The Annals of Applied Statistics
2017, Vol. 11, No. 2, 868–897
DOI: 10.1214/17-AOAS1021

BAYESIAN INFERENCE OF HIGH-DIMENSIONAL, CLUSTER-STRUCTURED ORDINARY DIFFERENTIAL EQUATION MODELS WITH APPLICATIONS TO BRAIN CONNECTIVITY STUDIES

BY TINGTING ZHANG[*,1,2], QIANNAN YIN[*], BRIAN CAFFO[†], YINGE SUN[*] AND DANA BOATMAN-REICH[†,1,3]

*University of Virginia * and Johns Hopkins University†*

We build a new ordinary differential equation (ODE) model for the directional interaction, also called effective connectivity, among brain regions whose activities are measured by intracranial electrocorticography (ECoG) data. In contrast to existing ODE models that focus on effective connectivity among only a few large anatomic brain regions and that rely on strong prior belief of the existence and strength of the connectivity, the proposed high-dimensional ODE model, motivated by statistical considerations, can be used to explore connectivity among multiple small brain regions. The new model, called the modular and indicator-based dynamic directional model (MIDDM), features a cluster structure, which consists of modules of densely connected brain regions, and uses indicators to differentiate significant and void directional interactions among brain regions. We develop a unified Bayesian framework to quantify uncertainty in the assumed ODE model, identify clusters, select strongly connected brain regions, and make statistical comparison between brain networks across different experimental trials. The prior distributions in the Bayesian model for MIDDM parameters are carefully designed such that the ensuing joint posterior distributions for ODE state functions and the MIDDM parameters have well-defined and easy-to-simulate posterior conditional distributions. To further speed up the posterior simulation, we employ parallel computing schemes in Markov chain Monte Carlo steps. We show that the proposed Bayesian approach outperforms an existing optimization-based ODE estimation method. We apply the proposed method to an auditory electrocorticography dataset and evaluate brain auditory network changes across trials and different auditory stimuli.

1. Introduction. This paper focuses on modeling and making inferences about directional interactions among human brain regions. The human brain is a continuous time dynamic system, so it is biophysically natural to use ordinary

Received June 2016; revised January 2017.

[1]Corresponding author.

[2]Supported by University of Virginia Quantitative Collaborative Seed grant.

[3]Supported by NIDCD Grant K24-DC010028, Army Research Organization Grant W911NF-12R001202, and Army Research Laboratory Grant W911NF-1020022.

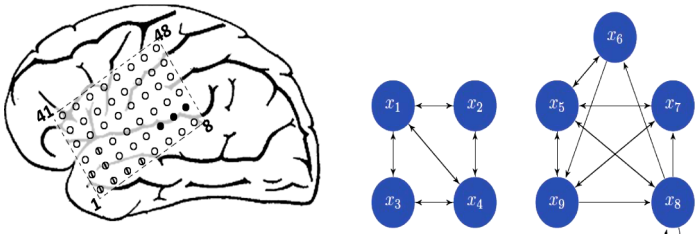
Key words and phrases. Bayesian inference, ODE models, cluster structure, directional brain networks, network edge selection.

differential equations (ODE) to model the directional effects exerted by each system component (i.e., regions) over others. More specifically, since studies have shown that interactions among brain regions occur at the neuronal level [Aertsen and Preissl (1991)], we use ODEs to model the brain's neuronal state changes and directional connectivity.

The dynamic causal models (DCM) for fMRI and EEG data are the most commonly cited ODE models in the literature [Daunizeau, David and Stephan (2011), David and Friston (2003), David et al. (2006), Friston, Harrison and Penny (2003), Kiebel, David and Friston (2006)] for directional interactions, also called effective connectivity, among brain regions. Both fMRI and EEG are noninvasive methods for measuring brain activity with large noise, hence associated DCMs rely on strong prior information of the existence and strength of connections among the brain regions under study, and the DCMs are usually focused on connectivity among only a few large anatomic regions. In contrast, we use electrocorticography (ECoG) data to study directional interactions among many small brain regions and evaluate how their interactions change across time and stimuli. As intracranial measurements of brain activity, ECoG provides unique information for studying the brain's directional connectivity, as explained below.

ECoG, also called intracranial EEG, uses electrodes placed directly on the cortical surface of the human brain to record its electrical activity for clinical purposes in the treatment of patients with medically intractable seizures or tumors. Figure 1(a) shows the spatial placement of ECoG electrodes on the epileptic patient whose ECoG data are analyzed in this paper. The special data collection technique leads to nice properties of ECoG data, including simultaneous recordings of many small brain regions' neuronal electrical activity, combined high temporal (data collected every 1 ms) and spatial (10 mm) resolution, strong signal-to-noise ratio (SNR), and highly reliable and reproducible measurements of brain activity [Cervenka et al. (2013)], which are unavailable in noninvasive measurements of brain activity, for example, fMRI and EEG.

We propose an ODE model for ECoG data, which is motivated by statistical considerations and widely applicable to explore directional connectivity among



(a) 6×8 electrode array (b) Cluster Structure

FIG. 1. (a) *Spatial placement of ECoG electrodes on an epileptic patient*. (b) *A network in a cluster structure*.

many different brain regions without relying on strong, specific prior knowledge of the regions. Specifically, we use bilinear ODEs to model directional interactions among brain regions for three reasons. First, the bilinear model, also the simplest ODE model as a low-order Taylor expansion of nonlinear ones, provides general applicability for approximating high-dimensional dynamic systems. This approach is similar to using linear regression models to approximate complex association relationships between various response and predictor variables. As such, linear or bilinear ODEs have been used to model complex dynamic systems in many scientific studies when the underlying dynamic mechanism is elusive and the ensuing detailed model specification is difficult. These studies include gene regulation network [Lu et al. (2011), Voit (2000)] and brain effective connectivity studies based on fMRI data [Friston (2009)]. Second, the simple bilinear form provides intuitive scientific interpretation of the model parameters, and enables fast computation for high-dimensional data. Third, taking advantage of ECoG's high temporal resolution, we study the brain activity in response to a simple, short auditory stimulus, and the proposed bilinear model can effectively approximate the brain dynamics within a short period of time.

Within the bilinear formulation, we assume that many model parameters denoting directional interactions among brain regions are zeroes. This is because interactions among brain regions are energy-consuming [Anderson (2005), Földiák and Young (1995), Olshausen and Field (2004)] and sparse connections would help the brain, a biological system, to conserve energy in order to survive and prosper [Bullmore and Sporns (2009), Micheloyannis (2012)]. Moreover, motivated by many reports of brain networks in a cluster structure [Milo et al. (2002, 2004), Newman (2006), Sporns (2011)], which consists of clusters, also called modules, of densely connected brain regions, as shown in Figure 1(b), we build a new bilinear ODE model, called modular and indicator-based dynamic directional model (MIDDM), to characterize sparse connections in the cluster structure in particular. This new ODE model not only has a scientific basis, but also provides intuitive interpretation of different functions of the brain regions in different modules.

Despite that an iterative optimization algorithm, called Potts-based iterated principal differential analysis (P-iPDA), has been developed by Zhang et al. (2015) to identify clusters within a bilinear Potts-based dynamic direction model (PDDM), we develop a new Bayesian approach to estimate the proposed MIDDM, an extension of the PDDM, for three major reasons. First, the Bayesian method provides a unified inference framework for evaluating the statistical significance of identified network edges, each associated with a significantly nonzero model parameter denoting the directional effect between a pair of brain regions. With the proposed Bayesian method, we can study the brain connectivity changes in response to repetitive events, an important research topic in neuroscience [Eliades et al. (2014), Garrido et al. (2009)]. In contrast, it is difficult to evaluate the statistical significance or compare network results of different trials by the P-iPDA (see Section 2.1

for more details). Second, the MIDDM, assuming different properties for connections within and between modules, essentially presents a hierarchical model for the brain network. It is natural and convenient to characterize this multilevel structure and to simultaneously address module identification and directional network edge selection within a unified Bayesian framework, similar to Bayesian methods [Dunson, Herring and Engel (2008), Kim, Tadesse and Vannucci (2006), Tadesse, Sha and Vannucci (2005)] for simultaneous variable selection and clustering in multiple regression. Third, quantification of the MIDDM model inadequacy for characterizing the complex brain system is natural within a Bayesian framework, an approach unique from most existing ODE estimating methods, as explained in detail below.

The proposed new ODE model, motivated by statistical considerations, is considered an approximation rather than a principle for the underlying brain mechanism, in contrast to many existing low-dimensional ODE models for simple, well-understood dynamic systems. As such, it is important to account for model uncertainty, defined as the discrepancy between the state functions of the assumed model and the true state functions of the complex brain system, when estimating the MIDDM. Model uncertainty quantification, though straightforward for standard statistical models, is not self-evident for ODEs, since the latter are essentially deterministic models for dynamic systems. Chkrebtii et al. (2016) and Conrad et al. (2015) developed approaches within the Bayesian framework by Kennedy and O'Hagan (2001) to quantify discretization uncertainty of ODE models, which is the discrepancy, caused by limited computation and coarse grids, between the state functions fitted by discretization methods and the exact state functions of the ODE model. However, very few methods in the statistical literature have been developed to quantify ODE model uncertainty. Existing approaches [Cheung et al. (2011), Oliver and Moser (2011)] for ODE model uncertainty quantification were mainly developed for specific low-dimensional dynamic systems. Here, we develop a new prior on high-dimensional ODE state functions to quantify the discrepancy between the assumed MIDDM and the underlying brain system.

In summary, this paper proposes a new ODE model for ECoG data to characterize the brain network of effective connectivity in a cluster structure, and develops a new Bayesian framework to quantify the ODE model uncertainty, identify clusters, select significant network edges, and evaluate brain network changes across time and stimulus types. Moreover, we carefully design new priors on the MIDDM parameters to ensure fast posterior simulation of the ensuing hierarchical Bayesian model for high-dimensional ECoG data.

The rest of the article is organized as follows. Section 2 introduces the new ODE model, MIDDM, for cluster-structured directional brain networks, and reviews existing methods for ODE model estimation and cluster identification. We develop a Bayesian hierarchical method to estimate the MIDDM based on basis representation of ODE state functions in Section 3, and develop a Markov chain Monte Carlo (MCMC) simulation algorithm for posterior inference in Section 4.

Then we apply the proposed MIDDM and Bayesian method to two simulated examples in Section 5, comparing the results by the Bayesian method with those by the existing ODE estimation method P-iPDA, and demonstrating the advantages of the former over the latter. In Section 6, we analyze ECoG data collected in an auditory experiment, and evaluate brain network changes across trials and auditory stimuli. The analysis results not only confirm existing results, but also bring new insights into understanding brain connectivity changes in response to repetitive events. Section 7 discusses future development of ODE models and Bayesian methods for ECoG data analysis.

2. MIDDM for ECoG data. Let $\mathbf{y}(t) = (y_1(t), \ldots, y_d(t))'$ be the observed ECoG measurements of d brain regions' neuronal activity at time t. The observed data $\mathbf{y}(t)$ are measured at discrete time points $t = 1, 2, \ldots, T$.

Let $\mathbf{x}(t) = (x_1(t), \ldots, x_d(t))'$ be the neuronal state functions of the d brain regions at time t. For ECoG data, the observation model is given by

$$\mathbf{y}(t) = \mathbf{x}(t) + \boldsymbol{\varepsilon}(t), \tag{2.1}$$

where $\boldsymbol{\varepsilon}(t) = (\varepsilon_1(t), \ldots, \varepsilon_d(t))'$ is a d-dimensional vector of errors with mean zeroes.

The brain system consisting of the d regions under study received a stimulus input. We let $u(t)$ be an experimental input function taking values 1 and 0 only, for example, a boxcar or stick stimulus function. The input function indicates whether the stimulus is present at time t. Since brain regions interact with each other at the neuronal level, we use the following bilinear ODEs for $\mathbf{x}(t)$ to characterize directional interactions among the d regions:

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{A}\mathbf{x}(t)\,\big(1 - u(t)\big) + \mathbf{B}\mathbf{x}(t)\,u(t) + \mathbf{C}u(t) + \mathbf{D}, \tag{2.2}$$

where $\mathbf{A} = (A_{ij})_{d\times d}$ with entry A_{ij} denoting the effect of region j on region i exerted at the current state without the stimulus; $\mathbf{B} = (B_{ij})_{d\times d}$ with B_{ij} denoting the stimulus-dependent effect exerted by region j on region i; $\mathbf{C} = (C_1, \ldots, C_d)$ with C_i denoting the stimulus effect on region i; and $\mathbf{D} = (D_1, \ldots, D_d)$ denoting the intercepts for the d regions. As a low-order Taylor approximation of the underlying system, the simple bilinear model (2.2) provides general applicability and facilitates fast computation for a high-dimensional dynamic system.

