A Dynamic Directional Model for Effective Brain Connectivity using Electrocorticographic (ECoG) Time Series

Tingting Zhang** Jingwei Wu* Fan Li Brian Caffo* Dana Boatman-Reich* [1]

ABSTRACT

We introduce a dynamic directional model (DDM) for studying brain effective connectivity based on intracranial electrocorticographic (ECoG) time series. The DDM consists of two parts: a set of differential equations describing neuronal activity of brain components (state equations), and observation equations linking the underlying neuronal states to observed data. When applied to functional MRI or EEG data, DDMs usually have complex formulations and thus can accommodate only a few regions, due to limitations in spatial resolution and/or temporal resolution of these imaging modalities. In contrast, we formulate our model in the context of ECoG data. The combined high temporal and spatial resolution of ECoG data result in a much simpler DDM, allowing investigation of complex connections between many regions. To identify functionally segregated sub-networks, a form of biologically economical brain networks, we propose the Potts model for the DDM parameters. The neuronal states of brain components are represented by cubic spline bases and the parameters are estimated by minimizing a log-likelihood criterion that combines the state and observation equations. The Potts model is converted to the Potts penalty in the penalized regression approach to achieve sparsity in parameter estimation, for which a fast iterative algorithm is developed. The methods are applied to an auditory ECoG dataset.

KEY WORDS: brain mapping, ordinary differential equation (ODE), dynamic system, effective connectivity, Potts model

[1]⋆ co-first authors; ∗ co-corresponding authors. Zhang is assistant professor and Wu is PhD candidate at Department of Statistics, University of Virginia, Charlottesville, VA, USA (email:tz3b@virginia.edu). Li is assistant professor at Department of Statistical Science, Duke University, Durham, NC, USA. Boatman-Reich is professor at Department of Neurology and Caffo is professor at Department of Biostatistics, Johns Hopkins University, Baltimore, MD, USA.

1 Introduction

A useful nominal taxonomy of brain connectivity relies on three broad categories: anatomical, functional and effective (Friston, 1994). Anatomical refers to the network architecture, whereas functional and effective connectivity refer to network engagement. Specifically, functional refers to relationships (usually via correlations or synchrony) in activity while effective refers to directed effect of components on each other. Here we focus on effective connectivity, aiming to improve the understanding of functional interaction among brain regions, a topic of core interest in neuroscience (Swanson, 2003).

The evaluation of effective connectivity relies on modeling interactions among brain regions, and the assumed model depends on the method used to measure brain activity. As examples, functional MRI (fMRI), yielding indirect measurements of neuronal activity (the blood oxygenation level dependent, BOLD, signal) under an unknown hemodynamic response function, possesses relatively high spatial resolution but low temporal resolution (1 - 2 seconds between images). Electroencephalograph (EEG), in contrast, has high temporal resolution (1-2 ms), but relatively poor spatial resolution. Other modalities include magnetoencephalography (MEG) and electrocorticography (ECoG, discussed below) in humans and a variety of others available for animal studies. Each possesses benefits, compromises, processing details and intricacies for the feasibility and methodology for studying effective connectivity.

Electrocorticography (ECoG) involves intracranial electrophysiology recordings from subdural electrodes implanted directly on the cortical surface for clinical purposes in neurosurgical patients with medically intractable seizures or tumors. The combined high spatial (diameter 2.3 mm) resolution and temporal resolution (data collected every 1 ms) of ECoG data make it an ideal candidate for building effective connectivity models (Korzeniewska *et al.*, 2011). Of course, it is not without its limitations, notably including the very restricted population available for study, necessarily low subject sample sizes and subject-dependent and varying electrode placement locations, all impacting the generalizability of ECoG results. Nonetheless, for studying effective connectivity, ECoG offers a unique complement to traditional scalp EEG recordings methods (see Bressler and Ding, 2002; Boatman-Reich *et al.*, 2010, for a detailed comparison between

EEG and ECoG).

Effective connectivity is usually characterized by a model on the dynamic interactions between brain components (Aertsen and Preissl, 1991; Friston *et al.*, 2004), and the most commonly used models include Structural Equation Models (SEM, McIntosh and Gonzalez-Lima, 1994) and the closely related Dynamic Causal Models (DCM, Friston *et al.*, 2003). Here we use a model that can be thought of as a special case of DCM, as it attempts to describe the biophysical mechanism of the brain system building from the neuronal level. In contrast, most standard applications of SEM evaluate connection strength based on the variance-covariance structure of the observed data.

A DCM requires two parts: (1) neuronal state equations consisting of a set of ordinary differential equations (ODE), which describe how instantaneous changes of the neuronal activities of system components are modulated jointly by the immediate states of the components and experimental inputs; and (2) observation equations linking the underlying neuronal states of brain components to the observed data. A DCM can be viewed as a continuous time state-space model, parameterizing effective connectivity as coupling between the neuronal states of the brain system under the influence of experimental inputs.

Although DCM has been widely used in brain connectivity research, existing implementations, primarily within the setting of fMRI, EEG and MEG data (Friston *et al.*, 2003; David and Friston, 2003; David *et al.*, 2006; Kiebel *et al.*, 2006; Daunizeau *et al.*, 2011), have two major complications. First, parameter estimation of the DCM is computationally difficult, due to the complicated model formulation. Thus the number of brain regions included in the model is usually limited. Second, identifiability issues can arise, even with only a moderate number of brain components. The current practice in addressing this problem is to conduct Bayesian inference, using a highly informative prior, introducing subjective knowledge of the existence and strength of connections, and thus imposing regularization on the coupled dynamic system. However, a strong prior increases the risk of bias, raising concerns on the reliability of the results. These drawbacks are alleviated in ECoG, which has high temporal and spatial resolution, and a strong signal-to-noise ratio (SNR), to evaluate the effective connectivity among many brain regions. We

propose a new ODE-based model, hinging on the unique properties of ECoG data, and develop efficient methods to estimate the model. We refer to this model as a dynamic *directional* model (DDM) to delineate from the general DCM and to avoid confusion with the widely used Rubin Causal Model (Rubin, 1974, 1978; Holland, 1986). Though we use ECoG as an application of the proposed methods, we note that they have potential applicability in many other network studies where multivariate time series data measuring temporal changes of system components are collected, and the focus is on investigating directional interactions among them.

Anatomical and functional connections between brain regions are commonly believed to be biologically expensive, as they take up space and consume energy (Földiák and Young, 1995; Olshausen and Field, 2004; Anderson, 2005). Therefore, it is reasonable to assume that connections between the components of a complex brain system are sparse (Bullmore and Sporns, 2009; Micheloyannis, 2012). Sparsely connected brain networks can arise in different forms, and we here focus on the one that is decomposable into several functionally-segregated sub-networks/modules, a network structure called modularity and most relevant to brain organization (Tononi *et al.*, 1994; Newman, 2004). A main thrust of this paper is to propose a new DDM jointly with a Potts model (Potts, 1952; Graner and Glazier, 1992)—the Potts-based DDM (PDDM)—for the ODE parameters in the state equations to characterize modularity.

To solve for the proposed PDDM, we adopt the log-likelihood based criterion proposed by Varah (1982) and Ramsay *et al.* (2007), in which the neuronal state and observation equations are combined into one formula. The time-varying neuronal states of brain components are represented by spline bases. The parameters for the state and observation equations are estimated simultaneously by optimizing the log-likelihood criterion. To achieve sparsity, we employ a popular penalized log-likelihood based approach in regression analysis. This is intuitive, since the ODEs can be viewed as a set of special regression models, where temporal functions of neuronal states are predictors and their derivatives are the responses. In particular, we convert the Potts model to a penalty term to penalize large modules, and identify small functionally segregated sub-networks by minimizing the penalized criterion.

The proposed PDDM for the ECoG data is a special ODE model. There is an extensive sta-

tistical literature on solving ODEs from noisy data (e.g. Li *et al.*, 2002; Huang and Wu, 2006; Huang *et al.*, 2006; Ramsay *et al.*, 2007; Chen and Wu, 2008). However, these methods are mostly effective for low-dimensional cases, and are not directly applicable to the ECoG data because either the model is highly case-specific or the associated computation is too expensive. By decomposing high-dimensional differential equations into several independent low-dimensional ones using the Potts penalty, we greatly reduce the computational demand and increase estimation efficiency. As such, besides advancing the scientific research in effective brain connectivity, this article also contributes to statistical methodology for inference of high-dimensional ODEs.

The rest of the article is organized as follows. In Section 2, we first introduce the general DDM for the ECoG data and then propose the Potts-based DDM. Section 3 presents the log-likelihood based criterion, as well as the induced Potts penalty, for parameter estimation. Corresponding optimization strategies for selecting penalty parameters are also proposed. We apply the methods to analyze a real ECoG study in Section 4 and conduct simulations in Section 5. Section 6 concludes.

2 Directional dynamical model

2.1 The general DCM framework

Before proposing the DDM for ECoG data, we first introduce the general form of the DCMs for neurophysiological data, and then describe its existing examples in the context of fMRI and EEG/MEG data. Because the brain is a continuous-time physical system changing rapidly over time, it is intuitive to model its dynamics by characterizing its instantaneous changes. In this line of thought, the neuronal state equations are a set of ODEs, linking the derivatives of neuronal states $\mathbf{x}(t) = (x_1(t), \ldots, x_d(t))'$ of d brain components/regions to themselves under the influence of experimental inputs. We omit the subscript for subject, because the analysis is conducted subject by subject. Among all possible ODE-based dynamical models, the one with the Markovian property that the instantaneous changes of the system depend only on system states and experimental inputs at that same moment of time, is the simplest with least model complexity.

The Markovian property is a reasonable assumption for a brain system performing a simple task, such as visual, auditory, and motor functions, for a short period of time. For more complicated brain functions such as memory, a more complicated model may be needed to accommodate the time effect. Let $\mathbf{u}(t) = (u_1(t), \ldots, u_J(t))$ be J experimental input functions corresponding to designed causes (e.g., boxcar or stick stimulus functions). The first part of a DCM is the ODEs characterizing dynamic changes of the neuronal states:

$$\frac{d\mathbf{x}(t)}{dt} = F_1(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta}_1), \tag{1}$$

where F_1 is usually a set of unknown nonlinear functions describing the neurophysiological influences that the present activity of brain regions $\mathbf{x}(t)$ and experimental inputs $\mathbf{u}(t)$ exert upon changes in the others, and the vector $\boldsymbol{\theta}_1$ contains all the unknown parameters defining the system. Model (1) is deterministic, because each smooth function $x_i(t)$ is an average of a large number of neuron activities in the local region. The second part of a DCM consists of observation equations describing how the underlying neuronal activity causes changes in the observed data $\mathbf{y}$ in each region:

$$\mathbf{y}(t) = F_2(\mathbf{x}(t), \boldsymbol{\theta}_2, \boldsymbol{\epsilon}(t)), \tag{2}$$

where F_2 is some known function, $\boldsymbol{\theta}_2$ are unknown parameters, and $\boldsymbol{\epsilon}(t)$ are the error terms.

