J.M., Smith, S., 2004b. Fully Bayesian spatio-temporal modelling of fMRI data. IEEE Trans. Med. Imag. 23, 213–231.
Worsley, K.J., 2003. Detecting activation in fMRI data. Stat. Methods Med. Res. 12, 401–418.