As discussed in Introduction, it is reasonable to believe that many brain regions are not directly connected and the ensuing sparse brain network with each network edge denoting one directional effect between a pair of regions is in a cluster structure. To characterize the cluster structure, we introduce module labels $\mathbf{m} = \{m_1, \ldots, m_d\}$ for d regions, which take integer values between 1 and d. In addition, following the formulation in the Bayesian stochastic search variable selection (SSVS) framework [Brown, Vannucci and Fearn (1998), George and McCulloch (1993, 1997), Yi, George and Allison (2003)] that uses the "spike and

slab" prior [Ishwaran and Rao (2005), Miller (2002), Theo and Mike (2004)], we use indicators γ_{ij}^A and γ_{ij}^B—which take values 1 and 0 only for $i, j = 1, \ldots, d$—to differentiate strong and void directional effects without and with the stimulus, respectively. We modify Model (2.2) and propose the following ODE model:

$$\begin{aligned}\frac{dx_i(t)}{dt} &= \sum_{j=1}^{d} \delta(m_i, m_j)\, \gamma_{ij}^A\, A_{ij}\, x_j(t)\, \big(1 - u(t)\big) \\ &\quad + \sum_{j=1}^{d} \delta(m_i, m_j)\, \gamma_{ij}^B\, B_{ij}\, x_j(t)\, u(t) + C_i\, u(t) + D_i,\end{aligned} \tag{2.3}$$

where $\delta(m_i, m_j)$ is the Kronecker delta, which equals 1 whenever $m_i = m_j$ and 0 otherwise. Under Model (2.3), component j has a nonzero directional effect on i or a directional network edge from j to i exists if and only if the two components are in the same cluster, that is, $m_i = m_j$, and either γ_{ij}^A or γ_{ij}^B is nonzero.

The ODE (2.3) together with the observation model (2.1) is referred to as the MIDDM. The ODE (2.2) and the sparse network, in which each node is sparsely connected with the rest of nodes and all nodes are connected directly or indirectly, are special cases of the MIDDM with only one single module and with either mostly nonzero or mostly zero indicators. The MIDDM is also an extension of the existing ODE model for ECoG, PDDM Zhang et al. (2015), as the latter assumes all regions within clusters to be pairwise connected while the former uses indicators to distinguish nonzero directional interactions within clusters from void ones.

Under the MIDDM, the inference of the brain's effective connectivity is equivalent to identifying modules, selecting statistically significant directional connections, and estimating the model parameters $\mathbf{A}$ and $\mathbf{B}$, which denote the strength of effective connectivity among brain regions.

2.1. *Existing ODE model estimation methods.* In the statistical literature, three major approaches have been developed for estimating ODE models: basis-function-expansion approaches in which state functions $\mathbf{x}(t)$ are represented by functional bases [Bhaumik and Ghosal (2015), Brunel (2008), Deuflhard and Bornemann (2002), Poyton et al. (2006), Qi and Zhao (2010), Ramsay and Silverman (2005), Ramsay et al. (2007), Varah (1982)], discretization methods using numerical approximation [Bard (1974), Biegler, Damiano and Blau (1986), Campbell (2007), Cao, Huang and Wu (2012), Gelman, Bois and Jiang (1996), Girolami (2008), Hemker (1972), Huang, Liu and Wu (2006), Huang and Wu (2006), Li, Osborne and Prvan (2005), Mattheij and Molenaar (2002), Xue, Miao and Wu (2010)], and Bayesian approaches using a Gaussian process prior for state functions [Calderhead, Girolami and Lawrence (2008), Chkrebtii et al. (2016), Stuart (2010)].

The approaches mentioned above usually concern low-dimensional dynamic systems with only a few ODEs. For high-dimensional systems, Lu et al. (2011), Wu et al. (2014a, 2014b) proposed to use penalization-based variable selection methods, which were originally developed for high-dimensional regression problems [Fan and Li (2001), Tibshirani (1996), Wang and Leng (2008), Yuan and Lin (2006), Zou (2006), Zou and Hastie (2005)], to estimate sparse ODEs. These methods are suitable for identifying sparse networks in which each component has only a few connections with other components and all components are connected directly or indirectly.

An optimization algorithm P-iPDA was developed by Zhang et al. (2015) to search for the optimal modules/clusters of densely connected brain regions. However, the P-iPDA has several limitations. First, the P-iPDA, relying on a crucial assumption that regions within the same cluster are all pairwise connected, does not distinguish connected brain regions from disconnected ones within the same cluster, and thus may lead to many false positives whenever the P-iPDA fails to identify the smallest clusters and groups separate clusters together. Second, the P-iPDA essentially minimizes a log-likelihood based criterion with an L_0 penalty; statistical inference of the ensuing parameter estimates is challenging, and thus the P-iPDA cannot be used to compare brain networks across different trials and stimulus types. Third, the network results by the P-iPDA are highly sensitive to the used penalty parameters. As such, the P-iPDA needs to perform a time-consuming cross-validation procedure on quite a few candidate values to select ideal penalty parameters. In contrast, the proposed Bayesian method can address all the three limitations of the P-iPDA, as explained in the following sections.

3. Hierarchical Bayesian model for MIDDM. We propose a Bayesian method to make inferences about the MIDDM. We elaborate on building the hierarchical Bayesian model in the following section, and present an MCMC simulation algorithm for posterior inference in Section 4.

3.1. *Model construction.* First, represent the state function $x_i(t)$ of each component i by a vector of B-spline functions $\mathbf{b}(t) = (b_1(t), \ldots, b_L(t))'$ defined on an equally spaced partition $\{t_1 = 1, t_2, \ldots, t_q = T\}$ of the interval $[1, T]$:

$$x_i(t) = \mathbf{b}(t)'\eta_i, \tag{3.1}$$

where η_i is an $L \times 1$ vector consisting of the basis coefficients of $x_i(t)$.

Nonparametric model for observed data. Let $Y_i = (y_i(1), y_i(2), \ldots, y_i(T))'$ and $\mathbf{Y} = (Y_1', Y_2', \ldots, Y_d')'$. With representation (3.1), we assume that the Y_is are independently distributed with multivariate normal distributions:

$$Y_i|\eta_i \overset{\text{ind}}{\sim} \text{MN}(\mathbf{\Phi}\eta_i, \sigma_i^2\mathbf{I}_T) \qquad \text{for } i = 1, 2 \ldots, d, \tag{3.2}$$

where $\mathrm{MN}(\mu, \Omega)$ stands for a multivariate normal distribution with mean μ and variance matrix Ω, and $\mathbf{\Phi}$ is a $T \times L$ matrix with element $\mathbf{\Phi}[t, l] = b_l(t)$ for $t = 1, 2, \ldots, T$ and $l = 1, 2, \ldots, L$. Though an AR(1) or AR(2) model can be assumed for $\varepsilon_i(t)$, $t = 1, 2, \ldots, T$, in the observation model (2.1), for simplicity we assume them to be independently Gaussian distributed with zero mean. For data with a strong SNR, such as ECoG [Boatman-Reich et al. (2010), Bressler and Ding (2002), Zhang et al. (2015)], accounting for autocorrelation in the model does not improve estimation much.

Prior specification for basis coefficients. Let $\boldsymbol{\gamma}^A = \{\gamma_{ij}^A, i, j = 1, 2, \ldots, d\}$, $\boldsymbol{\gamma}^B = \{\gamma_{ij}^B, i, j = 1, 2, \ldots, d\}$, $\boldsymbol{\theta} = \{\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}\}$, and $\mathbf{\Theta}_I$ denote all the MIDDM parameters:

$$\mathbf{\Theta}_I = \{\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}, \mathbf{m}, \boldsymbol{\gamma}^A, \boldsymbol{\gamma}^B\}.$$

We propose a prior for vectorized basis coefficients $\boldsymbol{\eta} = (\eta_1', \ldots, \eta_d')'$, which depends on the MIDDM parameters $\mathbf{\Theta}_I$ through the MIDDM model-fitting error:

$$p(\boldsymbol{\eta}|\mathbf{\Theta}_I, \tau) \propto \exp\left\{-\frac{1}{2\tau}\mathrm{R}(\boldsymbol{\eta}, \mathbf{\Theta}_I)\right\}, \tag{3.3}$$

where τ is a pre-specified positive constant and $\mathrm{R}(\boldsymbol{\eta}, \mathbf{\Theta}_I)$ is the MIDDM model fitting error:

$$\mathrm{R}(\boldsymbol{\eta}, \mathbf{\Theta}_I) = \sum_{i=1}^{d} \int_0^T \left[\frac{dx_i(t)}{dt} - \sum_{j=1}^{d} \delta(m_i, m_j)\, \gamma_{ij}^A\, A_{ij}\, x_j(t)\, (1 - u(t)) \right.$$
$$\left. - \sum_{j=1}^{d} \delta(m_i, m_j)\, \gamma_{ij}^B\, B_{ij}\, x_j(t)\, u(t) - C_i\, u(t) - D_i \right]^2 dt.$$

Note that $x_i(t)$, $i = 1, \ldots, d$, in $\mathrm{R}(\boldsymbol{\eta}, \mathbf{\Theta}_I)$ is represented by basis functions as in (3.1), and $\frac{dx_i(t)}{dt} = \mathbf{b}^{(1)}(t)'\eta_i$.

The prior (3.3) quantifies the deviation of the state functions from the assumed ODE model by a probability measure and suggests a preference for state functions with a small model-fitting error given $\mathbf{\Theta}_I$.

With the linear basis representation of $\mathbf{x}(t)$, $\mathrm{R}(\boldsymbol{\eta}, \mathbf{\Theta}_I)$ given $\mathbf{\Theta}_I$ is quadratic of $\boldsymbol{\eta}$:

$$\mathrm{R}(\boldsymbol{\eta}, \mathbf{\Theta}_I) = \boldsymbol{\eta}'\mathbf{\Omega}_{\mathbf{\Theta}_I}\boldsymbol{\eta} - 2\mathbf{\Lambda}_{\mathbf{\Theta}_I}'\boldsymbol{\eta} + \Xi_{\mathbf{\Theta}_I}, \tag{3.4}$$

where $\mathbf{\Omega}_{\mathbf{\Theta}_I}$, $\mathbf{\Lambda}_{\mathbf{\Theta}_I}$, and $\Xi_{\mathbf{\Theta}_I}$, respectively, are a $dL \times dL$ matrix, a $dL \times 1$ vector, and a scalar, whose values depend on the MIDDM parameters $\mathbf{\Theta}_I$. As such, the prior (3.3) is equivalent to a normal distribution:

$$\boldsymbol{\eta}|\mathbf{\Theta}_I, \tau \sim \mathrm{MN}(\mathbf{\Omega}_{\mathbf{\Theta}_I}^{-1}\mathbf{\Lambda}_{\mathbf{\Theta}_I}, \tau\, \mathbf{\Omega}_{\mathbf{\Theta}_I}^{-1}),$$

and the prior (3.3) on basis coefficients $\boldsymbol{\eta}$ is equivalent to a Gaussian process prior on the state functions $\mathbf{x}(t)$.

The exact formulas of $\boldsymbol{\Omega}_{\boldsymbol{\Theta}_I}$, $\boldsymbol{\Lambda}_{\boldsymbol{\Theta}_I}$, and $\boldsymbol{\Xi}_{\boldsymbol{\Theta}_I}$, as functions of $\boldsymbol{\Theta}_I$, are provided in the Appendix B.6.

Prior specification for MIDDM parameters. We specify a joint prior for MIDDM parameters $\boldsymbol{\Theta}_I$:

$$
\begin{aligned}
p(\boldsymbol{\Theta}_I|\tau) \propto{}& \det(\Omega_{\boldsymbol{\Theta}_I})^{-1/2} \exp\left\{\frac{1}{2\tau}\left(\boldsymbol{\Lambda}'_{\boldsymbol{\Theta}_I}\boldsymbol{\Omega}^{-1}_{\boldsymbol{\Theta}_I}\boldsymbol{\Lambda}_{\boldsymbol{\Theta}_I} - \Xi_{\boldsymbol{\Theta}_I}\right)\right\} \\
&\times \exp\left\{-\mu \sum_{i,j=1}^{d} \delta(m_i, m_j)\right\} p_0^{\sum_{i,j}\gamma^A_{ij}+\sum_{i,j}\gamma^B_{ij}} \\
&\times (1-p_0)^{2d^2-\sum_{i,j}\gamma^A_{ij}-\sum_{i,j}\gamma^B_{ij}} \\
&\times \prod_{i,j=1}^{d} \phi\left(\frac{A_{ij}}{\xi_0}\right) \prod_{i,j=1}^{d} \phi\left(\frac{B_{ij}}{\xi_0}\right) \prod_{i=1}^{d} \phi\left(\frac{C_i}{\xi_0}\right) \prod_{i=1}^{d} \phi\left(\frac{D_i}{\xi_0}\right),
\end{aligned}
\tag{3.5}
$$

where the prior probability p_0 is pre-specified by the user to input the prior belief of the average degree of connections within clusters, μ is a given nonnegative constant, ϕ is the standard normal density function, and ξ_0 is a large positive constant to give an almost flat prior for $\boldsymbol{\theta}$ in a wide domain. It is possible to use different p_0 for $\boldsymbol{\gamma}^A$ and $\boldsymbol{\gamma}^B$. For simplicity, we assume identical prior probabilities for them. The discussion of choosing μ and p_0 is deferred to Section 4.2.