The formulation of F_1 and F_2 depends on the targeted imaging modality. For fMRI data, the neuronal states equations F_1 are usually approximated by the first and part of the second order Taylor expansions, with a bilinear form that will be specified later (equation (3) in Section 2.2). Strong restrictions are imposed on the parameter space to ensure that the underlying system is stable over time (hundreds of seconds). The observation equations F_2 include several differential equations characterizing the relationships between the BOLD signal, the normalized total deoxy-hemoglobin content, the normalized blood volume fraction, and the underlying neuronal activity (Friston *et al.*, 2003; Penny *et al.*, 2004; Friston, 2009; Stephan *et al.*, 2007). For EEG/MEG data, F_1 describes the interactions between three neuron subpopulations (pyramidal, spiny-stellate and inhibitory interneurons, respectively, in one of three cortical layers) at multiple signal-source locations (David and Friston, 2003; David *et al.*, 2005, 2006), while F_2 maps pyramidal cell

activities—the neuronal subpopulations assumed to give rise to the observed EEG/MEG data—linearly to scalp data. With both fMRI and EEG/MEG data, estimation of the DCMs is conducted within a Bayesian framework, where strong prior distributions of the parameters are imposed to encode the requisite constraints and ensure model identifiability.

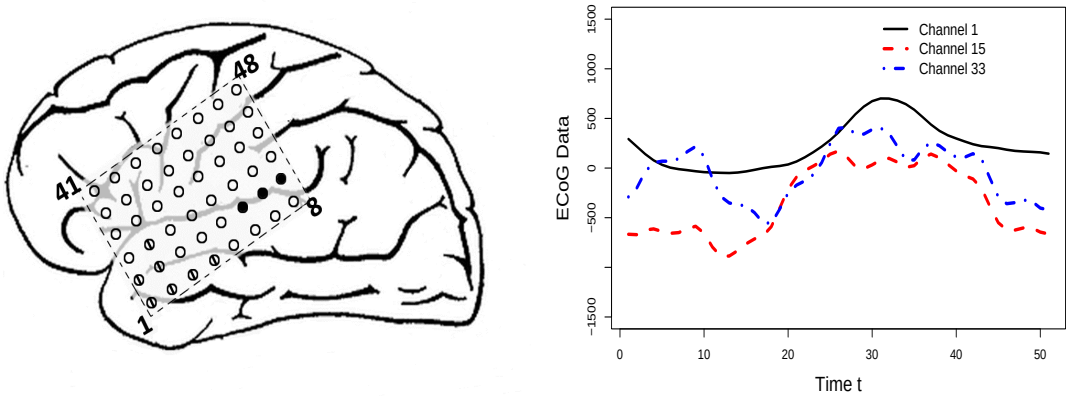
2.2 The DDM for ECoG time series

ECoG signals are recorded from electrodes implanted directly over the cortical surface of the brain. Electrodes are 2-3 mm in diameter and evenly spaced at 10 mm, center-to-center, in 6×8 or 8×8 arrays. The ECoG time series is recorded from all electrodes simultaneously. Figure 1(b) shows a 50-ms sample of ECoG data recorded from three electrode channels in our application. ECoG recordings are characterized by high signal-to-noise ratios and excellent temporal (1-2 ms) and spatial (10 mm) resolution. ECoG recordings from the human auditory cortex have been shown recently to be highly reliable and reproducible (Cervenka *et al.*, 2013). Resting on the unique properties of the ECoG, below we propose a new dynamic directional model (DDM) with a simple formulation to evaluate the effective connectivity between many brain components.

At the neuronal state, we use a bilinear approximation to the unknown F_1:

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{A}\,\mathbf{x}(t) + \sum_{j=1}^{J} u_j(t) \cdot \boldsymbol{B}_j \mathbf{x}(t) + \mathbf{C}\,\mathbf{u}(t) + \mathbf{D}, \tag{3}$$

where $\mathbf{A} = (A_{i_1 i_2})_{d\times d}$ with entry $A_{i_1 i_2}$ denoting the effect of component i_2 on component i_1 exerted at the current state; $\mathbf{B}_j = (B_{j,i_1 i_2})_{d\times d}$, $j = 1, \ldots, J$, couples the jth stimulus with the neuronal states and non-zero $B_{j,i_1 i_2}$ implies that the effect exerted by component i_2 on component i_1 depends on stimulus j; $\mathbf{C} = (C_{ij})_{d\times J}$ with C_{ij} denoting the effect of stimulus j on component i; and $\mathbf{D} = (D_1, \ldots, D_d)'$ with D_i denoting the intercept for component i. Model (3) is referred to as a "directional" or "causal" model, because it specifies two separate parameters $A_{i_1 i_2}$ and $A_{i_2 i_1}$ for the effect from component i_1 to i_2 and the effect from i_2 to i_1, respectively, which are allowed to take different values. The same applies to the parameters $\mathbf{B}$ encoding the stimulus-dependent effective connectivity between components. This is distinct from the association studies where usually only one parameter is specified to quantify the non-directional



(a) 6×8 electrode array

(b) ECoG Time Series

Figure 1: (a) 2D schematic of left hemisphere with 8×6 electrode array superimposed. Electrode locations are estimated from intra-operative photographs and post-implantation CT scans. Filled electrodes (nodes 14, 15, and 16) represent sites where auditory responses were elicited. Electrodes 1-4, 9-10, and 18 corresponding to circles with vertical lines inside are epileptic areas. (b) Plot of a short segment of ECoG time series collected at Channel 1, 15, and 33, respectively.

relationship between two components. The functional form (3) is not new: it was previously introduced in the context of DCM for fMRI data. However, for fMRI data, the model is assumed in conjunction with strong restrictions on the parameter space to ensure system stability over a long period of time (hundreds of seconds). In contrast, for ECoG data, we do not impose any restrictions, and instead assume Model (3) only for short periods of time, say less than 0.5 s. Also, to ensure that the bilinear approximation to F_1 is effective, we allow the parameters to vary across different periods of time. At the observation level, since each ECoG electrode directly measures the electrical activity of neurons in a local region with a small random noise, we assume F_2 in (2) has the form

$$\mathbf{y}(t) = \mathbf{x}(t) + \boldsymbol{\epsilon}(t), \tag{4}$$

where $\boldsymbol{\epsilon}(t) = (\epsilon_1(t), \ldots, \epsilon_d(t))'$ is a d-dimensional vector of measurement errors with mean zeroes. The errors $\epsilon_i(t)$ are assumed to be independently Gaussian distributed with mean zero and variance σ_i^2. Though the errors can be autocorrelated, it is not necessary to take autocorrelation into consideration given the strong SNR of the ECoG data. The observed data $\mathbf{y}(t)$ are measured

at discrete times $t = 1 \cdot \hbar, \ldots, T \cdot \hbar$, with $\hbar$ equalling one millisecond. Because each $y_i(t)$ is collected at one recording channel of ECoG, we use component, brain region, and channel exchangeably hereafter.

2.3 Potts-based DDM for functionally segregated system

Estimated parameters of Model (3) are often unreliable with very large variance when the number of components, d, is large for two reasons. First, though the observed data $\mathbf{y}(t)$ have a large SNR, the estimated $d\hat{\mathbf{x}}(t)/dt$ still has considerable errors. This is a prevalent problem in ODE model estimation, which is sensitive to noise. Second, to ensure the bilinear approximation effective, we fit the model to very short periods of ECoG time series with only a few hundred time points. Notice that the number of model parameters, $(J+1) \cdot d^2 + (J+1) \cdot d$, is of quadratic order of d, and thus large d leads to a large number of parameters, the estimates of which would have large variances given the limited data. This problem can be addressed by imposing sparsity assumptions on the parameters, that is, forcing many entries of $\mathbf{A}$ and $\mathbf{B}$ to be zero. Such sparsity-inducing regularization has been widely used in regression problems with a large number of candidate predictors (for a review, see Shao, 1998; Fan and Lv, 2010). Since given $\mathbf{x}(t)$, the parameter estimation for Model (3) is equivalent to solving d linear regression models, the sparsity assumption would likewise help to increase estimation efficiency of the DDM. More importantly, the sparsity assumption has a scientific motivation. Connections among brain regions are energy consuming (Földiák and Young, 1995; Olshausen and Field, 2004; Anderson, 2005), and biological units tend to minimize energy-consuming activities unless necessary in order to survive and prosper (Bullmore and Sporns, 2009; Micheloyannis, 2012). As such, it is reasonable to believe that the connections among many brain components should be sparse, especially when the human brain is performing a simple function within a short period of time. Here, among different possible forms of sparse networks, we focus on the one that is decomposable into several functionally segregated sub-networks/modules, a network structure known as "modularity", for two reasons. First, modularity has been widely reported in the literature on brain networks (Girvan and Newman, 2002; Milo *et al.*, 2002; Newman, 2003; Milo *et al.*,

2004; Newman, 2006). The modules form building blocks for large network systems and "*the modularity of the brain's architecture*" "*effectively insulates functionally bound subsystems from spreading perturbations due to small fluctuations in structure or dynamics*" (Sporns, 2011). Second, the modularity/cluster structure provides intuitive interpretation of independent functions of brain regions in different modules, and supports functional segregation and specialization, an important principle of the brain's functional organization (Friston *et al.*, 2004; Sporns, 2013).

To characterize the modularity, we first assign module labels to brain components. Let m_i be the module label of the ith system component, which can take any integer values from 1 to d, and $\delta(m_{i_1}, m_{i_2})$ be the Kronecker delta, which equals 1 whenever $m_{i_1} = m_{i_2}$ and 0 otherwise. For each brain component i_1 ($i_1 = 1, \cdots, d$), we assume the following model, a generalization of Model (3):

$$\begin{aligned} \frac{dx_{i_1}(t)}{dt} &= \sum_{i_2=1}^{d} \delta(m_{i_1}, m_{i_2}) \cdot A_{i_1 i_2} \cdot x_{i_2}(t) \\ &\quad + \sum_{j=1}^{J} u_j(t) \sum_{i_2=1}^{d} \delta(m_{i_1}, m_{i_2}) \cdot B_{j,i_1 i_2} \cdot x_{i_2}(t) + \sum_{j=1}^{J} C_{i_1 j} \cdot u_j(t) + D_{i_1}. \end{aligned} \tag{5}$$

Then we assume the Potts model on the module labels m_i:

$$\mathcal{P}(m_1, \ldots, m_d) \propto \exp\{-\mu \cdot \sum_{i_1, i_2=1}^{d} \delta(m_{i_1}, m_{i_2})\}, \tag{6}$$

where $\mathcal{P}$ denotes a probability measure, and μ is a positive constant. The above is referred to as the Potts-based DDM (PDDM). From (5) it is clear that the larger a module (cluster) is, the more parameters are required to characterize the directional effects between its elements, and thus the more complex the PDDM is. Indeed, the most complex PDDM is the one with all the components grouped into one single module, as Model (3), and the most parsimonious PDDM is the one with all components independent from each other, and each component forming one module. This is opposite to the standard clustering scenario, where the elements in one cluster are assumed identical and fewer and larger clusters imply a more parsimonious model. The Potts model (6) assigns small weights to large modules, and thus favors economic systems decomposable to small modules, within which the components are all connected, and between which the components are functionally independent.