The above prior essentially puts together a multivariate Bernoulli distribution [George and McCulloch (1997)] for indicators $\boldsymbol{\gamma}^A$ and $\boldsymbol{\gamma}^B$, the Potts model [Graner and Glazier (1992), Potts (1952)] for module labels $\mathbf{m}$, and almost flat prior distributions for parameters $\boldsymbol{\theta}$. We specify the prior for MIDDM parameters in the above form, a main thrust of the proposed Bayesian framework, for two reasons. First, using the prior (3.5), the full posterior conditional distributions of parameters $\boldsymbol{\eta}$ and $\boldsymbol{\theta}$ are multivariate normal, which are easy to simulate from; this is a crucial advantage of the proposed Bayesian framework, especially for analyzing high-dimensional ECoG data. Second, in contrast to independent priors on the MIDDM parameters, the prior (3.5) makes use of the proposed model information by incorporating the model-fitting error. A similar prior has been proposed by Yuan and Lin (2005) in an empirical Bayes approach for variable selection and estimation in linear models.

Priors for data variances. We impose an uninformative prior on $\boldsymbol{\sigma}^2 = \{\sigma_i^2, i = 1, 2, \ldots, d\}$:

$$
p(\boldsymbol{\sigma}^2) \propto \prod_{i=1}^{d} 1/\sigma_i^2, \qquad \text{for } i = 1, \ldots, d. \tag{3.6}
$$

Joint posterior distribution. In summary, equations (3.2), (3.3), (3.5), and (3.6) jointly define a hierarchical Bayesian model for the MIDDM, referred to as Bayesian MIDDM (BMIDDM) in the following. The joint posterior distribution of the BMIDDM is given by

$$
\begin{aligned}
&p(\boldsymbol{\eta}, \boldsymbol{\Theta}_I, \boldsymbol{\sigma}^2 | \mathbf{Y}, \tau, \mu) \\
&\quad \propto \prod_{i=1}^{d} \frac{1}{\sigma_i^T} \exp\left\{ -\frac{(Y_i - \boldsymbol{\Phi}\eta_i)^2}{2\sigma_i^2} \right\} \exp\left\{ -\frac{1}{2\tau} \mathrm{R}(\boldsymbol{\eta}, \boldsymbol{\Theta}_I) \right\} \\
&\quad \times \exp\left\{ -\mu \sum_{i,j=1}^{d} \delta(m_i, m_j) \right\} p_0^{\sum_{i,j} \gamma_{ij}^A + \sum_{i,j} \gamma_{ij}^B} \\
&\quad \times (1 - p_0)^{2d^2 - \sum_{i,j} \gamma_{ij}^A - \sum_{i,j} \gamma_{ij}^B} \\
&\quad \times \prod_{i,j=1}^{d} \phi\left(\frac{A_{ij}}{\xi_0} \right) \prod_{i,j=1}^{d} \phi\left(\frac{B_{ij}}{\xi_0} \right) \prod_{i=1}^{d} \phi\left(\frac{C_i}{\xi_0} \right) \prod_{i=1}^{d} \phi\left(\frac{D_i}{\xi_0} \right) \prod_{i=1}^{d} \frac{1}{\sigma_i^2}.
\end{aligned} \tag{3.7}
$$

It can be shown that with a fixed positive constant τ and nonnegative μ, the above posterior is proper as long as $T > L$. The proof is provided in the Appendix A.

4. Posterior simulations. Let $\mathbf{m}_{-i} = \mathbf{m} \setminus \{m_i\}$, $\boldsymbol{\gamma}_{-ij}^A = \boldsymbol{\gamma}^A \setminus \{\gamma_{ij}^A\}$, and $\boldsymbol{\gamma}_{-ij}^B = \boldsymbol{\gamma}^B \setminus \{\gamma_{ij}^B\}$ for $i, j = 1, 2, \ldots, d$. We use a partially collapsed Gibbs Sampler [PCGS; van Dyk and Park (2008)] to sample from (3.7) with given μ and τ (omitted in the posterior conditional distributions below). Specifically, $\boldsymbol{\theta}$ is integrated out when drawing posterior samples of $\mathbf{m}$ and indicators $\boldsymbol{\gamma}^A$ and $\boldsymbol{\gamma}^B$, and the PCGS is performed in the following order to maintain the target stationary distribution:

1. Sequentially update m_i by a draw from $p(m_i | \mathbf{m}_{-i}, \boldsymbol{\eta}, \boldsymbol{\sigma}^2, \boldsymbol{\gamma}^A, \boldsymbol{\gamma}^B, \mathbf{Y})$ for $i = 1, 2, \ldots, d$.
2. Sequentially update γ_{ij}^A by a draw from $p(\gamma_{ij}^A | \mathbf{m}, \boldsymbol{\eta}, \boldsymbol{\sigma}^2, \boldsymbol{\gamma}_{-ij}^A, \boldsymbol{\gamma}^B, \mathbf{Y})$ for $i, j = 1, 2, \ldots, d$.
3. Sequentially update γ_{ij}^B by a draw from $p(\gamma_{ij}^B | \mathbf{m}, \boldsymbol{\eta}, \boldsymbol{\sigma}^2, \boldsymbol{\gamma}^A, \boldsymbol{\gamma}_{-ij}^B, \mathbf{Y})$ for $i, j = 1, 2, \ldots, d$.
4. Draw $\boldsymbol{\theta}$ from $p(\boldsymbol{\theta} | \mathbf{m}, \boldsymbol{\eta}, \boldsymbol{\sigma}^2, \boldsymbol{\gamma}^A, \boldsymbol{\gamma}^B, \mathbf{Y})$, which is a multivariate normal distribution.
5. Draw $\sigma_1^2, \ldots, \sigma_d^2$ from $p(\boldsymbol{\sigma}^2 | \boldsymbol{\Theta}_I, \boldsymbol{\eta}, \mathbf{Y})$, which is a product of independent inverse-gamma distributions.
6. Draw $\boldsymbol{\eta}$ from $p(\boldsymbol{\eta} | \boldsymbol{\Theta}_I, \boldsymbol{\sigma}^2, \mathbf{Y})$, which is a multivariate normal.

We defer technical derivations of posterior $p(\mathbf{m}, \boldsymbol{\eta}, \boldsymbol{\gamma}^A, \boldsymbol{\gamma}^B, \boldsymbol{\sigma}^2 | \mathbf{Y}, \tau, \mu)$ with $\boldsymbol{\theta}$ being integrated out and the posterior conditional distributions of each parameter to the Appendix B.

4.1. *Parallel computing*. To speed up MCMC posterior simulations of the BMIDDM, we employ a parallel computing scheme similar to that developed by Caffo et al. (2011) in three major MCMC steps: (a) simulation of basis coefficients η_is for regions in different clusters, (b) calculation of the posterior conditional probabilities of m_i, and (c) simulation from the posterior conditional distributions of indicators $\boldsymbol{\gamma}^A$ and $\boldsymbol{\gamma}^B$. Specifically, in (a), since given $\mathbf{m}$, η_is of brain regions in different clusters are conditionally independent, we employ the same number of process cores as the number of different clusters of $\mathbf{m}$, and use each core to simulate η_is in one unique cluster. In (b), we use the same number of process cores as the number of different values that m_i can take given the rest of the parameters, with each core computing one posterior conditional probability for m_i taking one unique value. In (c), since indicator variables $\tilde{\boldsymbol{\gamma}}_i = \{\gamma_{ij}^A, \gamma_{ij}^B, j = 1, 2, \ldots, d\}$, conditional on the rest of the parameters, are independent for $i = 1, 2, \ldots, d$, we employ d process cores, each sequentially simulating every element of one $\tilde{\boldsymbol{\gamma}}_i$ from the element's posterior conditional probability.

The reduction of computational time by using parallel computing in Steps (a) and (b) depends mostly on the size of the largest cluster at each iteration, and parallel computing is most efficient when all the clusters are small and of similar sizes. The use of parallel computing in Step (c) reduces the computational time for the posterior simulation of indicators from $O(d^2)$ to $O(d)$.

Parallel computing can also be used in other MCMC steps in a similar manner, including simulation of MIDDM parameters $\boldsymbol{\theta}$ of regions in different clusters and simulation of σ_i^2 for $i = 1, 2, \ldots, d$.

4.2. *Hyperparameter selection*. The choice of hyperparameter τ is most crucial, because it balances between the data and model information for inferring directional connections among brain regions. Specifically, a small τ, compared to the data variance, can impose an incorrect strong prior belief that the assumed model fits the underlying dynamic system well. We found that with a small τ, only a few regions' temporal activity can be jointly fitted by the assumed model, and thus only a few brain regions are identified to be connected. On the other hand, if τ is too large, the model information in the posterior is too weak to be useful for differentiating strong directional connections from weak or void ones. Then all the brain regions are identified to be connected. In summary, an appropriate value of τ depends on how well the assumed ODE model can fit the data.

Given the above consideration, we decide to choose τ based on the model-fitting error of the observed data. Though cross-validation-based methods [Reiss and Ogden (2007) (2007, 2009), Wahba (1990)] are straightforward for choosing hyperparameters, they are time consuming within a Bayesian framework. Instead, we propose an easy-to-implement approach to determine the value of hyperparameter τ. Since τ can be regarded as the variance of ODE fitting errors, we set it to the variance of estimated ODE fitting errors. Specifically, we fit $\mathbf{x}(t)$ nonparametrically with $\mathbf{b}(t)$ to the observed data; regress estimated $d\hat{x}_i(t)/dt$ versus

$\hat{\mathbf{x}}(t)\ (1-u(t))$ and $\hat{\mathbf{x}}(t)\ u(t)$; and obtain the regression mean squared error, denoted by $\hat{\tau}_i$, for $i=1,\ldots,d$. The range of $\hat{\tau}_i$, $i=1,\ldots,d$, gives the range of the variances of model fitting errors for the observed data. Then we choose τ to be $\max\{\hat{\tau}_i\}_{i=1}^d$, which leads to the least informative prior for basis coefficients among all the candidate values. Through simulation and real data analysis, we found that this value can help us effectively identify the underlying modules, select true network edges, and provide scientifically interpretable results.

We let $\mu=0$ to give a noninformative prior on the cluster structure $\mathbf{m}$. For choosing the prior probability p_0 for nonzero network edges within modules, we have tried different p_0 values, from 0.9 to 0.7, which reflect the prior belief that regions within the same cluster are densely connected. We evaluated the network edge selection performance through simulation studies, and found that $p_0=0.9$ is most effective with the highest power for selecting network edges, especially for the cluster structure where all regions within the same cluster are pairwise connected. Considering that the connections within modules are usually short-range, strong, and dense [Park and Friston (2013)], we let $p_0=0.9$ to ensure a high power for selecting within-module network edges. Though it is possible to assign a prior to p_0 or tune its value based on the data, we choose to specify its value, because this approach directly uses the existing scientific knowledge of the brain network and thus reduces uncertainty in the model estimation.

The choice of the number of bases L directly affects the posterior computational time: The larger L, the more computational time needed for simulating the state functions, which is the most computationally intensive MCMC step. Considering this, we choose a small L without compromising the flexibility of representing the state function $\mathbf{x}(t)$, as suggested in Ramsay et al. (2007). For the real data under study, we found that data at every three consecutive points take similar values, and thus we chose $L=\lceil T/3\rceil$.

5. Simulations. Given hyperparameters (μ,τ) and data $\mathbf{y}(t)$, we conduct posterior simulations of the BMIDDM. Let S be the total number of MCMC iterations excluding the burn-in time, and $\theta^{(s)}$ be the value of BMIDDM parameter θ simulated at the sth iteration. Based on the posterior draws of BMIDDM parameters, for each pair of regions (i,j), we estimate the posterior clustering probability of the two regions being in the same cluster and the posterior probabilities of nonzero directional effects exerted by region j on i, also called the posterior selection probabilities of directional network edges from j to i, without and with the stimulus, by $\hat{P}_{ij}^m=\frac{1}{S}\sum_{s=1}^S\delta(m_i^{(s)},m_j^{(s)})$, $\hat{P}_{ij}^A=\frac{1}{S}\sum_{s=1}^S\delta(m_i^{(s)},m_j^{(s)})\ (\gamma_{ij}^A)^{(s)}$, and $\hat{P}_{ij}^B=\frac{1}{S}\sum_{s=1}^S\delta(m_i^{(s)},m_j^{(s)})\ (\gamma_{ij}^B)^{(s)}$, respectively.

We use $\hat{P}_{ij}^m$ to identify clusters of components and $\hat{P}_{ij}^A$ with $\hat{P}_{ij}^B$ to select directional network edges. Specifically, for module identification, we first rank $\hat{P}_{ij}^m$ for all $i,j=1,2,\ldots,d$ and select a set of pairs of components $\mathcal{S}=\{(i,j):$

$\text{rank}(\hat{P}_{ij}^m) > \mathbf{h}, i, j = 1, 2, \ldots, d\}$ for some predetermined threshold $\mathbf{h}$, for example, a top 5% rank. Given $\mathcal{S}$, we identify modules $\mathcal{C}_k$, $k = 1, 2, \ldots, K$, also a partition, of the d components, such that for any two components i and j in the same module, there exists a set of pairs $\{(k_0 = i, k_1), (k_1, k_2), \ldots, (k_{l-1}, k_l = j)\}$, which is a subset of $\mathcal{S}$. By using a high threshold $\mathbf{h}$, this procedure identifies clusters of regions that have high posterior probabilities of being connected directly or indirectly.

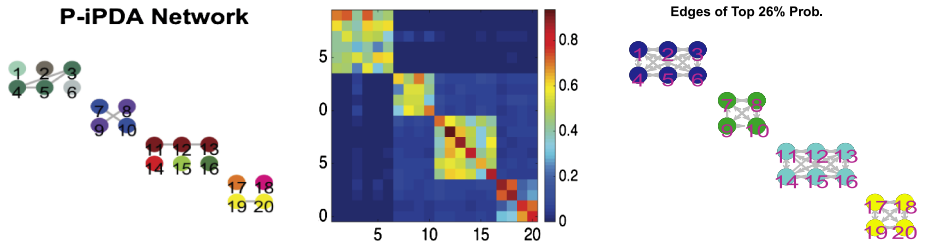
For network edge selection, which is performed after the module identification with a threshold $\mathbf{h}$, we first set $\hat{P}_{ij}^A$, $\hat{P}_{ji}^A$, $\hat{P}_{ij}^B$, and $\hat{P}_{ji}^B$ for i and j in two different clusters to zeroes. Then we order all the $\hat{P}_{ij}^A$ and $\hat{P}_{ij}^B$ for $i, j = 1, 2, \ldots, d$ and select network edges whose $\hat{P}_{ij}^A$ and $\hat{P}_{ij}^B$ have ranks higher than the threshold $\mathbf{h}$.

Similar to Bayesian variable selection for the linear regression, we use the ROC curve to summarize the performance of the proposed network edge selection procedure, which is computed as follows: for each given threshold, the percentages of true directional edges and null directional edges whose posterior probabilities are greater than the threshold are calculated as true positive rate (TPR) and false positive rate (FPR) of the selection procedure; the ROC curve summarizes pairs of TRPs and FPRs for different thresholds. The higher the ROC curve, that is, the larger TPR for each FPR, the better performance of the selection procedure.

In Section 5.1, we apply the proposed BMIDDM to one simulated dataset in Zhang et al. (2015), where components within the same cluster are all pairwise connected, and compare the results with those by the P-iPDA. We generate another time series data of the same dimension in Section 5.2 and use the BMIDDM to detect difference in the networks of these two simulated examples.

5.1. *Example* 1: *Data from a bilinear model.* The simulated dynamic system has 4 clusters of size 6, 4, 6, and 4. We let $T = 250$ and $u(t) = 1$ for $100 \le t \le 150$ and 0 otherwise, which are identical to those of the real data. For simplicity, parameter $\mathbf{B}$ is twice of $\mathbf{A}$. We generated 20 time series $\mathbf{x}(t)$ by using numerical approximation based on discretized bilinear model (2.3) with given parameters $\boldsymbol{\Theta}_I$, and generated 20 independent error time series $\boldsymbol{\varepsilon}(t)$, each following an AR(1) model with a lag-one correlation of 0.5. The SNR—defined as $\text{var}(x_i(t))/\text{var}(\varepsilon_i(t))$—of each time series $y_i(t)$, the sum of $x_i(t)$ and $\varepsilon_i(t)$, was set at 10. Before applying the proposed Bayesian approach, we standardized the observed time series to unit variance, such that time series of different components are in the same scale. In the following, $\mathbf{y}(t)$ is referred to the standardized data.

We applied the proposed BMIDDM with $\mu = 0$ and $\tau = \max_{i=1}^d \hat{\tau}_i$ to one simulated dataset for which the P-iPDA failed to identify all the clusters. The network by the P-iPDA is shown in Figure 2(a), which has only 36.5% TRP. In contrast, the BMIDDM identified all the nonzero directional effects with zero FPR, as shown in Figure 2(c), in which directional edges are corresponding to the posterior probabilities $\hat{P}_{ij}^B$ with top 26% ranks, the exact percentage of true edges among all



(a) Network by P-iPDA (b) Sel. Prob. with Stimulus (c) Network with Stimulus

FIG. 2. (a) *The network result by the P-iPDA in Example* 1. (b) *Shows* $\hat{P}_{ij}^{B}$ *for* $i, j = 1, \ldots, 20$. (c) *Shows network edges with top* 26% (*i.e., the percentage of true directional edges among all possible ones*) *posterior selection probabilities under the stimulus. Nodes in the same color correspond to regions in the same cluster identified by the used method, either the P-iPDA or the Bayesian method.*

possible ones. In this example, the posterior selection probabilities of true network edges and null ones have a sharp difference, as illustrated by Figure 2(b) of $\hat{P}_{ij}^{B}$ for $i, j = 1, \ldots, d$. The plot of $\hat{P}_{ij}^{A}$ is similar and not shown here.

The better performance of the Bayesian method than the P-iPDA for this example is possibly due to three reasons. First, though components within the same cluster are indeed pairwise connected, the average degree of nodes, however, in the network is small. Consequently, the BMIDDM, allowing for more sparsity, has a better selection efficiency. Second, the P-iPDA, an optimization algorithm for an L_0 penalized criterion, very likely outputs a network corresponding to a local mode of the criterion, especially since the ODE model estimation is sensitive to noise. In contrast, the Bayesian method, evaluating the posterior probabilities of different cluster structures, outputs a posterior "average," and thus is more stable. Third, the performance of the P-iPDA is sensitive to the choice of penalty parameters. It is possible that the penalty parameters used by the P-iPDA are not optimal for the presented dataset, while the Bayesian method using $\tau = \max_{i=1}^{d} \hat{\tau}_i$ is more adaptive to the specific data being analyzed, and thus has a better performance.

We have compared the computational complexities and times of the proposed Bayesian method with the P-iPDA. MCMC simulations of module labels $\mathbf{m}$ and basis coefficients $\boldsymbol{\eta}$ from their posterior conditional distributions, the two most time-consuming steps in the PCGS, have similar computational complexity as the two optimization steps updating $\mathbf{m}$ and $\boldsymbol{\eta}$ in the P-iPDA, because they all require matrix inversion of large matrices of similarly high dimensions. In particular, conditional on the same cluster structure $\mathbf{m}$ in the previous step, the computational complexity in the following iteration of the P-iPDA and PCGS is the same.

As searching for a mode is computationally much easier than exploring the entire posterior distribution, the P-iPDA takes much fewer iterations to converge than the PCGS. However, the P-iPDA relies on computationally extensive cross-validation, calculating prediction errors for every left-out data point and for a large

number of candidate penalty parameter values, to find the best combination of penalty parameters with the smallest prediction error. Specifically, for the simulated dynamic system in Example 1, at least 6 cross-validations at 50 left-out points were needed to find the optimal penalty parameters. One usually needs to perform 35 iterations of the P-iPDA for each left-out point given one combination of penalty parameters. As such, the total number of iterations needed for penalty parameter selection is 10,500, close to the number of iterations in the PCGS. For dynamic systems of larger dimensions and with longer time series, more iterations of the P-iPDA and cross-validation on more candidate penalty parameters using more left-out points are needed, because the number of potential cluster structures is larger and the network result by the P-iPDA is more sensitive to the choice of penalty parameters. Overall, if accounting for the penalty parameter selection time, the PCGS and P-iPDA use similar amounts of computational time. For the simulation example under study, it took 1.2 hours for the proposed Bayesian method to finish 10,000 MCMC iterations on a personal laptop using one i7 core.

We let $E_{ij} = \delta(m_i, m_j)\ \gamma_{ij}^A\ A_{ij}$ and $G_{ij} = \delta(m_i, m_j)\ \gamma_{ij}^B\ B_{ij}$. With S posterior draws of BMIDDM parameters, for each pair of regions i and j, we estimated the directional effects E_{ij} and G_{ij} exerted by region j over i without and with the stimulus, respectively, by their posterior means, which are given by

$$\hat{E}_{ij} = \frac{1}{S}\sum_{s=1}^{S}(\gamma_{ij}^A)^{(s)}\ \delta\big(m_i^{(s)}, m_j^{(s)}\big)\ A_{ij}^{(s)}$$

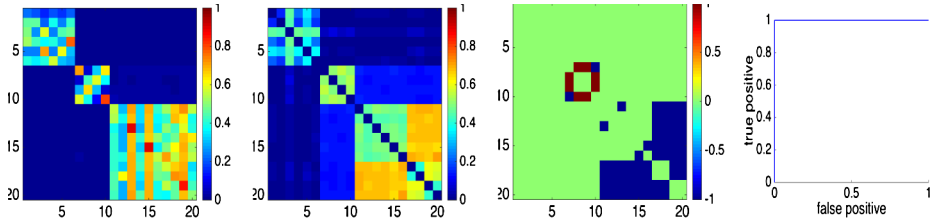
and

$$\hat{G}_{ij} = \frac{1}{S}\sum_{s=1}^{S}(\gamma_{ij}^B)^{(s)}\ \delta\big(m_i^{(s)}, m_j^{(s)}\big)\ B_{ij}^{(s)}.$$

Then we evaluated the mean squared errors (MSE) of $\mathbf{E}$ and $\mathbf{G}$: $\mathrm{MSE}(\mathbf{E}) = \sum_{i,j=1}^{d}(\hat{E}_{ij} - E_{ij})^2/d^2$ and $\mathrm{MSE}(\mathbf{G}) = \sum_{i,j=1}^{d}(\hat{G}_{ij} - G_{ij})^2/d^2$, which are summarized in Table 1. For comparison, we also present the MSEs of the estimates by the P-iPDA. Though the underlying cluster structure exactly matches the model assumption of the P-iPDA, that is, the regions in the same cluster are all pairwise connected, the Bayesian method gives much smaller errors, suggesting that incorporating indicator variables for significant directional effects in the model improves both selection accuracy of network edges and estimation efficiency of model parameters.

TABLE 1
The MSEs of estimated model parameters by the P-iPDA and the Bayesian method in Example 1

	Bayesian	**P-iPDA**		**Bayesian**	**P-iPDA**
MSE(**E**)	0.09	0.11	MSE(**G**)	0.11	0.20



(a) Estimated $\hat{P}_{ij}^A$ (b) P_{ij}^d (c) P_{ij}^d with a cutoff 0.5 (d) ROC Curve

FIG. 3. (a) *Estimated* $\hat{P}_{ij}^A$ *by the BMIDDM in Example* 2. (b) *Shows* P_{ij}^d *for* $i, j = 1, \ldots, 20$. (c) *Shows identified pairs of components with significantly different clustering probabilities in networks* 1 *and* 2 *using a cutoff* 0.5 *on* P_{ij}^d. *Specifically, if* $P_{ij}^d < 0.5$, *the corresponding cells* (*a pair*) *in the Figure are green*; *if* $P_{ij}^d > 0.5$ *and* $p_1 > p_2$, *the corresponding cells are red*; *and if* $P_{ij}^d > 0.5$ *and* $p_1 < p_2$, *the corresponding cells are blue*. (d) *Shows the ROC curve for selecting pairs of components with different clustering probabilities in the two networks by using various cutoffs on* P_{ij}^d.

5.2. *Example* 2: *Network comparison*. We generated a dynamic system of 20 dimension with 3 clusters of size 6, 4, and 10 from the bilinear model (2.3). The state functions $\mathbf{x}(t)$ of the first two clusters were generated using exactly the same parameters as those in Example 1. The third cluster consists of 10 densely connected components. Figure 3(a) shows estimated network edge selection probabilities $\hat{P}_{ij}^A$ by the BMIDDM for this example.

We develop a simple approach to compare networks of Examples 1 and 2. Let m_{1i} and m_{2i} denote module labels of the ith component in networks 1 and 2, respectively, $\mathbf{Y}_1$ and $\mathbf{Y}_2$ be the observed time series data of networks 1 and 2, respectively, $p_1 = P(\delta(m_{1i}, m_{1j}) = 1|\mathbf{Y}_1)$, and $p_2 = P(\delta(m_{2i}, m_{2j}) = 1|\mathbf{Y}_2)$. We compare the two components' clustering probabilities in two separate networks through evaluating the probability $P_{ij}^d = P(\delta(m_{1i}, m_{1j}) \neq \delta(m_{2i}, m_{2j})|\mathbf{Y}_1, \mathbf{Y}_2)$. Since the time series data of two networks are analyzed independently, $\delta(m_{1i}, m_{1j})$ and $\delta(m_{2i}, m_{2j})$ are independent, and thus

$$P_{ij}^d = P\big(\delta(m_{1i}, m_{1j}) = 0 \text{ and } \delta(m_{2i}, m_{2j}) = 1|\mathbf{Y}_1, \mathbf{Y}_2\big) + P\big(\delta(m_{1i}, m_{1j}) = 1 \text{ and } \delta(m_{2i}, m_{2j}) = 0|\mathbf{Y}_1, \mathbf{Y}_2\big) = p_1 + p_2 - 2p_1p_2.$$

In practice, we evaluate p_1 and p_2 by the corresponding clustering probabilities $\hat{P}_{ij}^m$ in their respective network. Figure 3(b) shows P_{ij}^d for comparing networks in Examples 1 and 2. Following typical Bayesian decision procedure [Gelman et al. (2004)], we use 0.5 as the threshold for P_{ij}^d, and the two components i and j are deemed to have different clustering probabilities in the two networks if $P_{ij}^d > 0.5$. The green cells in Figure 3(c) indicate a pair of components i and j with P_{ij}^d smaller than 0.5; the blue cells indicate components with P_{ij}^d larger than 0.5 and $p_1 < p_2$; and the red cells indicate components with P_{ij}^d larger than 0.5 and $p_1 > p_2$. The proposed approach identified components 11 to 20 to have

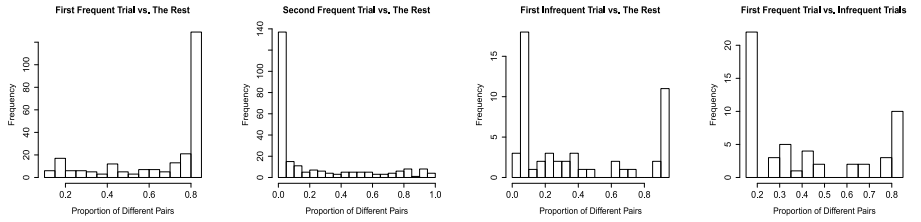
stronger connectivity in Example 2 than those in Example 1. Overall, using 0.5 as the threshold for P_{ij}^d, 70 edges (400 total edges) are identified significantly different, among which 48 edges are true positives with 100% TPR and 6.25% FPR. If using 0.6 as the threshold, we get 100% TPR and 0% FPR. Figure 3(d) shows the ROC curve for selecting component pairs with different clustering probabilities in the two simulated examples by using different thresholds on P_{ij}^d.