3 Estimation of the DDM

There are two main approaches in the literature of ODE model estimation: discretization methods using numerical approximation (Biegler *et al.*, 1986; Gelman *et al.*, 1996; Campbell, 2007) and basis function expansion (Varah, 1982; Deuflhard and Bornemann, 2000; Ramsay and Silverman, 2005; Poyton *et al.*, 2006; Ramsay *et al.*, 2007). We adopt the latter for PDDM estimation for three reasons. First, since each $x_i(t)$ is an average of a large number of neuron activities in the local region, and temporally-dense observations of $x_i(t)$ are available, $x_i(t)$ is smooth (as shown in Figure 1(b)) and can be well approximately by functional bases. Second, the approach using basis expansions provides closed forms of $\mathbf{x}(t)$ and $d\mathbf{x}(t)/dt$. Third, under the bilinear model (3), given the estimated $\hat{\mathbf{x}}(t)$ and $d\hat{\mathbf{x}}(t)/dt$, the estimation of model parameters is straightforward, equivalent to solving d linear regressions, where $d\hat{\mathbf{x}}(t)/dt$ are responses, and $\hat{\mathbf{x}}(t)$ and $\hat{\mathbf{x}}(t) \cdot u(t)$ are predictors. In summary, the estimation of PDDM parameters using functional basis representation has computational advantages under the proposed bilinear formulation for the ECoG data. We elaborate the details of PDDM estimation below.

3.1 Log-Likelihood based criterion

We first explain the standard procedure based on cubic spline basis expansion for solving the DDM (3) and (4) without the Potts term. Let $\mathbf{x}(t)$ be represented by a set of cubic spline bases:

$$\mathbf{x}(t) = \mathbf{\Gamma}\, \boldsymbol{\phi}(t), \tag{7}$$

where $\boldsymbol{\phi}(t) = (\phi_1(t), \ldots, \phi_p(t))'$ is a vector of basis functions, and $\mathbf{\Gamma}$ is a d by p matrix whose ith row, denoted by $\mathbf{\Gamma}(i)$, consists of the basis coefficients of $x_i(t)$. We use equally-spaced cubic spline bases to represent $\mathbf{x}(t)$ since the time series data under study are equally spaced. Through simulations we found that choice of the number of spline bases does not affect the results much, as long as the number is not too far from the number of observations T.

A simple two-stage procedure proposed by Varah (1982) can be used to estimate DDM: First fit $\mathbf{\Gamma}\boldsymbol{\phi}(t)$ to the observed data $\mathbf{y}(t)$ using nonparametric spline smoothing regularized by a roughness penalty, and then estimate parameters $\boldsymbol{\theta}_1$ in (1) by fitting $d\hat{\mathbf{x}}(t)/dt$ to $F_1(\hat{\mathbf{x}}(t), \mathbf{u}(t), \boldsymbol{\theta}_1)$ if F_1

is known. Ramsay and Silverman (2005) and Poyton *et al.* (2006) replaced the roughness penalty on the fitted $\hat{\mathbf{x}}(t)$ with model-fitting errors $\int (d\hat{\mathbf{x}}(t)/dt - F_1(\hat{\mathbf{x}}(t), \mathbf{u}(t), \boldsymbol{\theta}_1))^2 dt$, also referred to as fidelity to differential equations, and proposed a log-likelihood based criterion that combines the curve-fitting errors of $\mathbf{x}(t)$ and fidelity to differential equations. These authors further developed an iterative algorithm—"iterated refined principal differential analysis (iPDA)"—to minimize this criterion to solve the ODEs. They showed that iPDA converges very fast and outperforms the original two-stage procedure. For the DDM (3) and (4) under study, the log-likelihood based criterion has the following form:

$$\begin{aligned} \mathrm{H}(\boldsymbol{\theta}) \;=\; & \sum_{i=1}^{d}\sum_{t=1}^{T}(y_i(t) - \boldsymbol{\Gamma}(i)\,\boldsymbol{\phi}(t))^2 \\ & +\lambda \cdot \sum_{i=1}^{d}\int \left(\boldsymbol{\Gamma}(i)\frac{d\boldsymbol{\phi}(t)}{dt} - \mathbf{A}(i)\,\mathbf{x}(t) - \sum_j u_j(t)\cdot \mathbf{B}_j(i)\,\mathbf{x}(t) - \mathbf{C}(i)\,\mathbf{u}(t) - \mathbf{D}\right)^2 dt, \end{aligned} \tag{8}$$

where $d\boldsymbol{\phi}(t)/dt \;=\; (d\phi_1(t)/dt, \ldots, d\phi_p(t)/dt)'$, and $\boldsymbol{\theta}$ contains all the parameters including $\boldsymbol{\Gamma}, \mathbf{A}, \mathbf{B}_j$, $\mathbf{C}$ and $\mathbf{D}$. The two steps of the iPDA for minimizing $\mathrm{H}(\boldsymbol{\theta})$ in (8) are given below:

A. Solve for the minimizer $\boldsymbol{\Gamma}$ of $\mathrm{H}(\boldsymbol{\theta})$ given $\mathbf{A}, \mathbf{B}_j$, $\mathbf{C}$, and $\mathbf{D}$;

B. Solve for the ordinary least square (OLS) estimates of $\mathbf{A}, \mathbf{B}_j$, $\mathbf{C}$, and $\mathbf{D}$ given estimated $\hat{\mathbf{x}}(t) = \hat{\boldsymbol{\Gamma}}\boldsymbol{\phi}(t)$ and $\frac{d\hat{\mathbf{x}}(t)}{dt} = \hat{\boldsymbol{\Gamma}}\frac{d\boldsymbol{\phi}(t)}{dt}$ with $\hat{\boldsymbol{\Gamma}}$ obtained from Step A.

Step A fits $\mathbf{x}(t)$ to the observed data regularized by the fidelity to the neuronal state equations, where the extent of regularization is controlled by the parameter λ. At the start of the iPDA, one may first estimate $\boldsymbol{\Gamma}$ in a fully nonparametric manner with $\lambda = 0$. The analytic form for the minimizer $\boldsymbol{\Gamma}$ of $\mathrm{H}(\boldsymbol{\theta})$ in Step A, given in the Appendix, is straightforward to derive, because $\mathrm{H}(\boldsymbol{\theta})$ is quadratic of $\boldsymbol{\Gamma}$ given the rest parameters.

3.2 Estimation with the Potts penalty

The Potts model (6), a prior distribution on the module labels in the view of the Bayesian paradigm, is mathematically equivalent to a penalty in the optimization framework. The Potts penalty, defined as the log of the probability of the Potts model, can be naturally included in

the log-likelihood based criterion (8). Then the curve fitting of $\mathbf{x}(t)$, module identification, and PDDM parameter estimation can be simultaneously obtained through minimizing the following Potts-penalized log-likelihood based criterion:

$$\mathrm{PH}_P(\boldsymbol{\theta}) = \mathrm{H}_P(\boldsymbol{\theta}) + \lambda \cdot \mu \cdot \sum_{i_1,i_2=1}^{d} \delta(m_{i_1}, m_{i_2}), \tag{9}$$

where $\mathrm{H}_P(\boldsymbol{\theta})$ is a modified $\mathrm{H}(\boldsymbol{\theta})$ with the second term changed according to (5). We here suppose that the penalty parameters λ and μ are given, and defer the associated parameter selection to Section 3.3.

We propose an iterative procedure—Potts-based iPDA (P-iPDA)—for minimizing $\mathrm{PH}_P(\boldsymbol{\theta})$, which is modified based on iPDA to accommodate the extra Potts penalty in (9). P-iPDA iterates between two major steps: solving for the minimizer $\boldsymbol{\Gamma}$ of $\mathrm{PH}_P(\boldsymbol{\theta})$ given $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{D}$, and given $\boldsymbol{\Gamma}$ searching for the optimal cluster and estimating associated parameters $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{D}$ that lead to the smallest $\mathrm{PH}_P(\boldsymbol{\theta})$. The first step is the same as Step A of iPDA. For the second step, we use a stepwise procedure to search for the optimal clusters. Specifically, let

$$\begin{aligned}\mathrm{R}(\boldsymbol{\theta}_2) = \sum_{i_1=1}^{d} \int \Bigg(& \frac{d\hat{x}_{i_1}(t)}{dt} - \sum_{i_2=1}^{d} \delta(m_{i_1}, m_{i_2}) \cdot A_{i_1 i_2} \cdot \hat{x}_{i_2}(t) \\ & - \sum_{j=1}^{J} u_j(t) \sum_{i_2=1}^{d} \delta(m_{i_1}, m_{i_2}) \cdot B_{j,i_1 i_2} \cdot \hat{x}_{i_2}(t) - \sum_{j=1}^{J} C_{i_1 j} \cdot u_j(t) - D_{i_1} \Bigg)^2 dt \\ & + \mu \cdot \sum_{i_1,i_2=1}^{d} \delta(m_{i_1}, m_{i_2}),\end{aligned}$$

where $\boldsymbol{\theta}_2 = (\mathbf{A}, \mathbf{B}_1, \cdots, \mathbf{B}_J, \mathbf{C}, \mathbf{D})$. Let $\mathcal{M}^s = \{m_1^s, \ldots, m_d^s\}$ be the collection of module labels of all d system components at step s, $\mathcal{M}_{-i}^s = \{m_1^s, \ldots, m_{i-1}^s, m_{i+1}^s, \ldots, m_d^s\}$ be the collection of module labels excluding component i, and G_{-i}^s be the set of distinct values in $\mathcal{M}_{-i}^s$. Setting $\mathcal{M}^0 = \{1, \ldots, d\}$ as the initial values, P-iPDA iterates between the following steps.

I.A Identical to Step A of iPDA.

I.B Given the estimated $\hat{\mathbf{x}}(t)$ and $\frac{d\hat{\mathbf{x}}(t)}{dt}$, repeat the following procedure for each component i.

B.1 Set $\mathcal{M}_{-i}^s = \mathcal{M}_{-i}^{s-1}$. Given $\mathcal{M}_{-i}^s$ and G_{-i}^s, repeat the following procedure for each

$z \in \{G^s_{-i}, g^s_{-i}\}$, where g^s_{-i} is any positive integer value not belonging to G^s_{-i} but smaller than $d+1$.

B.1.a Let $\mathcal{M}^s = \{m^s_1, \ldots, m^s_{i-1}, m^s_i = z, m^s_{i+1}, \ldots, m^s_d\}$.