6. Application to ECoG data.

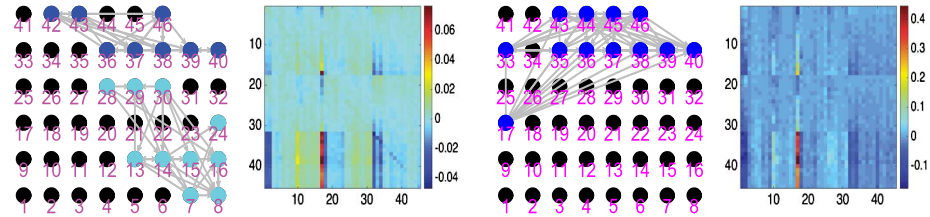
6.1. *Data acquisition*. The auditory ECoG data acquisition and processing methods have been described previously [Zhang et al. (2015)]. Briefly, continuous ECoG signals were recorded simultaneously from a 6×8 array of electrodes (2.3 cm diameter, 9 cm spacing) implanted over the lateral left hemisphere of an adult epilepsy patient, as shown in Figure 1, for clinical purposes of localizing seizures prior to resection surgery. The experimental paradigm was a 300-trial passive listening task using pure tones (50 ms duration). Since electrodes 47 and 48 are used as the reference and ground electrodes and electrode 32 contains excessive noise, recordings from a total of $d = 45$ electrode channels were analyzed. Auditory responses, computed using time-domain and time-frequency analyses [Boatman-Reich et al. (2010), Durka et al. (2001), Franaszczuk and Bergey (1998), Sinai et al. (2009)] were identified at three electrode sites (electrodes 14–16) consistent with the location of auditory cortex in the posterior temporal lobe. Seven electrode sites 1–4, 9–10, and 18 located in the inferior anterior temporal lobe were identified as the primary seizure focus based on clinical recordings.

The recording data contain 246 trials of ECoG recordings using a frequently presented, standard 1000 Hz tone stimulus of 50 ms duration and 54 trials using a different, infrequently repeated, that is, deviant, 1200 Hz tone stimulus also of 50 ms duration. All tone stimuli were presented sequentially at 1400 ms interstimulus interval. Following common practice in the literature, we focus on brain activity in an early cortical auditory processing time range, which corresponds to the first 150 ms after the stimulus onset. As such, each trial of data is of 250 ms duration: 100 ms pre-stimulus (0–100 ms), 50 ms for stimulus presentation (100–150 ms), and 100 ms post-stimulus (200–250 ms). We applied the BMIDDM to each 250-ms window independently to allow for variation of brain networks across trials and also to ensure that the assumed bilinear ODE model can approximate the underlying nonlinear system effectively. As such, in the MIDDM for the ECoG data, $d = 45$, $T = 250$, and $u(t) = 1$ for $100 \le t \le 150$.

6.2. *Data analysis*. We compared the first five trials with the rest of trials using 1000 Hz stimulus by calculating numbers of region pairs, denoted by $R_{l_1 l_2}$ for the pairwise comparison of trials l_1 and l_2, with different clustering probabilities (threshold 0.5 on P_{ij}^d), and found that the brain network in response to the first stimulus is distinct from those in response to the rest of the stimulus sequence.



(a) 1000 Hz Trial 1 vs. Rest (b) 1000 Hz Trial 2 vs. Rest (c) 1200 Hz Trial 1 vs. Rest (d) 1000 Hz Trial 1 vs. 1200 Hz Trials



(e) 1000 Hz Average Network (f) 1000 Hz Average **E** (g) 1200 Hz Average Network (h) 1200 Hz Average **E**

FIG. 4. (a) *and* (b) *Show histograms of proportions of region pairs with significantly different clustering probabilities* (*i.e.*, $P_{ij}^d > 0.5$) *in the network comparison of the first two trials with the rest trials using* 1000 *Hz stimulus*. (c) *Shows the histogram of proportions of region pairs with significantly different clustering probabilities in the network comparison of the first trial with the rest trials using* 1200 *Hz stimulus*. (d) *Shows the proportions of region pairs with significantly different clustering probabilities in the network comparison of the first trial using* 1000 *Hz stimulus with the trials using* 1200 *Hz stimulus*. (e) *and* (g) *Show the network edges with top* 5% *average ranks of* $\hat{P}_{ij}^A$ *across trials using* 1000 *Hz and* 1200 *Hz stimuli, respectively. Nodes in the same colors of either light blue, blue, or dark blue, correspond to regions in the same cluster identified by the BMIDDM, and nodes in black correspond to regions in the clusters with only one component*. (f) *and* (h) *Show average posterior means of* **E** *across trials using* 1000 *Hz and* 1200 *Hz stimuli, respectively.*

Figure 4(a) and (b) show histograms of the percentages of region pairs (among all possible pairs) with P_{ij}^d larger than 0.5 in the network comparison between the first two trials ($l_1 = 1, 2$) and the rest of trials using 1000 Hz stimulus, that is, histograms of $R_{l_1 l_2}/d(d-1)$ for $l_1 = 1$ or 2 and $l_2 \neq l_1$. Most pairs of regions in the first trial have much larger clustering probabilities than those in the rest of trials, in line with the discovery by [Garrido et al. (2009)] that connectivity strength between regions is the strongest in the first of several repetitive auditory events. In addition, by exploring connectivity among many regions, we found that the brain network of the first trial has more strong connections than the networks in the rest of trials (with 1000 Hz tone stimulus).

We evaluated networks in response to infrequent, deviant 1200 Hz stimulus. Figure 4(c) shows the histogram of the proportions of region pairs with P_{ij}^d larger than 0.5 in the network comparison between the first and the rest trials using 1200 Hz stimulus. In comparison with distinct networks of the first and the rest

trials using 1000 Hz stimulus, the difference among networks in response to the deviant 1200 Hz stimulus is much less pronounced. We have compared the network of the first trial using 1000 Hz stimulus with those of the 54 trials using 1200 Hz stimulus, as shown in Figure 4(d): the network of the first trial is similar to networks in most trials using infrequent, deviant stimuli. Similar results have been reported in the literature [Eliades et al. (2014), Herrmann, Henry and Obleser (2013), Herrmann, Schlichting and Obleser (2014)] that the observation of decreasing responses to repetitive stimuli does not apply to different or deviant stimuli.

To summarize analysis results of trials associated with the two different stimulus types, we calculated average ranks of posterior network-edge-selection probabilities $\hat{P}_{ij}^{A}$ of 246 trials using 1000 Hz stimulus and of 54 trials using 1200 Hz stimulus, respectively, and presented the two average networks with top 5% average ranks in Figure 4(e) and (g). We selected directional edges with top 5% posterior probabilities to identify most closely connected components with a small FPR. Moreover, with this high threshold, the identified clusters are small in size, and thus, easier to examine visually.

For trials using regular 1000 Hz stimuli, the identified network consists of two modules of closely connected brain regions that are believed to be specialized in different brain functions. The auditory responsive regions, electrodes 14–16, interact closely with regions in the posterior temporal lobe, involved in auditory perception. Regions in the inferior frontal lobe have dense interactions, which is in line with existing findings of short frontal lobe connections [Catani et al. (2012)] and this brain area's implication in predictive coding (generating expectations based on stimulus presentation probability) [Garrido et al. (2009)]. In comparison, for trials using 1200 Hz stimulus, regions in the frontal lobe show the strongest connections, consistent with the role of the frontal lobe in detecting novel or different auditory events [Näätänen et al. (2007), Schönwiesner et al. (2007)].

We estimated $\mathbf{E}$ and $\mathbf{G}$ of 300 trials, and the average $\hat{\mathbf{E}}$ of trials using two different stimuli are shown in Figure 4(f) and (h). Estimates of parameters denoting brain's effective connectivity in response to deviant stimuli are much larger in absolute values than those associated with regular, repetitive stimuli, suggesting stronger effective connectivity among brain regions in the former scenario. This result is in line with the finding of stronger brain responses to deviant stimuli in the literature [Eliades et al. (2014)].

7. Discussion. The BMIDDM brings three crucial advantages over the existing optimization method P-iPDA. First, one can detect different strengths of directional interactions among brain regions by the BMIDDM, and thus identify different levels of connectivity in the brain network. In the presented real data analysis of trials using 1000 Hz stimulus, given different posterior selection probabilities of network edges, the BMIDDM identified two small modules, which consist of

spatially close brain regions and are specialized for different functions. In contrast, the P-iPDA grouped the two modules into one cluster in almost all the trials and treated all connections within this cluster equally. Second, by using a high threshold for the posterior probabilities of network edges, the Bayesian method identifies most closely connected regions with a small FPR. Moreover, the resulting small clusters are much easier to understand and interpret than large clusters outputted by the P-iPDA, especially in connectivity studies of hundreds of brain regions. Third, given different posterior clustering probabilities of brain regions in different trials, the BMIDDM can be used to detect network changes.

The application of the proposed method to brain connectivity studies will potentially enhance our understanding of the brain's functional organization for two reasons. First, since brain components in the same module usually have a similar function, the module identification can be used to determine the module's functional role in the brain network, especially if the function of some brain regions in the module is already known. Taking the network result of the real data analysis, shown in Figure 4(e), as an example, the regions (electrode sites 7, 8, 13, 24, 28–30) connected with the auditory responsive regions (electrode sites 14–16) are mostly likely also specialized for primary auditory perception. Second, the module identification can significantly simplify the study of the brain's functional organization. It is much easier, both in terms of computation and interpretation of the results, to evaluate connections between and within modules separately than to evaluate connections between every pair of brain regions. Moreover, the proposed cluster-structured ODE model and the associated estimation method are scalable to a system with hundreds of components by first identifying modules locally and treating modules as components of the large-scale system.

In the brain network, connections within modules are usually short-range, strong, and dense, while those between modules tend to be sparse and long-range to ensure integration among different specialized areas [Park and Friston (2013)]. It is very likely that the two modules specialized for different functions shown in Figure 4(e) are integrated through weak, long-range connections. In the future study, we will extend the MIDDM to accommodate interactions among clusters. Then the ensuing ODE model uses two types of indicators for connectivity within and between modules, and the prior on indicators (3.5) is expanded to incorporate probabilities on between-module indicators. To distinguish between within- and between-module connections, the prior probability for the former should be much larger than that for the latter to reflect the prior belief that connections within modules are dense and connections between modules are sparse. The priors of the rest of model parameters are unchanged, and a similar PCGS algorithm can be developed for ensuing posterior simulations.

Since spatially close brain regions tend to have stronger connections, one can incorporate regions' spatial structure into the Bayesian framework for inferring effective connectivity between the regions. For example, the prior on module labels $\exp\{-\mu \sum_{ij} \delta(m_i, m_j)\}$ can be modified to $\exp\{-\mu \sum_{ij} W_{ij}\ \delta(m_i, m_j)\}$, where

the weight W_{ij} is proportional to the distance between regions i and j. Similarly, the prior probability for γ_{ij}^A and γ_{ij}^B can be changed to p_{ij}, which is proportional to the distance between regions i and j, leading to higher prior probabilities for short-range network edges. The choices of W_{ij} and p_{ij} and their effect on the ensuing posterior inference will be evaluated in the future research.

We select network edges based on the ranks rather than the exact values of their posterior selection probabilities, the same strategy used in Bayesian variable selection problems. This is because given $\mathbf{x}(t)$, identifying connected components is equivalent to simultaneously solving multiple high-dimensional variable selection problems. With limited data information and many candidate predictors, the difference between posterior selection probabilities of void and true variables is small. As such, the ranks are more informative than the exact values of the posterior selection probabilities regarding the underlying network structure.

One can select directional network edges based on original $\hat{P}_{ij}^A$ and $\hat{P}_{ij}^B$ without adjusting them for identified modules. However, we found that the network edge selection based on adjusted $\hat{P}_{ij}^A$ and $\hat{P}_{ij}^B$—obtained through the procedure described in Section 5—outperforms that based on unadjusted $\hat{P}_{ij}^A$ and $\hat{P}_{ij}^B$ by having a higher ROC curve. We attribute this finding to two possible reasons. First, within the ODE framework, the components in the same cluster tend to have similar temporal behaviors, because the instantaneous change of each component directly or indirectly depends on the states of others in the same cluster. Then this temporal similarity among components within the same cluster causes difficulty in the directional edge selection due to multicollinearity, but facilitates module identification. Consequently, module identification tends to be easier and more accurate than network edge selection, and utilizing the information from the former can enhance the accuracy of the latter. Second, the network edge selection based on unadjusted $\hat{P}_{ij}^A$ and $\hat{P}_{ij}^B$ treats each directional edge separately, while the selection based on adjusted probabilities utilizes information across different components.

If one is interested in identifying components on which the stimulus has a direct effect, he/she can include indicators γ_i^C for coefficient C_i, $i = 1, 2, \ldots, d$, in the MIDDM. Then the prior on indicator variables in (3.5) is modified to incorporate the prior probability p_c for γ_i^C for $i = 1, 2, \ldots, d$. The same PCGS algorithm can be used to draw samples from the ensuing posterior distribution. Note that if the goal is to identify nonzero C_is, the observed data $\mathbf{y}(t)$ should be standardized to unit norm without changing the center of the time series and the signs (positive, negative, or zero) of C_is in the ODE model for standardized data.

APPENDIX A: PROOF OF PROPER POSTERIOR DISTRIBUTION

Let $\mathfrak{N}$ be the normalizing constant for $p(\boldsymbol{\eta}, \boldsymbol{\Theta}_I, \boldsymbol{\sigma}^2 | \mathbf{Y}, \tau, \mu)$ in equation (3.7).

Because $\exp\{-\frac{1}{2\tau}\mathrm{R}(\boldsymbol{\eta}, \boldsymbol{\Theta}_I)\}, \exp\{-\mu \sum_{i,j=1}^{d} \delta(m_i, m_j)\} \le 1$, and $p_0 < 1$, we have:

$$
\begin{aligned}
&p(\boldsymbol{\eta}, \boldsymbol{\Theta}_I, \boldsymbol{\sigma}^2 | \mathbf{Y}, \tau, \mu) \\
&\quad \le \mathfrak{N} \prod_{i=1}^{d} \frac{1}{\sigma_i^T} \exp\left\{-\frac{(Y_i - \boldsymbol{\Phi}\eta_i)^2}{2\sigma_i^2}\right\} \prod_{i,j=1}^{d} \phi\left(\frac{A_{ij}}{\xi_0}\right) \\
&\qquad \times \prod_{i,j=1}^{d} \phi\left(\frac{B_{ij}}{\xi_0}\right) \prod_{i=1}^{d} \phi\left(\frac{C_i}{\xi_0}\right) \prod_{i=1}^{d} \phi\left(\frac{D_i}{\xi_0}\right) \prod_{i=1}^{d} \frac{1}{\sigma_i^2}.
\end{aligned} \tag{A.1}
$$

The above inequality gives an upper bound for the posterior joint density. In the following, we integrate out parameters in this upper bound step by step, and show that the upper bound is integrable.

Since $\mathbf{m}$, $\boldsymbol{\gamma}^A$, and $\boldsymbol{\gamma}^B$ are discrete and take a finite number of different values, after integrating these parameters out in (A.1), the ensuing joint posterior of $\boldsymbol{\theta}$ and $\boldsymbol{\sigma}^2$ follows:

$$
\begin{aligned}
&p(\boldsymbol{\eta}, \boldsymbol{\theta}, \boldsymbol{\sigma}^2 | \mathbf{Y}, \tau, \mu) \\
&\quad \le \mathfrak{C}_d \, \mathfrak{N} \prod_{i=1}^{d} \frac{1}{\sigma_i^{T+2}} \exp\left\{-\frac{(Y_i - \boldsymbol{\Phi}\eta_i)^2}{2\sigma_i^2}\right\} \\
&\qquad \times \prod_{i,j=1}^{d} \phi\left(\frac{A_{ij}}{\xi_0}\right) \prod_{i,j=1}^{d} \phi\left(\frac{B_{ij}}{\xi_0}\right) \prod_{i=1}^{d} \phi\left(\frac{C_i}{\xi_0}\right) \prod_{i=1}^{d} \phi\left(\frac{D_i}{\xi_0}\right),
\end{aligned} \tag{A.2}
$$

where $\mathfrak{C}_d$ is some positive constant depending on d. After integrating out $\boldsymbol{\theta}$ in (A.2), we have

$$
p(\boldsymbol{\eta}, \boldsymbol{\sigma}^2 | \mathbf{Y}, \tau, \mu) \le \xi_0^{2d^2+2d} \, \mathfrak{C}_d \, \mathfrak{N} \prod_{i=1}^{d} \sigma_i^{-T-2} \exp\left\{-\frac{(Y_i - \boldsymbol{\Phi}\eta_i)^2}{2\sigma_i^2}\right\}.
$$

Then as long as the number of basis coefficients L is smaller than the number of time points T for each component, the formula on the right of the above inequality is integrable.

APPENDIX B: TECHNICAL DETAILS OF PCGS ALGORITHM

B.1. Derive the joint posterior distribution $p(\mathbf{m}, \boldsymbol{\eta}, \boldsymbol{\sigma}^2, \boldsymbol{\gamma}^A, \boldsymbol{\gamma}^B | \mathbf{Y}, \boldsymbol{\tau}, \boldsymbol{\mu})$. In the following, we use $p(\theta|-)$ to denote the full posterior conditional distribution of θ. Based on the formulation of the joint distribution (3.7), given the rest of the parameters, $\{A_{ij}, B_{ij}, C_i, D_i\}_{j=1}^{d}$ are independent for $i = 1, 2, \ldots, d$, so we will first derive the posterior conditional distribution of $\{A_{ij}, B_{ij}, C_i, D_i\}_{j=1}^{d}$.

Let $\mathcal{G}_i^A = \{j, \delta(m_i, m_j)\, \gamma_{ij}^A \neq 0 \text{ and } j = 1, \ldots, d\}$ and $\mathcal{G}_i^B = \{j, \delta(m_i, m_j)\, \gamma_{ij}^B \neq 0 \text{ and } j = 1, \ldots, d\}$. Define a $d \times d$ diagonal matrix $\mathbf{I}_i^A$ where diagonal en-

tries corresponding to $\mathcal{G}_i^A$ equal 1, and the rest diagonal entries equal 0. We define $\mathbf{I}_i^B$ associated with $\mathcal{G}_i^B$ in the same manner. We use $\mathbf{M}[\mathcal{G}_1, \mathcal{G}_2]$ to denote the matrix consisting of elements in the rows indexed by $\mathcal{G}_1$ and columns indexed by $\mathcal{G}_2$ of $\mathbf{M}$, and use $\mathbf{M}[\mathcal{G},]$ to denote the matrix consisting of rows indexed by $\mathcal{G}$ of $\mathbf{M}$. Let $X_i^A(t) = \mathbf{I}_i^A \mathbf{x}(t)\ (1 - u(t))$ and $X_i^B(t) = \mathbf{I}_i^B \mathbf{x}(t)\ u(t)$, so $X_i^A(t)$ and $X_i^B(t)$ are vectors whose elements are functions of time t. Let $\Lambda_i(t) = ((X_i^A(t))', (X_i^B(t))', u(t), 1)$ and $\boldsymbol{\theta}_i = (\mathbf{A}[i, \mathcal{G}_i^A], \mathbf{B}[i, \mathcal{G}_i^B], C_i, D_i)'$. We have

$$p(\mathbf{A}[i,], \mathbf{B}[i,], C_i, D_i | -) \propto \exp\left\{-\frac{1}{2\tau}\int_0^T \left(\Lambda_i(t)\boldsymbol{\theta}_i - \frac{dx_i(t)}{dt}\right)^2 dt\right\} \times \prod_{j=1}^d \phi\left(\frac{A_{ij}}{\xi_0}\right) \prod_{j=1}^d \phi\left(\frac{B_{ij}}{\xi_0}\right) \phi\left(\frac{C_i}{\xi_0}\right) \phi\left(\frac{D_i}{\xi_0}\right),$$

where $dx_i(t)/dt = (\mathbf{b}^{(1)}(t))'\eta_i$.

After integrating out A_{ij} and B_{ij} corresponding to zero indicator values in the above equation, we have

$$\begin{aligned} &p(\boldsymbol{\theta}_i | \mathbf{m}, \boldsymbol{\eta}, \boldsymbol{\sigma}^2, \boldsymbol{\gamma}^A, \boldsymbol{\gamma}^B, \mathbf{Y}, \tau, \mu) \\ &\quad \propto \exp\left\{-\frac{1}{2}\boldsymbol{\theta}_i'\left(\frac{1}{\tau}\int_0^T \Lambda_i'(t)\Lambda_i(t)\,dt + \frac{1}{\xi_0^2}\mathbf{I}\right)\boldsymbol{\theta}_i\right\} \\ &\qquad \times \exp\left\{\frac{1}{\tau}\int \frac{dx_i(t)}{dt}\Lambda_i(t)\,dt\boldsymbol{\theta}_i\right\} \exp\left\{-\frac{1}{2\tau}\int_0^T \left(\frac{dx_i(t)}{dt}\right)^2 dt\right\}, \end{aligned} \tag{B.1}$$

where $\mathbf{I}$ denotes an identity matrix.

Let $\mathbf{M}_i = \frac{1}{\tau}\int_0^T \Lambda_i'(t)\Lambda_i(t)\,dt + \frac{1}{\xi_0^2}\mathbf{I}$ and $\mathbf{V}_i = \frac{1}{\tau}\int_0^T \frac{dx_i(t)}{dt}\ \Lambda_i'(t)\,dt$. Based on equation (B.1), we have

$$\begin{aligned} p(\mathbf{m}, \boldsymbol{\eta}, \boldsymbol{\sigma}^2, \boldsymbol{\gamma}^A, \boldsymbol{\gamma}^B | \mathbf{Y}, \tau, \mu) \propto{}& \prod_{i=1}^d \sigma_i^{-T-2} \exp\left\{-\frac{(Y_i - \boldsymbol{\Phi}\eta_i)^2}{2\sigma_i^2}\right\} \\ &\times \prod_{i=1}^d \det(\mathbf{M}_i)^{-1/2} \exp\left\{\sum_{i=1}^d \mathbf{V}_i'\mathbf{M}_i^{-1}\mathbf{V}_i/2\right\} \\ &\times \exp\left\{-\frac{1}{2\tau}\sum_{i=1}^d \int_0^T \left(\frac{dx_i(t)}{dt}\right)^2 dt\right\} \\ &\times \exp\left\{-\mu \sum_{i,j=1}^d \delta(m_i, m_j)\right\} \\ &\times p_0^{\sum_{i,j}\gamma_{ij}^A + \sum_{i,j}\gamma_{ij}^B} (1 - p_0)^{2d^2 - \sum_{i,j}\gamma_{ij}^A - \sum_{i,j}\gamma_{ij}^B}. \end{aligned}$$

We have $p(\mathbf{m}, \boldsymbol{\gamma}^A, \boldsymbol{\gamma}^B|\boldsymbol{\eta}, \boldsymbol{\sigma}^2, \mathbf{Y}, \tau, \mu) \propto \mathbb{J}(\mathbf{m}, \boldsymbol{\gamma}^A, \boldsymbol{\gamma}^B, \boldsymbol{\eta}, \tau, \mu)$, where

$$\begin{aligned}
&\mathbb{J}(\mathbf{m}, \boldsymbol{\gamma}^A, \boldsymbol{\gamma}^B, \boldsymbol{\eta}, \tau, \mu) \\
&\quad = \prod_{i=1}^{d} \det(\mathbf{M}_i)^{-1/2} \exp\left\{\sum_{i=1}^{d} \mathbf{V}_i' \mathbf{M}_i^{-1} \mathbf{V}_i/2\right\} \\
&\qquad \times \exp\left\{-\mu \sum_{i,j=1}^{d} \delta(m_i, m_j)\right\} p_0^{\sum_{i,j}\gamma_{ij}^A + \sum_{i,j}\gamma_{ij}^B} \\
&\qquad \times (1-p_0)^{2d^2 - \sum_{i,j}\gamma_{ij}^A - \sum_{i,j}\gamma_{ij}^B}.
\end{aligned}$$

B.2. Sequentially simulate m_i from $p(m_i|\mathbf{m}_{-i}, \boldsymbol{\eta}, \boldsymbol{\sigma}^2, \boldsymbol{\gamma}^A, \boldsymbol{\gamma}^B, \mathbf{Y}, \boldsymbol{\tau}, \boldsymbol{\mu})$ for $i = 1, 2, \ldots, d$. Let $\mathcal{G}_{-i}$ be the set of distinct values in $\mathbf{m}_{-i}$, and g_{-i} be any positive integer smaller than $d+1$ and not belonging to $\mathcal{G}_{-i}$. Then the posterior conditional distribution of m_i is discrete and has a support of $\{\mathcal{G}_{-i}, g_{-i}\}$. In addition, for each $z \in \{\mathcal{G}_{-i}, g_{-i}\}$,

$$P(m_i = z|\mathbf{m}_{-i}, \boldsymbol{\eta}, \boldsymbol{\sigma}^2, \boldsymbol{\gamma}^A, \boldsymbol{\gamma}^B, \mathbf{Y}) \propto \mathbb{J}(m_i = z, \mathbf{m}_{-i}, \boldsymbol{\gamma}^A, \boldsymbol{\gamma}^B, \boldsymbol{\eta}, \tau, \mu).$$

B.3. Sequentially simulate γ_{ij}^As and γ_{ij}^Bs from their posterior conditional probabilities. Given parameter values $\mathbf{m}$, $\boldsymbol{\gamma}_{-ij}^A$, $\boldsymbol{\gamma}^B$, and $\boldsymbol{\eta}$, γ_{ij}^A for $i, j = 1, 2, \ldots, d$ follows a Bernoulli distribution with probability:

$$\frac{\mathbb{J}(\mathbf{m}, \gamma_{ij}^A = 1, \boldsymbol{\gamma}_{-ij}^A, \boldsymbol{\gamma}^B, \boldsymbol{\eta}, \tau, \mu)}{\mathbb{J}(\mathbf{m}, \gamma_{ij}^A = 1, \boldsymbol{\gamma}_{-ij}^A, \boldsymbol{\gamma}^B, \boldsymbol{\eta}, \tau, \mu) + \mathbb{J}(\mathbf{m}, \gamma_{ij}^A = 0, \boldsymbol{\gamma}_{-ij}^A, \boldsymbol{\gamma}^B, \boldsymbol{\eta}, \tau, \mu)}.$$

Note that if $m_i \neq m_j$, the above probability equals p_0. Similarly, we sequentially simulate γ_{ij}^B conditional on the rest of the parameters.