B.1.b For each component $l \in \{1, \ldots, d\}$, based on $\mathcal{M}^s$ from Step B.1.a, identify the associated components $h_{l,1}, \ldots, h_{l,n_l}$ that fall into the same module as l, that is, $m^s_{h_{l,1}} = \ldots = m^s_{h_{l,n_l}} = m^s_l$. Obtain the OLS estimates of A_{lh}, $B_{j,lh}$, $C_{l,j}$, and D_l by regressing $\frac{d\hat{x}_l(t)}{dt}$ versus $\hat{x}_h(t)$, $u_j(t) \cdot \hat{x}_h(t)$ and $u_j(t)$ for $j = 1, \ldots, J$ and $h = h_{l,1}, \ldots, h_{l,n_l}$. Set A_{lh}, $B_{j,lh}$ for the rest h as zero.

B.1.c. Based on $\mathcal{M}^s$ and estimated $\hat{\mathbf{A}}$, $\hat{\mathbf{B}}_j$, $\hat{\mathbf{C}}$, and $\hat{\mathbf{D}}$ from Step B.1.b, calculate the associated $\mathrm{R}(\hat{\boldsymbol{\theta}}_2)$, denoted by $R^s_{i,z}$.

I.C Let $(i^\star, z^\star_{i^\star}) = \text{argmin}\{R^s_{i,z}, i = 1, \ldots, d, z \in (G^s_{-i}, g^s_{-i})\}$. Let $\mathcal{M}^s_{-i^\star} = \mathcal{M}^{s-1}_{-i^\star}$ and $m^s_{i^\star} = z^\star_{i^\star}$, update $\hat{\mathbf{A}}, \hat{\mathbf{B}}_j, \hat{\mathbf{C}}$ and $\hat{\mathbf{D}}$ given the component labels $\mathcal{M}^s$.

Step B.1 proposes all possible module-label changes for component i given the module labels $\mathcal{M}^{s-1}_{-i}$ of the rest components. The extra element g^s_{-i} is used to allow component i to form an module (containing component i only) independent of the rest components. Then sub-steps B.1.a and B.1.b calculate the associated $\mathrm{R}(\hat{\boldsymbol{\theta}}_2)$ given the module labels of d system components $\mathcal{M}^s = \{m^s_1, \ldots, m^s_{i-1}, m^s_i = z, m^s_{i+1}, \ldots, m^s_d\}$, where z is the proposed module label for component i. Step I.C selects the optimal one with the smallest $\mathrm{R}(\hat{\boldsymbol{\theta}}_2)$ among all one-component label changes.

3.3 Penalty parameter selection

The criterion $\mathrm{PH}_P(\boldsymbol{\theta})$ depends on two penalty parameters, λ and μ, the former regularizing the roughness of the estimated $\hat{\mathbf{x}}(t)$ and the latter controlling the size of the modules. Existing approaches for penalty parameter selection include ordinary leave-one-out cross-validation (LOOCV), K-fold cross-validation, generalized cross-validation (GCV, Wahba, 1990), GCV for functional data (Reiss and Ogden, 2007, 2009), and restricted maximum likelihood (Wood, 2011). For the particular multivariate time series and ODE models under study, we consider a modified LOOCV: at each round, leave out the observations at one time point, say the $(v+1)th$

time point, of the d time series, and then apply the P-iPDA with the candidate λ and μ to the rest of the data, and obtain estimates $\hat{\mathbf{x}}(t)$ and $\hat{\mathbf{A}}, \hat{\mathbf{B}}, \hat{\mathbf{C}}, \hat{\mathbf{D}}$. The predicted $\mathbf{x}$ at the left-out time $v+1$ is given by

$$\tilde{\mathbf{x}}(v+1) = \hat{\mathbf{x}}(v) + \left(\hat{\mathbf{A}}\, \hat{\mathbf{x}}(v) + \sum_{j=1}^{J} u_j(v) \cdot \hat{\mathbf{B}}_j\, \hat{\mathbf{x}}(v) + \hat{\mathbf{C}}\, \mathbf{u}(v) + \hat{\mathbf{D}} \right) \cdot \hbar.$$

And we use the sum of prediction errors as the criterion to select λ and μ,

$$\text{SPE}(\lambda, \mu) = \sum_{v=1}^{T-1} \sum_{i=1}^{d} (y_i(v+1) - \tilde{x}_i(v+1))^2.$$

The LOOCV is very time-consuming for two reasons. First, if μ is small, it will result in many components falling into the same module and consequently the P-iPDA requires many iterations to converge. Second, when many components fall into the same module, computation for the coefficients, $\mathbf{\Gamma}$, of the cubic spline bases representing $\mathbf{x}(t)$ in the module involves inversion of a large matrix, which is computationally expensive. We propose to conduct LOOCV, instead of at every time point, at only a small sub-sample of the data, say n equally-spaced time points. We call this procedure sub-sample cross validation (SSCV).

The optimal choice of the size of the sub-sample in SSCV—the number of time points n—is an empirical study for subsequent work and our work has the same issues in choosing cross-validation size as every other method that uses it. A major issue to be considered when choosing n is the ensuing computational cost, which depends on the length of time series T, the number of system components d, and the number of stimulus types J. The larger n, the closer the estimated prediction error to that of LOOCV, and also the more computational cost. The choice of n also depends on the volatility of $\mathbf{x}(t)$. If $\mathbf{x}(t)$ changes slowly, then the values of $\mathbf{x}(t)$ at close times are similar, and consequently a small value n is needed to estimate the prediction error. For our application, we found that the data at every five consecutive time points take similar values, so conducting LOOCV on 50 equally-spaced time points provides reliable comparison of prediction errors using different combinations of λ and μ without incurring prohibitive computational burden.

Even with SSCV, it is still computationally challenging to explore a large number of candidate penalty parameters, so we propose to first select a small number of pairs of λ and μ solely based

on the curve-fitting error

$$\text{SSE}(\lambda, \mu) = \sum_{i=1}^{d} \sum_{t=1}^{T} (y_i(t) - \hat{\mathbf{\Gamma}}(i)\, \hat{\boldsymbol{\phi}}(t))^2$$

and the fidelity to ODE models

$$\text{Fid}(\lambda, \mu) = \sum_{i=1}^{d} \int \left(\hat{\mathbf{\Gamma}}(i) \frac{d\boldsymbol{\phi}(t)}{dt} - \hat{\mathbf{A}}(i)\, \hat{\mathbf{x}}(t) - \sum_{j} u_j(t) \cdot \hat{\mathbf{B}}_j(i)\, \hat{\mathbf{x}}(t) - \hat{\mathbf{C}}(i)\, \mathbf{u}(t) - \hat{\mathbf{D}} \right)^2 dt$$

before proceeding cross validation. For each combination of λ and μ, we apply P-iPDA to the entire data, and keep records of $\text{SSE}(\lambda, \mu)$ and $\text{Fid}(\lambda, \mu)$ from the last step. Then screen out the parameters associated with either $\text{SSE}(\lambda, \mu)$ or $\text{Fid}(\lambda, \mu)$ much larger than the smallest values, and those associated with very many, say d, clusters or very few, say only 1, cluster. Then among the rest of a few parameters with both small $\text{SSE}(\lambda, \mu)$ and $\text{Fid}(\lambda, \mu)$, we use SSCV to select the pair of penalty parameters for the final data analysis.

4 Application to a Real ECoG Study

4.1 Data acquisition

We now apply the proposed method to analyze the ECoG data collected from a right-handed adult female epilepsy patient, who had subdural electrodes implanted for clinical purposes of seizure localization and functional mapping prior to surgery for treatment of medically intractable seizures. The data were recorded from a 6×8 electrode array implanted over the left hemisphere, including the posterior temporal lobe (auditory cortex) of the patient (see Figure 1(a)). Electrodes were 2.3 mm in diameter and spaced 9 mm center-to-center. Recordings were performed four days after electrode implantation while the patient was awake and fully responsive. The patient participated in several research studies and provided informed written consent for all research testing in compliance with the Johns Hopkins Institutional Review Board.

Event-related ECoG recordings were acquired simultaneously from all electrodes using an established 300-trial passive auditory oddball paradigm (Sinai *et al.*, 2009; Cervenka *et al.*, 2013). Two 50 ms duration single-frequency tones were presented: a frequently repeated 1000 Hz tone

(N=246 trials) and an infrequently and pseudo-randomly presented (no consecutive repetitions) 1200 Hz tone (N=54 trials). Tone stimuli were presented binaurally at a comfortable listening level through insert earphones at inter-stimulus intervals of 1450 ms. To reduce attention effects, the patient watched an animated movie with no sound. The continuous ECoG signal was amplified (5$\times$1000) and recorded digitally using a referential montage, 1000 Hz A/D sampling and a bandwidth of 0.1-300 Hz, as previously described (Cervenka *et al.*, 2013). Two electrodes (channels 47 and 48) in the top corner of the array, outside perisylvian cortex, were assigned as the reference and ground electrodes. Stimulus onset markers were recorded simultaneously to separate EEG channels.

ECoG recordings from the 46 active electrode channels were reviewed visually to identify any with excessive artifact for exclusion from analysis. One channel was identified as noisy and excluded (channel 32). The remaining $d = 45$ channels of ECoG time series data were analyzed. For each channel, the ECoG signal was segmented into 300 trials of 250 ms duration: 100 ms pre-stimulus (0-100 ms), 50 ms for stimulus presentation (100-150 ms), and 100 ms post-stimulus (50-150 ms). We use relatively short segments to maintain an efficient bilinear approximation of the nonlinear connectivity relationships among components. Since the 1000 Hz tone was presented far more frequently than 1200 Hz, we focus on the analysis of 246 trials under 1000 Hz. For each 250-ms trial, let $u(t) = 1$ for $100 \leq t \leq 150$. We omit the subscript for matrix $\mathbf{B}$, as only one stimulus is considered in the model.

The presence of cortical auditory evoked and spectral power responses was used to identify electrode sites responsive to auditory stimulation. Evoked responses were derived by trial averaging of the phase-locked signal components in the time domain, where the phase lock refers to neuron firing at or near a particular phase angle of the sinusoidal stimulus sound wave; event-related changes in spectral power were determined by using time-frequency analysis. We focused on the evoked N1 response that occurs around 100 ms post-stimulus and is a large, obligatory cortical response to sound stimulation that is prominent in ECoG recordings from auditory cortex (Edwards *et al.*, 2005; Sinai *et al.*, 2009). For the spectral power analysis, we used a time-frequency matching pursuit algorithm (Mallat and Zhang, 1993; Franaszczuk and Bergey, 1998;

Durka *et al.*, 2001; Boatman-Reich *et al.*, 2010). A total of 3 electrode sites were identified as auditory responsive based on the presence of measurable auditory evoked N1 responses and increased spectral power (> 30 Hz). The three electrode sites were located in the posterior superior temporal gyrus, consistent with the location of auditory cortex (Figure 1(a)). Based on clinical intracranial EEG recordings, seven electrode channels located in the inferior-anterior portion of the temporal lobe were associated with epileptiform activity and identified as the primary seizure focus (Figure 1(a)).