B.4. Simulate $\boldsymbol{\theta}$ from $p(\boldsymbol{\theta}|\mathbf{m}, \boldsymbol{\eta}, \boldsymbol{\sigma}^2, \boldsymbol{\gamma}^A, \boldsymbol{\gamma}^B, \mathbf{Y}, \boldsymbol{\tau}, \boldsymbol{\mu})$. Based on the joint posterior distribution (3.7) and posterior conditional distribution of $\boldsymbol{\theta}_i$ (B.1),

$$\begin{aligned}
A_{ij}|\delta(m_i, m_j)\gamma_{ij}^A &= 0 \overset{\text{i.i.d.}}{\sim} \mathrm{N}(0, \xi_0^2), \\
B_{ij}|\delta(m_i, m_j)\gamma_{ij}^B &= 0 \overset{\text{i.i.d.}}{\sim} \mathrm{N}(0, \xi_0^2), \\
\boldsymbol{\theta}_i|\mathbf{m}, \boldsymbol{\sigma}^2, \boldsymbol{\gamma}^A, \boldsymbol{\gamma}^B, \mathbf{Y}, \tau, \mu &\overset{\text{ind}}{\sim} \mathrm{MN}(\mathbf{M}_i^{-1}\mathbf{V}_i, \mathbf{M}_i^{-1}) \qquad \text{for } i = 1, 2, \ldots, d.
\end{aligned}$$

B.5. Simulate $\boldsymbol{\sigma}^2$ from $p(\boldsymbol{\sigma}^2|\boldsymbol{\Theta}_I, \boldsymbol{\eta}, \mathbf{Y}, \boldsymbol{\tau}, \boldsymbol{\mu})$. From the joint posterior distribution (3.7), we have

$$\sigma_i^2|\boldsymbol{\Theta}_I, \boldsymbol{\eta}, \mathbf{Y}, \tau, \mu \overset{\text{ind}}{\sim} \text{Inv-Gamma}\left(\frac{T}{2}, \frac{(Y_i - \boldsymbol{\Phi}\eta_i)^2}{2}\right) \qquad \text{for } i = 1, 2, \ldots, d.$$

B.6. Simulate η from $p(\eta|\Theta_I, \sigma^2, \mathbf{Y}, \tau, \mu)$. Define a dT-by-dL matrix

$$\mathbf{Q} = \begin{pmatrix} \mathbf{\Phi} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \mathbf{\Phi} \end{pmatrix},$$

where $\mathbf{\Phi}$ is defined in (3.2). Let $\mathbf{U}$ be a dT-by-dT diagonal matrix with $(i-1)T$ to iT diagonal entries equalling $1/\sigma_i^2$, $i = 1, 2, \ldots, d$. Then

$$\begin{aligned} p(\eta|\Theta_I, \sigma^2, \mathbf{Y}, \tau, \mu) &\propto \exp\left\{-\frac{1}{2}(\mathbf{Y} - \mathbf{Q}\eta)'\mathbf{U}(\mathbf{Y} - \mathbf{Q}\eta)\right\} \\ &\quad \times \exp\left\{-\frac{1}{2\tau}(\eta'\mathbf{\Omega}_{\Theta_I}\eta - 2\mathbf{\Lambda}'_{\Theta_I}\eta + \mathbf{\Xi}_{\Theta_I})\right\} \\ &\propto \exp\left\{-\frac{1}{2}(\eta - \boldsymbol{\psi})'\mathbf{H}(\eta - \boldsymbol{\psi})\right\}, \end{aligned} \tag{B.2}$$

where $\mathbf{H} = \mathbf{Q}'\mathbf{U}\mathbf{Q} + \mathbf{\Omega}_{\Theta_I}/\tau$, and $\boldsymbol{\psi} = \mathbf{H}^{-1}(\mathbf{Q}'\mathbf{U}\mathbf{Y} + \mathbf{\Lambda}_{\Theta_I}/\tau)$. From (B.2),

$$\eta|\Theta_I, \sigma^2, \mathbf{Y}, \tau, \mu \sim \mathrm{MN}(\boldsymbol{\psi}, \mathbf{H}^{-1}).$$

Notation $\mathbf{\Omega}_{\Theta_I}$, $\mathbf{\Lambda}_{\Theta_I}$, and $\mathbf{\Xi}_{\Theta_I}$ are introduced in equation (3.4), and we derive their formulas depending on Θ_I in the following.

Define vectors with dL elements:

$$\begin{aligned} \mathbf{\Delta}_i(t) = \big(&A_{i1}\,\delta(m_i, m_1)\,\gamma_{i1}^A\,b_1(t)\big(1 - u(t)\big), \ldots, \\ &A_{i1}\,\delta(m_i, m_1)\,\gamma_{i1}^A\,b_L(t)\,\big(1 - u(t)\big), \\ &A_{i2}\,\delta(m_i, m_2)\,\gamma_{i2}^A\,b_1(t)\big(1 - u(t)\big), \ldots, \\ &A_{id}\,\delta(m_i, m_d)\,\gamma_{id}^A\,b_L(t)\,\big(1 - u(t)\big)\big), \\ \mathbf{\Upsilon}_i(t) = \big(&B_{i1}\,\delta(m_i, m_1)\,\gamma_{i1}^B\,b_1(t)\,u(t), \ldots, \\ &B_{i1}\,\delta(m_i, m_1)\,\gamma_{i1}^B\,b_L(t)\,u(t), \\ &B_{i2}\,\delta(m_i, m_2)\,\gamma_{i2}^B\,b_1(t)\,u(t), \ldots, \\ &B_{id}\,\delta(m_i, m_d)\,\gamma_{id}^B\,b_L(t)\,u(t)\big), \\ \mathbf{E}_i(t) = \Big(&\mathbf{0}_L, \ldots, \left(\frac{d\mathbf{b}(t)}{dt}\right)', \ldots, \mathbf{0}_L\Big), \end{aligned}$$

where $\mathbf{0}_L$ is a zero vector with L elements, and the $(i-1)L + 1$th to iLth elements of $\mathbf{E}_i(t)$ are nonzero. Then with basis representation, MIDDM (2.3) can be rewritten as $\mathbf{E}_i(t)\eta - \mathbf{\Delta}_i(t)\eta - \mathbf{\Upsilon}_i(t)\eta - C_i\,u(t) - D_i = 0$. Let $\mathbf{S}_i(t) =$

$\mathbf{E}_i(t) - \mathbf{\Delta}_i(t) - \mathbf{\Upsilon}_i(t)$. Then we have

$$\mathbf{R}(\boldsymbol{\eta}, \mathbf{\Theta}_I) = \sum_{i=1}^{d} \Big[\boldsymbol{\eta}' \int \mathbf{S}_i'(t)\mathbf{S}_i(t)\,dt\ \boldsymbol{\eta} - 2 \int \big(C_i u(t) + D_i\big)\, \mathbf{S}_i(t)\,dt\ \boldsymbol{\eta} + \int \big(C_i u(t) + D_i\big)\, \big(C_i u(t) + D_i\big)\,dt \Big].$$

Comparing the above to equation (3.4), we have

$$\mathbf{\Omega}_{\mathbf{\Theta}_I} = \sum_{i=1}^{d} \int \mathbf{S}_i'(t)\mathbf{S}_i(t)\,dt, \qquad \mathbf{\Lambda}_{\mathbf{\Theta}_I} = \sum_{i=1}^{d} \int \big(C_i u(t) + D_i\big)\, \mathbf{S}_i'(t)\,dt,$$

$$\mathbf{\Xi}_{\mathbf{\Theta}_I} = \sum_{i=1}^{d} \int \big(C_i u(t) + D_i\big)\, \big(C_i u(t) + D_i\big)\,dt.$$

Acknowledgments. Part of the project was conducted when Dr. Zhang was visiting the U.S. Statistical and Applied Mathematical Sciences Institute.

REFERENCES

AERTSEN, A. and PREISSL, H. (1991). Dynamics of activity and connectivity in physiological neuronal networks. In *Nonlinear Dynamics and Neuronal Networks* (H. Schuster, ed.) 281–302. VCH publishers Inc, New York.

ANDERSON, J. (2005). Learning in sparsely connected and sparsely coded system. Ersatz Brain Project Working Note.

BARD, Y. (1974). *Nonlinear Parameter Estimation*. Academic Press, New York. MR0326870

BHAUMIK, P. and GHOSAL, S. (2014). Bayesian estimation in differential equation models. Preprint. Available at arXiv:1403.0609.

BIEGLER, L., DAMIANO, J. and BLAU, G. (1986). Nonlinear parameter estimation: A case study comparison. *AIChE J.* **32** 29–45.

BOATMAN-REICH, D., FRANASZCZUK, P. J., KORZENIEWSKA, A., CAFFO, B., RITZL, E. K., COLWELL, S. and CRONE, N. E. (2010). Quantifying auditory event-related responses in multichannel human intracranial recordings. *Front. Comput. Neurosci.* **4** 4.

BRESSLER, S. and DING, M. (2002). Event-related potentials. In *The Handbook of Brain Theory and Neural Networks* 412–415. Wiley, New York.

BROWN, P. J., VANNUCCI, M. and FEARN, T. (1998). Multivariate Bayesian variable selection and prediction. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **60** 627–641. MR1626005

BRUNEL, N. J.-B. (2008). Parameter estimation of ODE's via nonparametric estimators. *Electron. J. Stat.* **2** 1242–1267. MR2471285

BULLMORE, E. and SPORNS, O. (2009). Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nat. Rev., Neurosci.* **10** 186–198.

CAFFO, B., PENG, R., DOMINICI, F., LOUIS, T. A. and ZEGER, S. (2011). Parallel MCMC for analyzing distributed lag models with systematic missing data for an application in environmental epidemiology. In *Handbook of Markov Chain Monte Carlo* (S. Brooks, A. Gelman, G. Jones and X. Meng, eds.) 493–511. CRC Press, Boca Raton, FL.

CALDERHEAD, B., GIROLAMI, M. and LAWRENCE, N. (2008). Accelerating Bayesian inference over nonlinear differential equations with Gaussian processes. *Adv. Neural Inf. Process. Syst.* **22**.

CAMPBELL, D. A. (2007). *Bayesian Collocation Tempering and Generalized Profiling for Estimation of Parameters from Differential Equation Models*. ProQuest LLC, Ann Arbor, MI. Thesis (Ph.D.)–McGill University (Canada). MR2711737

CAO, J., HUANG, J. Z. and WU, H. (2012). Penalized nonlinear least squares estimation of time-varying parameters in ordinary differential equations. *J. Comput. Graph. Statist.* **21** 42–56. MR2913355

CATANI, M., DELLACQUA, F., VERGANI, F., MALIK, F., HODGE, H., ROY, P., VALABREGUE, R. and THIEBAUT DE SCHOTTEN, M. (2012). Short frontal lobe connections of the human brain. *Cortex* **48** 273–291.

CERVENKA, M. C., FRANASZCZUK, P. J., CRONE, N. E., HONG, B., CAFFO, B. S., BHATT, P., LENZ, F. A. and BOATMAN-REICH, D. (2013). Reliability of early cortical auditory gamma-band responses. *Clin. Neurophysiol.* **124** 70–82.

CHEUNG, S., OLIVER, T., PRUDENCIO, E., PRUDHOMME, S. and MOSER, R. (2011). Bayesian uncertainty analysis with applications to turbulence modeling. *Reliab. Eng. Syst. Saf.* **96** 1137–1149.

CHKREBTII, O. A., CAMPBELL, D. A., CALDERHEAD, B. and GIROLAMI, M. A. (2016). Bayesian solution uncertainty quantification for differential equations. *Bayesian Anal.* **11** 1239–1267. MR3577378

CONRAD, P., GIROLAMI, M., SÄRKKÄ, S., STUART, A. and ZYGALAKIS, K. (2015). Probability Measures for Numerical Solutions of Differential Equations.