4.2 Data analysis

Given one trial of data, we first standardized each time series to unit variance, applied P-iPDA to the standardized data and evaluated SSE(λ, μ) and Fid(λ, μ) for all combinations of $\lambda = (0.1, 0.25, 0.5, 1, 2.5, 5, 10, 25, 50, 100, 250, 500, 1000)$ and $\lambda{\cdot}\mu = (0.0001, 0.001, 0.01, 0.1, 1, 10, 50, 100)$, which cover a wide range of values. Based on the outputted SSE(λ, μ) and Fid(λ, μ) for each combination, we screened out parameters with either too large SSE(λ, μ) or Fid(λ, μ) and those resulting in either too many, say d, clusters or only 1 cluster. Then we conducted SSCV on the rest pairs of parameters to select final $(\lambda, \lambda \cdot \mu)$. We applied this selection procedure to five randomly chosen trials, collected at five different periods of the ECoG recording process, ranging from the beginning to the end. We found that the same parameters $\lambda = 0.25$ and $\lambda \cdot \mu = 0.01$ were selected. As such, we used the same pair of penalty parameters for analyzing data across different trials. We want to stress that though brain activities may vary across trials, this does not necessarily mean that the corresponding penalty parameters selected would vary significantly. In fact, the selection of penalty parameters does not directly depend on the temporal values of the state functions, but rather the SNR of the data and the most significant causal interactions among different components, which may be stable across trials. In this application, there are two potential reasons for identical penalty parameters being selected by SSCV. First, the parameter λ—used to control the roughness of the fitted curves—depends most on the SNR of the data. Since the SNR of ECoG data is consistently high, smaller values of λ are consistently chosen, inducing a weak regularization effect. Second, the Potts penalty parameter μ is

used to balance the ODE model size and fitting errors, and its value depends on the significance of the directional effects between components and/or the strength of the association between the instantaneous changes of components' state functions and the functions themselves. Even if the state functional curves vary across trials, the most significant connections between components can still be stable. An analogy is a Markov chain with a constant transition probability but temporally varying states. Since here we study connectivity within a small brain area involved in a basic brain auditory function, it is very likely that the most significant interactions among brain components are stable (Flinker *et al.*, 2010). Consequently, very similar (or identical) values of penalty parameters are selected.

We applied P-iPDA to each trial independently with $(\lambda = 0.25, \lambda \cdot \mu = 0.01)$, allowing the cluster structure and model parameters to vary across trials. The colored matrix in Figure 2(a) summarizes the percentage of every pair of channels being clustered together across 246 clustering results, each obtained with one trial of data. The color scale is arbitrary, with dark red indicating high percentage and dark blue indicating low percentage. Based on this matrix, we constructed networks of closely-connected regions, and presented them in Figures 2(b) and 2(c) with different thresholds for clustering frequencies. Each node represents one recording channel and each edge in 2(b) and 2(c) between two channels respectively indicates that they were clustered into the same module by P-iPDA in more than 90% and between 70% and 90% of trials. We found that channels 33-46 and 17 are most closely connected, with clustering percentage higher than 90% (corresponding to the darkest red areas in Figure 2(a)). This is possibly because these channels all reside in the same brain local area, inferior frontal lobe. Then the connections among them are "local" and thus strongest. As shown in Figure 2(c), the auditory responsive regions, especially channels 15 and 16, which are located at adjacent sites along the posterior superior temporal gyrus and inferior parietal lobe, proximal to auditory cortex, are closely connected to the inferior frontal lobe. This result is in keeping with the findings that the inferior frontal lobe is involved in auditory processing, specifically phonological and syntactic processing (Burton, 2001), and music perception (Steven *et al.*, 2006). In addition, the small clustering frequencies between channels 1-6, 9-10 and 18 in the epileptic area and channels 14-

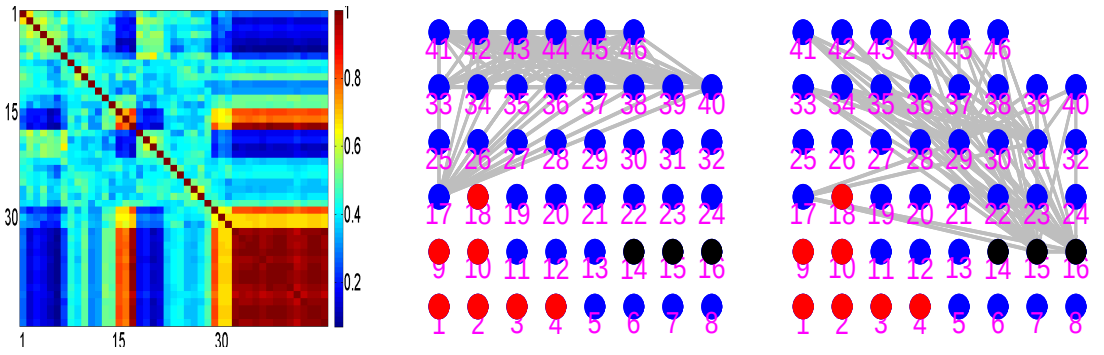
16 in auditory cortex (the dark blue areas in Figure 2(a)) indicate that there is no or very weak interaction between them. This is possibly because the brain sub-network involved in the auditory function was unaffected by the activity in brain epileptic areas during the data collection.

Figures 3(a) and 3(b) show the average of $\hat{\mathbf{A}}$ and $\hat{\mathbf{B}}$ estimated by P-iPDA across all trials. Row i ($1 \leq i \leq d$) of matrices $\hat{\mathbf{B}}$ and $\hat{\mathbf{A}}$ respectively represent interactive effects exerted by other channels on channel i with and without tone stimuli, and column i of matrices $\hat{\mathbf{B}}$ and $\hat{\mathbf{A}}$ respectively represent interactive effects exerted by channel i on other channels with and without tone stimuli. The columns corresponding to channels 14-16, the auditory-responsive regions, had values close to zero in the averaged $\hat{\mathbf{A}}$, indicating that no notable effects of the three channels were observed over other channels when tone stimulus was not evoked. The effect of channel 17 over other channels stood out in $\hat{\mathbf{A}}$, revealing that channel 17, located in the inferior frontal lobe, strongly affected all three auditory-responsive electrodes in the first module. Moreover, estimates of $\mathbf{A}$ from each of the 246 trials show that the effect of channel 17 was stable over time. This suggests that although channel 17 showed no identifiable auditory response itself, it may serve to monitor activity in those auditory responsive sites located more posteriorly. The top-down monitoring role of the frontal lobe has previously been reported by Stuss and Levine (2002).

5 Simulations

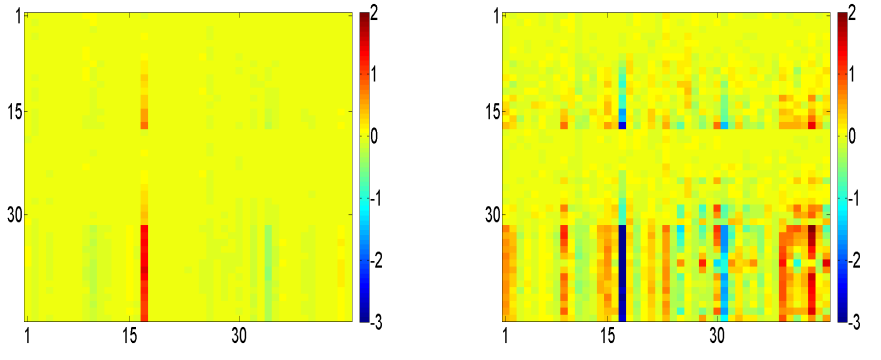
5.1 Example 1: two clusters with medium size

We further investigate the operating characteristics of P-iPDA in comparison to iPDA through simulations. Example 1 used data generated from a dynamic system of 32 channels. To mimic the real data, the stimulus function $u(t)$ is identical to that of the real data, and there are two functionally segregated sub-networks, each containing 16 channels. For simplicity, we use identical $\mathbf{A}$ and $\mathbf{B}$. The parameters are chosen such that $\mathbf{x}(t)$ have periodic temporal variations, and do not uni-directionally increase or decrease over time, as shown in Figure 4. Also, channels in the two modules have different frequencies of temporal variation, such that $\mathbf{x}(t)$ in two clusters



(a) Clustering matrix by P-iPDA (b) Edge denotes a clustering prob. higher than 90% (c) Edge denotes a clustering prob. between 70% and 90%

Figure 2: (a) Clustering matrix for all channel pairs by 1000 Hz tone stimulus. Each element in the symmetric matrix is the percentage of two regions clustered together by P-iPDA across 246 1000 Hz trials. (b-c) Networks constructed based on the clustering matrix (a). Each node represents one recording channel, the red ones are epileptic areas, and the black are auditory responsive areas. Each edge between two channels in (b) and (c) respectively indicates that they are clustered into the same module by P-iPDA in more than 90%, and between 70% and 90% of trials.



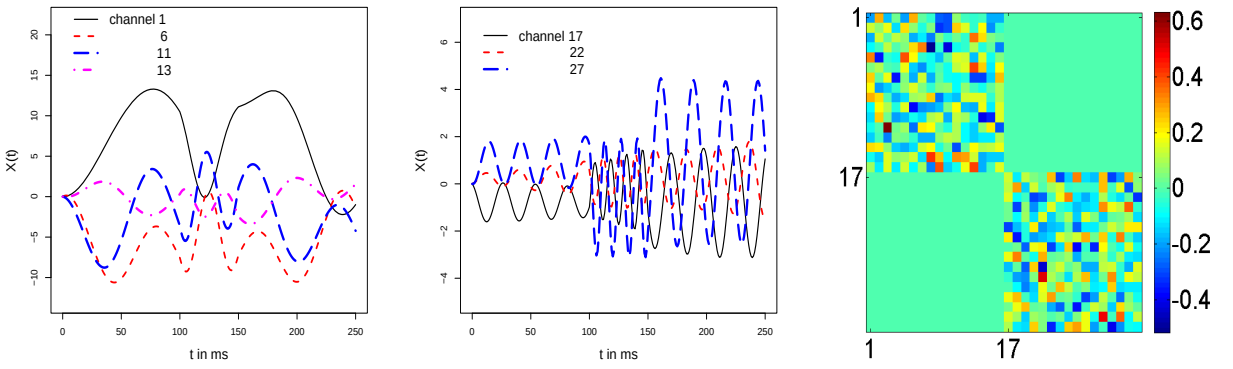
(a) Averaged $\hat{\mathbf{A}}$ estimated by P-iPDA (b) Averaged $\hat{\mathbf{B}}$ estimated by P-iPDA

Figure 3: The averaged estimates of $\mathbf{A}$ and $\mathbf{B}$ across 246 trials by P-iPDA in the real data analysis.

are *not* linearly dependent. All the channels in each module are pairwise connected, that is, none of the elements of $\mathbf{A}$ and $\mathbf{B}$ within each module is zero. The values of $\mathbf{A}/\mathbf{B}$ are shown in Figure 4(c). We conducted 100 independent simulations. Given the parameters $\boldsymbol{\theta}_2$, 32 time series $\mathbf{x}(t)$ are generated by discretizing the PDDM (5) using numerical approximation. Since the underlying system is deterministic, $\mathbf{x}(t)$ is identical across 100 simulations. Within each simulation, 32 independent error time series $\boldsymbol{\epsilon}(t)$, each following an AR(1) model with a lag-one correlation of 0.5, are generated. We adjusted the variance of each $\epsilon_i(t)$ such that the SNR—defined as $\text{var}(x_i(t))/\text{var}(\epsilon_i(t))$—equals 10. We set SNR at 10, because the estimated SNR of the real data, $\text{var}(\hat{x}_i(t))/\text{var}(\hat{\epsilon}_i(t))$, is above 10 across all trials and channels. Then for each i, the sum of $x_i(t)$ and $\epsilon_i(t)$ yields $y_i(t)$.

We first selected the penalty parameters based on one simulation through the same procedure as in the real data analysis, and used the selected λ and μ for analyzing 100 simulations. Figure 5(a) summarizes the percentage of each pair of channels clustered into the same module by P-iPDA, and Figures 5(b) and 5(c) respectively present the corresponding networks using different thresholds for frequencies: higher than 90% and between 70% and 90%. The true positive rate of P-iPDA (the frequency of correctly detecting nonzero values of $\mathbf{A}$ and $\mathbf{B}$) is 81.5%, and false positive rate (the frequency of estimating zero values of $\mathbf{A}$ and $\mathbf{B}$ incorrectly nonzero) is 0. Overall, P-iPDA successfully detected two clusters, though it occasionally missed clustering one or two channels, which is possibly due to the multicolinearity among $\mathbf{x}(t)$ in the same cluster.

We evaluated and compared the biases and variances of $\hat{\mathbf{A}}$ and $\hat{\mathbf{B}}$ estimated by P-iPDA and iPDA, which are summarized in Table 1. For easy comparison, we used the same λ in the two methods. In Example 1, P-iPDA produced estimates with slightly smaller biases and much smaller variances than those by iPDA. The reasons are two fold. First, if $\mathbf{x}(t)$ is known and P-iPDA correctly identifies all interactive channels, the regression models outputted from P-iPDA, where $dx_i(t)/dt$ of each channel i is the response, and $\mathbf{x}(t)$ in the same cluster as i are the predictors, are equivalent to those in iPDA. In addition, since estimated $\hat{\mathbf{x}}(t)$ by P-iPDA and iPDA based on identical λ take similar values, the estimates of the above mentioned regression coefficients, i.e., $\mathbf{A}$ and $\mathbf{B}$, by the two methods have similar means and biases. Second, with the



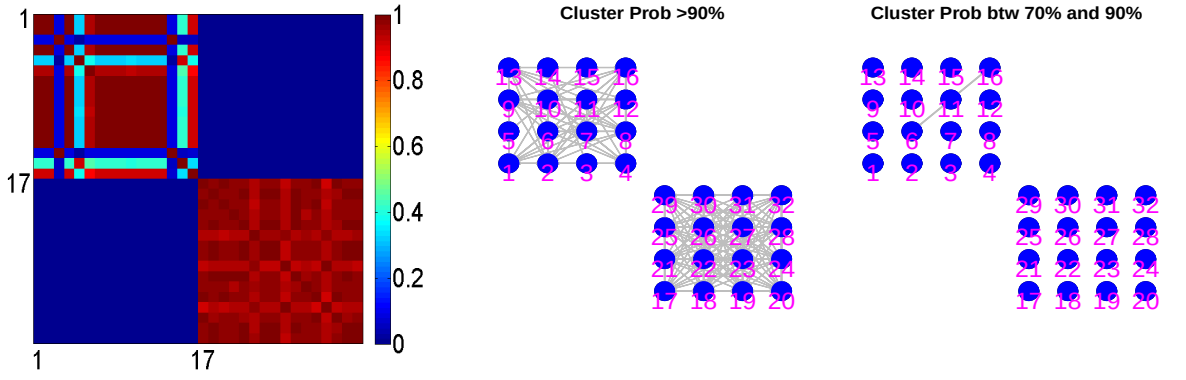
(a) First Module (b) Second Module (c) The matrix $\mathbf{A}/\mathbf{B}$

Figure 4: (a) and (b) show temporal changes of the simulated state function $x(t)$ of several channels in the first and second modules from Example 1. (c) displays the values of $\mathbf{A}/\mathbf{B}$ in Example 1.

sparsity constraint, P-iPDA uses much fewer predictors in the regression models than iPDA, and thus the ensuing estimates have much smaller variances. Other than achieving better estimation efficiency than iPDA, P-iPDA, by partitioning a large network into several independent smaller ones, also takes much less computational time.

Simulation Examples	Parameters	Average Bias		Average Standard Deviation	
		P-iPDA	iPDA	P-iPDA	iPDA
1	Matrix $\mathbf{A}$	0.08	0.10	0.26	0.36
	Matrix $\mathbf{B}$	0.14	0.15	1.18	1.02
2	Matrix $\mathbf{A}$	2.00	2.03	0.20	0.61
	Matrix $\mathbf{B}$	2.03	2.33	0.44	4.71

Table 1: The summaries of the biases and standard deviations of estimated $\mathbf{A}$ and $\mathbf{B}$ by P-iPDA and iPDA in two simulation examples.



(a) Clustering matrix by P-iPDA (b) Edge denotes clustering prob. larger than 90% (c) Edge denotes clustering prob. between 70% and 90%

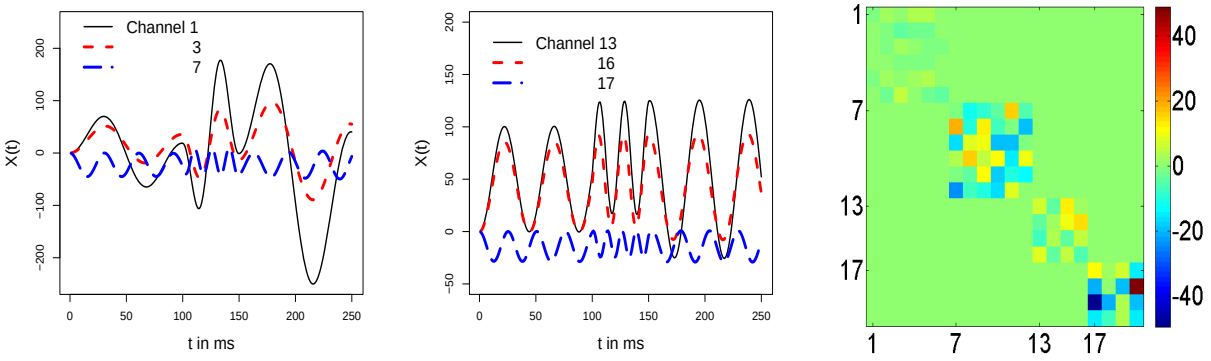
Figure 5: The (i, j)th (same as the (j, i)th) element of the 32×32 symmetric matrix in (a) represents the percentage of channels i and j in Example 1 clustered into the same module by P-iPDA among 100 simulations. Figures (b) and (c) are networks constructed based on the clustering matrix (a) with different thresholds: Each node represents one recording channel and each edge in (b) and (c) respectively indicates that the effect, exerted by one region on another, is estimated non-zero by P-iPDA in more than 90% and between 70% and 90% of the 100 simulations.

5.2 Example 2: multiple small clusters

The simulated dynamic system has 20 dimensions with 4 clusters of size 6, 6, 4, and 4. Figure 6(c) shows the values of $\mathbf{A} = \mathbf{B}$ used to generate 20 curves $\mathbf{x}(t)$, and Figures 6(a) and 6(b) show several $\mathbf{x}(t)$ from each cluster. Errors $\boldsymbol{\epsilon}(t)$ are generated in the same manner as those in Example 1 with SNRs of 10. Then for each i, the sum of $x_i(t)$ and $\epsilon_i(t)$ yields $y_i(t)$.

The parameter selection of λ and μ, and the analysis of data from 100 simulations follow the same procedure as in Example 1. Figure 7(a) summarizes the frequencies of each pair of components being clustered together by P-iPDA across 100 simulations and Figure 7(b) with 7(c) presents the associated networks constructed by using different thresholds for the clustering frequencies. Overall, the true positive rate of the P-iPDA in Example 2 is 97.5% and the false positive rate is 10.9%.

Comparing estimated $\hat{\mathbf{A}}$ and $\hat{\mathbf{B}}$ outputted from P-iPDA and iPDA, the former again achieved slightly smaller biases and much smaller variances than the latter, as shown in Table 1. We note



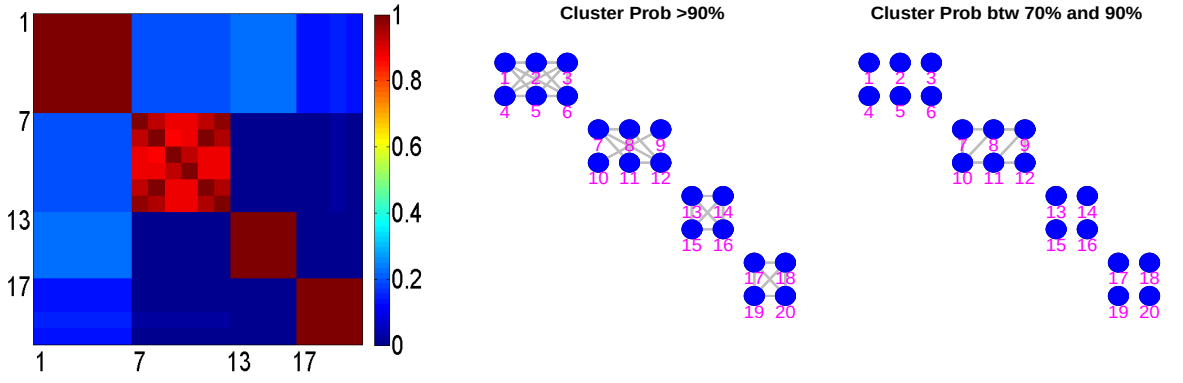
(a) First Two Modules (b) Second Two Modules (c) The matrix $\mathbf{A}/\mathbf{B}$

Figure 6: (a) and (b) show temporal changes of the simulated state functions $x(t)$ of several channels in Example 2. (c) displays the values of $\mathbf{A}/\mathbf{B}$ used to generate $\mathbf{x}(t)$ in Example 2.

that since Example 2 has a higher percentage of zero values in matrices $\mathbf{A}$ and $\mathbf{B}$ than that in Example 1, the reduction of estimation variability by P-iPDA, in comparison to iPDA, is more pronounced.