DAUNIZEAU, J., DAVID, O. and STEPHAN, K. (2011). Dynamic causal modelling: A critical review of the biophysical and statistical foundations. *NeuroImage* **58** 312–322.

DAVID, O. and FRISTON, K. (2003). A neural mass model for MEG/EEG: Coupling and neuronal dynamics. *NeuroImage* **20** 1743–1755.

DAVID, O., KIEBEL, S., HARRISON, L., MATTOUT, J., KILNER, J. and FRISTON, K. (2006). Dynamic causal modelling of evoked responses in EEG and MEG. *NeuroImage* **30** 1255–1272.

DEUFLHARD, P. and BORNEMANN, F. (2002). *Scientific Computing with Ordinary Differential Equations*. Springer, New York. MR1912409

DUNSON, D. B., HERRING, A. H. and ENGEL, S. M. (2008). Bayesian selection and clustering of polymorphisms in functionally related genes. *J. Amer. Statist. Assoc.* **103** 534–546. MR2523991

DURKA, P., IRCHA, D., NEUPER, C. and PFURTSCHELLER, G. (2001). Time-frequency microstructure of event-related electro-encephalogram eesynchronisation and synchronisation. *Med. Biol. Eng. Comput.* **39** 315–3211.

ELIADES, S., CRONE, N., ANDERSON, W., RAMADOSS, D., LENZ, F. and BOATMAN-REICH, D. (2014). Adaptation of high-gamma responses in human auditory association cortex. *J. Neurophysiol.* **112** 2147–2163.

FAN, J. and LI, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.* **96** 1348–1360. MR1946581

FÖLDIÁK, P. and YOUNG, M. P. (1995). Sparse coding in the primate cortex. In *The Handbook of Brain Theory and Neural Networks* 895–898. MIT Press, Cambridge.

FRANASZCZUK, P. J. and BERGEY, G. K. (1998). Application of the directed transfer function method to mesial and lateral onset temporal lobe seizures. *Brain Topogr.* **11** 13–21.

FRISTON, K. (2009). Causal modelling and brain connectivity in functional magnetic resonance imaging. *PLoS Biology* **7** 33.

FRISTON, K., HARRISON, L. and PENNY, W. (2003). Dynamic causal modelling. *NeuroImage* **19** 1273–1302.

GARRIDO, M., KILNER, J., KIEBEL, S., STEPHAN, K., BALDEWEG, T. and FRISTON, K. (2009). Comparative frequency analysis of single EEG-evoked potential records. *NeuroImage* **48** 269–279.

GELMAN, A., BOIS, F. and JIANG, J. (1996). Physiological pharamacokinetic analysis using population modeling and informative prior distributions. *J. Amer. Statist. Assoc.* **91** 1400–1412.

GELMAN, A., CARLIN, J. B., STERN, H. S. and RUBIN, D. B. (2004). *Bayesian Data Analysis*, 2nd ed. Chapman & Hall/CRC, Boca Raton, FL. MR2027492

GEORGE, E. and MCCULLOCH, R. (1993). Variable selection via Gibbs sampling. *J. Amer. Statist. Assoc.* **88** 881–889.

GEORGE, E. and MCCULLOCH, R. (1997). Approaches for Bayesian variable selection. *Statist. Sinica* **7** 339–373.

GIROLAMI, M. (2008). Bayesian inference for differential equations. *Theoret. Comput. Sci.* **408** 4–16. MR2460604

GRANER, F. and GLAZIER, J. A. (1992). Simulation of biological cell sorting using a two-dimensional extended Potts model. *Phys. Rev. Lett.* **69** 2013–2016.

HEMKER, P. (1972). Numerical methods for differential equations in system simulations and in parameter estimation. *Analysis and Simulation of Biochemical Systems* 59–80.

HERRMANN, B., HENRY, M. and OBLESER, J. (2013). Frequency-specific adaptation in human auditory cortex depends on the spectral variance in the acoustic stimulation. *J. Neurophysiol.* **109** 2086–2096.

HERRMANN, B., SCHLICHTING, N. and OBLESER, J. (2014). Dynamic range adaptation to spectral stimulus statistics in human auditory cortex. *J. Neurosci.* **34** 327–331.

HUANG, Y., LIU, D. and WU, H. (2006). Hierarchical Bayesian methods for estimation of parameters in a longitudinal HIV dynamic system. *Biometrics* **62** 413–423. MR2227489

HUANG, Y. and WU, H. (2006). A Bayesian approach for estimating antiviral efficacy in HIV dynamic models. *J. Appl. Stat.* **33** 155–174. MR2223142

ISHWARAN, H. and RAO, J. S. (2005). Spike and slab variable selection: Frequentist and Bayesian strategies. *Ann. Statist.* **33** 730–773. MR2163158

KENNEDY, M. C. and O'HAGAN, A. (2001). Bayesian calibration of computer models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **63** 425–464. MR1858398

KIEBEL, S., DAVID, O. and FRISTON, K. (2006). Dynamic causal modelling of evoked responses in EEG/MEG with lead-field parameterization. *NeuroImage* **30** 1273–1284.

KIM, S., TADESSE, M. G. and VANNUCCI, M. (2006). Variable selection in clustering via Dirichlet process mixture models. *Biometrika* **93** 877–893. MR2285077

LI, Z., OSBORNE, M. R. and PRVAN, T. (2005). Parameter estimation of ordinary differential equations. *IMA J. Numer. Anal.* **25** 264–285. MR2126204

LU, T., LIANG, H., LI, H. and WU, H. (2011). High-dimensional ODEs coupled with mixed-effects modeling techniques for dynamic gene regulatory network identification. *J. Amer. Statist. Assoc.* **106** 1242–1258. MR2896833

MATTHEIJ, R. and MOLENAAR, J. (2002). *Ordinary Differential Equations in Theory and Practice*. *Classics in Applied Mathematics* **43**. SIAM, Philadelphia, PA. MR1946758

MICHELOYANNIS, S. (2012). Graph-based network analysis in schizophrenia. *World J. Psychiatry* **2** 1–12.

MILLER, A. (2002). *Subset Selection in Regression*, 2nd ed. Chapman & Hall/CRC, Boca Raton, FL. MR2001193

MILO, R., SHEN-ORR, S., ITZKOVITZ, S., KASHTAN, N., CHKLOVSKII, D. and ALON, U. (2002). Network motifs: Simple building blocks of complex networks. *Science* **298** 824–827.

MILO, R., ITZKOVITZ, S., KASHTAN, N., LEVITT, R., SHEN-ORR, S., AYZENSHTAT, I., SHEFFER, M. and ALON, U. (2004). Superfamilies of evolved and designed networks. *Science* **303** 1538–1542.

NÄÄTÄNEN, R., PAAVILAINEN, P., RINNE, T. and ALHO, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clin. Neurophysiol.* **118** 2544–2590.

NEWMAN, M. E. J. (2006). Modularity and community structure in networks. *Proc. Natl. Acad. Sci. USA* **103** 8577–8696.

OLIVER, T. and MOSER, R. (2011). Bayesian uncertainty quantification applied to RANS turbulence models. *Int. J. Mod. Phys. Conf. Ser.* **318** 042032.

OLSHAUSEN, B. and FIELD, D. (2004). Sparse coding of sensor inputs. *Current Opinions in Neurobiology* **14** 481–487.

PARK, H.-J. and FRISTON, K. (2013). Structural and functional brain networks: From connections to cognition. *Science* **342** 1238411.

POTTS, R. B. (1952). Some generalized order-disorder transformations. *Math. Proc. Cambridge Philos. Soc.* **48** 106–109. MR0047571

POYTON, A., VARZIRI, M., MCAULEY, K., MCLELLAN, P. and RAMSAY, J. (2006). Parameter estimation in continuous dynamic models using principal differential analysis. *Computational Chemical Engineering* **30** 698–708.

QI, X. and ZHAO, H. (2010). Asymptotic efficiency and finite-sample properties of the generalized profiling estimation of parameters in ordinary differential equations. *Ann. Statist.* **38** 435–481. MR2589327

RAMSAY, J. O. and SILVERMAN, B. W. (2005). *Functional Data Analysis*, 2nd ed. Springer, New York. MR2168993

RAMSAY, J. O., HOOKER, G., CAMPBELL, D. and CAO, J. (2007). Parameter estimation for differential equations: A generalized smoothing approach. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **69** 741–796. MR2368570

REISS, P. T. and OGDEN, R. T. (2007). Functional principal component regression and functional partial least squares. *J. Amer. Statist. Assoc.* **102** 984–996. MR2411660

REISS, P. T. and OGDEN, R. T. (2009). Smoothing parameter selection for a class of semiparametric linear models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **71** 505–523. MR2649608

SCHÖNWIESNER, M., NOVITSKI, N., PAKARINEN, S., CARLSON, S., TERVANIEMI, M. and NÄÄTÄNEN, R. (2007). Heschl's gyrus, posterior superior temporal gyrus, and mid-ventrolateral prefrontal cortex have different roles in the detection of acoustic changes. *J. Neurophysiol.* **97** 2075–2082.

SINAI, A., CRONE, N., WIED, H., FRANASZCZUK, P., MIGLIORETTI, D. and BOATMAN-REICH, D. (2009). Intracranial mapping of auditory perception: Event-related responses and electrocortical stimulation. *Clin. Neurophysiol.* **120** 140–149.

SPORNS, O. (2011). *Networks of the Brain*. MIT Press, Cambridge, MA.

STUART, A. M. (2010). Inverse problems: A Bayesian perspective. *Acta Numer.* **19** 451–559. MR2652785

TADESSE, M. G., SHA, N. and VANNUCCI, M. (2005). Bayesian variable selection in clustering high-dimensional data. *J. Amer. Statist. Assoc.* **100** 602–617. MR2160563

THEO, H. and MIKE, E. (2004). Mapping multiple QTL using linkage disequilibrium and linkage analysis information and multitrait data. *Genet. Sel. Evol* **36** 261–279.

TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B* **58** 267–288. MR1379242

VAN DYK, D. A. and PARK, T. (2008). Partially collapsed Gibbs samplers: Theory and methods. *J. Amer. Statist. Assoc.* **103** 790–796. MR2524010

VARAH, J. M. (1982). A spline least squares method for numerical parameter estimation in differential equations. *SIAM J. Sci. Statist. Comput.* **3** 28–46. MR0651865

VOIT, E. (2000). *Computational Analysis of Biochemical Systems*: *A Practical Guide for Biochemists and Molecular Biologists*. Cambridge Univ. Press, Cambridge.

WAHBA, G. (1990). *Spline Models for Observational Data*. SIAM, Philadelphia, PA. MR1045442

WANG, H. and LENG, C. (2008). A note on adaptive group lasso. *Comput. Statist. Data Anal.* **52** 5277–5286. MR2526593

WU, H., LU, T., XUE, H. and LIANG, H. (2014a). Sparse additive ordinary differential equations for dynamic gene regulatory network modeling. *J. Amer. Statist. Assoc.* **109** 700–716. MR3223744

WU, S., XUE, H., WU, Y. and WU, H. (2014b). Variable selection for sparse high-dimensional nonlinear regression models by combining nonnegative garrote and sure independence screening. *Statist. Sinica* **24** 1365–1387. MR3241292

XUE, H., MIAO, H. and WU, H. (2010). Sieve estimation of constant and time-varying coefficients in nonlinear ordinary differential equation models by considering both numerical error and measurement error. *Ann. Statist.* **38** 2351–2387. MR2676892

YI, N., GEORGE, V. and ALLISON, D. B. (2003). Stochastic search variable selection for identifying multiple quantitative trait loci. *Genetics* **164** 1129–1138.

YUAN, M. and LIN, Y. (2005). Efficient empirical Bayes variable selection and estimation in linear models. *J. Amer. Statist. Assoc.* **100** 1215–1225. MR2236436

YUAN, M. and LIN, Y. (2006). Model selection and estimation in regression with grouped variables. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **68** 49–67. MR2212574

ZHANG, T., WU, J., LI, F., CAFFO, B. and BOATMAN-REICH, D. (2015). A dynamic directional model for effective brain connectivity using electrocorticographic (ECoG) time series. *J. Amer. Statist. Assoc.* **110** 93–106. MR3338489

ZOU, H. (2006). The adaptive lasso and its oracle properties. *J. Amer. Statist. Assoc.* **101** 1418–1429. MR2279469

ZOU, H. and HASTIE, T. (2005). Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **67** 301–320. MR2137327

T. ZHANG
Q. YIN
Y. SUN
UNIVERSITY OF VIRGINIA
148 AMPHITHEATER WAY
CHARLOTTESVILLE, VIRGINIA 22904-4135
USA
E-MAIL: tz3b@virginia.edu
qy2mn@virginia.edu
ys5pe@virginia.edu

B. CAFFO
JOHNS HOPKINS UNIVERSITY
615 N. WOLFE STREET
ROOM E3610
BALTIMORE, MARYLAND 21205
USA
E-MAIL: bcaffo@gmail.com

D. BOATMAN-REICH
JOHNS HOPKINS UNIVERSITY
601 N. CAROLINE STREET
BALTIMORE, MARYLAND 21287
USA
E-MAIL: dboatma@jhmi.edu