5.3 Example 3: ECoG of different lengths

We have also investigated how the clustering results by P-iPDA vary with different lengths of time series. Using the same $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$ in Example 2, two additional sets of $\mathbf{x}(t)$ were generated, one with $T = 100$ and another with $T = 500$. We let $u(t) = 1$ for $25 \leq t \leq 75$ in the former, and $u(t) = 1$ for $100 \leq t \leq 150$ and $400 \leq t \leq 450$ in the latter. For each set of $\mathbf{x}(t)$, we conducted 100 simulations in the same manner as in Example 1, summarized the clustering frequencies by P-iPDA for each of two sets of $\mathbf{x}(t)$ respectively in Figures 8(a) and 8(d), and presented networks constructed using different thresholds for clustering frequencies in Figures 8(b), 8(c), 8(e), and 8(f). When the length of time series is reduced by half, the true positive rate of P-iPDA is decreased to 91.8% and false positive rate is increased to 20.8%. On the other hand, when the length is doubled, the true positive rate is increased to 99.7% and the false positive rate is significantly reduced to 2.4%. Overall, the length of the data affects the false positive rate more than the true positive rate. This is possibly due to the fact that the model-fitting errors are



(a) Clustering matrix by P-iPDA (b) Edge denotes clustering prob. larger than 90% (c) Edge denotes clustering prob. between 70% and 90%

Figure 7: The (i, j)th (same as the (j, i)th) element of the 20×20 symmetric matrix in (a) represents the percentage of channels i and j in Example 2 clustered into the same module by P-iPDA among 100 simulations. Figures (b) and (c) are networks constructed based on the clustering matrix (a) with different thresholds: Each node represents one recording channel and each edge in (b) and (c) respectively indicates that the effect, exerted by one region on another, is estimated non-zero by P-iPDA in more than 90% and between 70% and 90% of the 100 simulations.

more affected by missing truly interactive channels, but less affected by including non-interactive channels, and thus P-iPDA tends to cluster more channels together when the data information is limited.

ln summary, P-iPDA achieved a higher true positive rate in Examples 2 and 3 than that in Example 1. In general, P-iPDA is most effective for cluster structures consisting of small clusters: the smaller the clusters, the fewer iterations that P-iPDA takes to identify the clusters, and less computation needed for estimating the spline-basis coefficients and model parameters.

6 Discussion

We propose a differential-equation-based dynamic model for ECoG data to study directional effects between brain regions, and introduce a new Potts model for the state equations in the DDM to identify functionally segregated brain networks. The high spatial and temporal resolution of

the ECoG data allows the dynamic model to have a simple structure that can accommodate a large number of brain components, unlike related DCMs in other modalities. We represent the neuronal states of brain components by cubic spline bases and estimate the model by minimizing a log-likelihood-based criterion, for which we have developed an iterative optimizing algorithm. The Potts model is converted to the new Potts penalty in the penalized likelihood approach.

Penalty parameter selection is a data-dependent process. As the ECoG recordings analyzed in this article cover only a small brain area under one experimental paradigm with a simple auditory stimulus, brain auditory responses measured by ECoG tend to be stable across trials (Flinker *et al.*, 2010), and consequently very similar penalty parameters were selected. However, in more common modalities such as fMRI and EEG, significant heterogeneity—due to the longer length or larger brain area covered—in the underlying effective connectivity across trials is widely reported (e.g. Duann *et al.*, 2002; Truccolo *et al.*, 2002; Turetsky *et al.*, 1988). In such cases, penalty parameter selection should be conducted separately for each trial, which will likely lead to different values being selected for different trials. Since parameter screening and cross validation are performed independently for each combination of parameter and each data point, this process can be parallelized, for example, using GPU computing to reduce the computational cost.

While the proposed method is motivated from and applied to the ECoG data, the statistical methodology, particularly the PDDM, can be applied to a broad range of applications using multivariate time series. First, the Potts penalty, in fact, does not rely on the linearity of the ODE model assumed in this analysis, and can be used in any dynamic models. Second, the Potts penalty is also applicable to settings where the observation time T is smaller than $(J+1)\cdot d^2$, the number of parameters characterizing pairwise interactions between components, analogous to the "small n large p" paradigm. Indeed, when the number of parameters in each module is believed to be much smaller than T, one can start with the most economic PDDM in which each component forms an independent module, and thus requires the least number of parameters. Through similar optimizing procedures as that in Section 3.2, at each step one node is selected to be clustered with one existing module according to the ensuing criterion. Then the size of modules will increase and the ensuing number of modules will decrease until the criterion cannot

be optimized anymore. Such a procedure is comparable to a stepwise linear regression that adds one variable at a time, and can be used in the "small n large p" paradigm as long as the number of selected variables is much smaller than n. Third, the Potts penalty can be used for time series that are observed in segments. Even when the parameters are allowed to differ between segments, the penalty can still be used to identify modules as long as they are assumed to remain the same over the time.

We fit our PDDM to the ECoG time series observed over a very short (less than 1s) period of time, different from the common practice of fitting DCMs to long time series (often in hundreds of seconds) in fMRI. Thanks to its high temporal resolution, such ECoG time series, with a large number of total observations still, offers a unique opportunity for studying effective connectivity, for the following reasons. First, a brain system may change dramatically over a short period of time, inducing non-ignorable temporal changes in parameter values, or even modules. Second, even if there is no dramatic change in the brain system over a long period of time, the underlying brain activity is likely to deviate significantly from the assumed linear system. As such, model assumptions based on first or second order Taylor approximation are relevant in the context of dynamic systems evolving over a short time. Third, analyzing short time series allows us to avoid making strong assumptions on the parameters of the PDDM, such as a negative definite parameter matrix—commonly assumed in fMRI-based connectivity studies—in order to ensure a stable system over an extended period. Nevertheless, it is still feasible to investigate brain effective connectivity using data measured over a long time. One possible approach is to first divide the data into several much shorter periods, within each of which a separate linear model is assumed. Then, identify functionally independent modules using the PDDM. Finally, within each identified module, apply nonparametric regression methods to approximate the nonlinear and unknown relationship between instantaneous changes of neuronal states with themselves and the experimental input.

The PDDM specifies two separate parameters for the connection in each of two directions between any two components within the same cluster, and the associated estimation algorithm P-iPDA clusters two components together if the connection in any one direction is strong. It is

possible that the connection between some components within the same cluster is void in one direction, but strong in the other direction. Our current method, however, does not evaluate the statistical significance of the directional effects and thus cannot distinguish which underlying directional effect within a cluster is void or nonzero. One potential approach to address this issue is to conduct hypothesis testing on the estimates of the directional effects from P-iPDA. This procedure must take into account of the uncertainty in identifying the clusters by P-iPDA, which is non-trivial in practice. Another potential approach is to impose both Potts penalty and L_1 penalty (Tibshirani, 1996) on the parameters within clusters in the log-likelihood criterion. Though achieving simultaneous clustering and sparsity within clusters, this approach is computationally demanding with three penalty parameters to be selected, and thus may require more iterations to converge. These will be the focus of our future directions in high-dimensional ODE model estimation.

There are several other directions for improving the PDDM and the P-iPDA algorithm. First, the spatial information of brain regions can be incorporated into the Potts model, so that spatially-close regions are more favored to be clustered into one module. Second, our current practice of using identical penalty parameters for all the regions may not be suitable for brain networks comprising modules with distinct interactive patterns. One potential solution is to use adjustable and region-dependent penalty parameters. Another possibility is to modify P-iPDA such that the already-identified clusters can be removed from the optimizing function, and thus do not affect the estimation of other clusters. Third, the PDDM estimation is formulated as an optimization problem in the article; statistical inference such as confidence interval construction and hypothesis testing on the model parameters is not straightforward. As elucidated before, the Potts model defines a prior distribution for the DDM parameters, and thus inference of the PDDM can be naturally carried out within a Bayesian framework. Finally, the PDDM can be modified to allow for very few channels that have interactive activity with several clusters and act as the "hub" of the brain network.

Appendix

Computation of Step A, I.A., & II.A. First, let $\mathbf{Y} = (y_1(1), \ldots, y_1(T), y_2(1), \ldots, y_d(T))'$, $\mathbf{\Phi} = (\boldsymbol{\phi}(1), \ldots, \boldsymbol{\phi}(T))'$, a Td-by-pd matrix $\mathbf{Q} = \begin{pmatrix} \mathbf{\Phi} & & \\ & \ddots & \\ & & \mathbf{\Phi} \end{pmatrix}$, and $\boldsymbol{\gamma} = (\mathbf{\Gamma}(1), \ldots, \mathbf{\Gamma}(d))'$. Then

$$\sum_{i=1}^{d}\sum_{t=1}^{T}(y_i(t) - \mathbf{\Gamma}(i)\,\boldsymbol{\phi}(t))^2 = (\mathbf{Y} - \mathbf{Q}\boldsymbol{\gamma})'(\mathbf{Y} - \mathbf{Q}\boldsymbol{\gamma}). \tag{10}$$

Under the spline representation,

$$\frac{dx_i(t)}{dt} - \mathbf{A}(i)\,\mathbf{x}(t) - \sum_j u_j(t)\cdot\mathbf{B}_j(i)\,\mathbf{x}(t) - \mathbf{C}(i)\,\mathbf{u}(t) - D_i =$$

$$\mathbf{\Gamma}(i)\frac{d\boldsymbol{\phi}(t)}{dt} - \mathbf{A}(i)\mathbf{\Gamma}\boldsymbol{\phi}(t) - \sum_j u_j(t)\cdot\mathbf{B}_j(i)\,\mathbf{\Gamma}\boldsymbol{\phi}(t) - \mathbf{C}(i)\,\mathbf{u}(t) - D_i. \tag{11}$$

Define vectors with dp elements: $\mathbf{\Lambda}_i(t) = (A_{i1}\phi_1(t), \ldots, A_{i1}\phi_p(t), A_{i2}\phi_1(t), \ldots, A_{id}\phi_p(t))$, $\mathbf{\Upsilon}_{ij}(t) = (u_j(t)\cdot B_{j,i1}\cdot\phi_1(t), u_j(t)\cdot B_{j,i1}\cdot\phi_2(t), \ldots, u_j(t)\cdot B_{j,i1}\cdot\phi_p(t), u_j(t)\cdot B_{j,i2}\cdot\phi_1(t), \ldots, u_j(t)\cdot B_{j,id}\cdot\phi_p(t))$, and $\boldsymbol{E}_i(t) = (\mathbf{0}_p, \ldots, (\frac{d\boldsymbol{\phi}(t)}{dt})', \ldots, \mathbf{0}_p)$ where $\mathbf{0}_p$ is a zero vector with p elements, and the $(i-1)\cdot p + 1$th to $i\cdot p$th elements of $\boldsymbol{E}_i(t)$ are non zero. Then (11) can be rewritten as $\boldsymbol{E}_i(t)\,\boldsymbol{\gamma} - \mathbf{\Lambda}_i(t)\,\boldsymbol{\gamma} - \sum_j \mathbf{\Upsilon}_{ij}(t)\,\boldsymbol{\gamma} - \mathbf{C}(i)\,\mathbf{u}(t) - D_i$. Let $\mathbf{X}_i(t) = \boldsymbol{E}_i(t) - \mathbf{\Lambda}_i(t) - \sum_j \mathbf{\Upsilon}_{ij}(t)$. Then we have

$$\int\left(\frac{dx_i(t)}{dt} - \mathbf{A}(i)\,\mathbf{x}(t) - \sum_j u_j(t)\cdot\mathbf{B}_j(i)\,\mathbf{x}(t) - \mathbf{C}(i)\,\mathbf{u}(t) - D_i\right)^2 dt = \boldsymbol{\gamma}'\int\mathbf{X}_i'(t)\mathbf{X}_i(t)dt\,\boldsymbol{\gamma}$$

$$-2\int(\mathbf{C}(i)\,\mathbf{u}(t) + D_i)\cdot\mathbf{X}_i(t)dt\,\boldsymbol{\gamma} + \int(\mathbf{C}(i)\,\mathbf{u}(t) + D_i)\cdot(\mathbf{C}(i)\,\mathbf{u}(t) + D_i)dt.$$

In the above $\mathbf{X}_i'(t)\mathbf{X}_i(t)$ is a dp-by-dp matrix, and $\int\mathbf{X}_i'(t)\mathbf{X}_i(t)dt$ is also a dp-by$-dp$ matrix with integral taken at very element of $\mathbf{X}_i'(t)\mathbf{X}_i(t)$, and $\int(\mathbf{C}(i)\,\mathbf{u}(t) + D_i)\cdot\mathbf{X}_i(t)dt$ is defined in the same way. Let $\mathbf{M} = \sum_{i=1}^{d}\int\mathbf{X}_i'(t)\mathbf{X}_i(t)dt$ and $\mathbf{W} = \sum_{i=1}^{d}\int(\mathbf{C}(i)\,\mathbf{u}(t) + D_i)\cdot\mathbf{X}_i(t)dt$. Then we have $\mathbf{H}(\boldsymbol{\theta}) =$

$$(\mathbf{Y} - \mathbf{Q}\boldsymbol{\gamma})'(\mathbf{Y} - \mathbf{Q}\boldsymbol{\gamma}) + \lambda\cdot\boldsymbol{\gamma}'\mathbf{M}\boldsymbol{\gamma} - 2\lambda\cdot\mathbf{W}\boldsymbol{\gamma} + \lambda\cdot\sum_{i=1}^{d}\int(\mathbf{C}(i)\,\mathbf{u}(t) + D_i)\cdot(\mathbf{C}(i)\,\mathbf{u}(t) + D_i)dt$$

$$= \boldsymbol{\gamma}'(\mathbf{Q}'\mathbf{Q} + \lambda\cdot\mathbf{M})\boldsymbol{\gamma} - 2(\mathbf{Y}'\mathbf{Q} + \lambda\cdot\mathbf{W})\boldsymbol{\gamma} + \mathbf{Y}'\mathbf{Y} + \lambda\cdot\sum_{i=1}^{d}\int(\mathbf{C}(i)\,\mathbf{u}(t) + D_i)\cdot(\mathbf{C}(i)\,\mathbf{u}(t) + D_i)dt.$$

Then given $\mathbf{A}, \mathbf{B}_j, j = 1, \ldots, J, \mathbf{C}$, and $\mathbf{D}$, the minimizer $\hat{\boldsymbol{\gamma}}$ of $\mathbf{H}(\boldsymbol{\theta})$ is given by

$$\hat{\boldsymbol{\gamma}} = (\mathbf{Q}'\mathbf{Q} + \lambda \cdot \mathbf{M})^{-1}(\mathbf{Q}'\mathbf{Y} + \lambda \cdot \mathbf{W}').$$

Acknowledgements

Zhang's research was partially funded by the U.S. NSF DMS grants 1209118 and 1120756. Li's research was partially funded by the U.S. NSF DMS grant 1208983. Boatman-Reich's auditory ECoG studies were supported by NIDCD grant K24-DC010028 and a grant from the Johns Hopkins Science of Learning Institute (DBR). The authors are grateful to Dr. Deepti Ramadoss for assistance with Figure 1(a), and to the associate editor and two reviewers for constructive comments.

References

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

Anderson, J. (2005). Learning in sparsely connected and sparsely coded system. In *Ersatz Brain Project working note*.

Biegler, L., Damiano, J., and Blau, G. (1986). Nonlinear parameter estimation: a case study comparison. *AIChE Journal*, **32**, 29–45.

Boatman-Reich, D., Franaszczuk, P., Korzeniewska, A., Caffo, B., Ritzl, E., Colwell, S., and Crone, N. (2010). Quantifying auditory event-related responses in multichannel human intracranial recordings. *Frontiers in Computational Neuroscience*, **4**(4).

Bressler, S. and Ding, M. (2002). Event-related potentials. In *The handbook of brain theory and neural networks*, pages 412–415. John Wiley & Sons, Inc.

Bullmore, E. and Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, **10**(3), 186–198.

Burton, M. (2001). The role of inferior frontal cortex in phonological processing. *Cognitive Science*, **25**, 695–709.

Campbell, D. (2007). *Bayesian collocation tempering and generalized profiling for estimation of parameters from differential equation models*. Ph.D. thesis, McGill University, Montreal.

Cervenka, M., Franaszczuk, P., Crone, N., Hong, B., Caffo, B., Bhatt, P., Lenz, F., and Boatman-Reich, D. (2013). Reliability of early cortical auditory gamma-band responses. *Clinical Neurophysiology*, **124**(1), 70–82.

Chen, J. and Wu, H. (2008). Efficient local estimation for time-varying coefficients in deterministic dynamic models with applications to hiv-1 dynamics. *Journal of the American Statistical Association*, **103**, 369–384.

Daunizeau, J., O, D., 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., Harrison, L., and Friston, K. (2005). Modelling event-related responses in the brain. *NeuroImage*, **25**, 756–770.

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. (2000). *Scientific Computing with Ordinary Differential Equations*. Springer, New York.

Duann, J., Jung, T., Kuo, W., Yeh, T., Makeig, S., Hsieh, J., and TJ, S. (2002). Single-trial variability in event-related bold signals. *Neuroimage*, **15**(4), 823–35.

Durka, P., Ircha, D., Neuper, C., and Pfurtscheller, G. (2001). Time-frequency microstructure of event-related electro-encephalogram desynchronisation and synchronisation. *Medical & Biological Engineering & Computing*, **39**, 315–3211.

Edwards, E., Soltani, M., Deouell, L., Berger, M., and Knight, R. (2005). High gamma activity in response to deviant auditory stimuli recorded directly from human cortex. *Journal of Neurophysiology*, **94**, 4269–4280.

Fan, J. and Lv, J. (2010). A selective overview of variable selection in high dimensional feature space. *Statistica Sinica*, **20**, 101–148.

Flinker, A., Chang, E., Kirsch, H., Barbaro, N., Crone, N., and Knight, R. (2010). Single-trial speech suppression of auditory cortex activity in humans. *The Journal of Neuroscience*, **30**(49), 16643–16650.

Földiák, P. and Young, M. P. (1995). Sparse coding in the primate cortex. In 895-898, editor, *The Handbook of Brain Theory and Neural Networks*. The MIT Press.

Franaszczuk, P. and Bergey, G. (1998). Application of the directed transfer function method to mesial and lateral onset temporal lobe seizures. *Brain Topogra*, **11**, 13–21.

Friston, K. (1994). Functional and effective connectivity in neuroimaging: A synthesis. *Humman Brain Mapping*, **2**, 56–78.

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

Friston, K., Harrison, L., and Penny, W. (2003). Dynamic causal modelling. *NeuroImage*, **19**, 1273–1302.

Friston, K., Frith, C., Dolan, R., Price, C., Zeki, S., Ashburner, J., and Penny, W. (2004). *Human Brain Function, Section 4*. Academic Press, 2 edition.

Gelman, A., Bois, F., and Jiang, J. (1996). Physiological pharamacokinetic analysis using population modeling and informative prior distributions. *Journal of the American Statistical Association*, **91**, 1400–1412.

Girvan, M. and Newman, M. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America*, **99**(12), 7821–7826.

Graner, F. and Glazier, J. (1992). Simulation of biological cell sorting using a two-dimensional extended potts model. *Physical Review Letters*, **69**, 2013–2016.

Holland, P. (1986). Statistics and causal inference (with discussion). *Journal of the American Statistical Association*, **81**, 945–970.

Huang, Y. and Wu, H. (2006). A bayesian approach for estimating antiviral efficacy in hiv dynamic models. *Journal of Applied Statistics*, **33**, 155–174.

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.

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.

Korzeniewska, A., Franaszczuk, P., Crainiceanu, C., Kuś, R., and NE, C. (2011). Dynamic of large-scale cortical interactions at high gamma frequencies during word production: Event related causality (erc) analysis of human electrocorticography (ecog). *NeuroImage*, **56**(4), 2218–2237.

Li, L., Brown, M., Lee, K., and Gupta, S. (2002). Estimation and inference for a spline-enhanced population pharmacokinetic model. *Biometrics*, **58**, 601–611.

Mallat, S. and Zhang, Z. (1993). Matching pursuits with time-frequency dictionaries. *IEEE Transactions Signal Processing*, **41**, 3397–3415.

McIntosh, A. and Gonzalez-Lima, F. (1994). Structural equation modeling and its application to network analysis in functional brain imaging. *Humman Brain Mapping*, **2**, 2–22.

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

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**(5594), 824–827.

Milo, R., Itzkovitz, S., Kashtan, N., Levitt, R., Shen-Orr, S., Ayzenshtat, I., M, S., and Alon, U. (2004). Superfamilies of evolved and designed networks. *Science*, **303**(5663), 1538–1542.

Newman, M. (2003). The structure and function of complex networks. *SIAM Rev*, **45**, 167–256.

Newman, M. (2004). Fast algorithm for detecting community structure in networks. *Physical Review E.*, **69**, 066133.

Newman, M. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences of the United States of America*, **103**(23), 8577–8696.

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

Penny, W., Stephan, K., Mechelli, A., and Friston, K. (2004). Modelling functional integration: a comparison of structural equation and dynamic causal models. *NeuroImage*, **23**, 264–274.

Potts, R. (1952). Some generalized order-disorder transformations. *Mathematical Proceedings*, **48**, 106–109.

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.

Ramsay, J. and Silverman, B. (2005). *Functional Data Analysis*. Springer, New York.

Ramsay, J., Hooker, G., Campbell, D., and Cao, J. (2007). Parameter estimation for differential equations: A generalized smoothing approach (with discussion). *Journal of the Royal Statistical Society, Ser. B*, **69**, 741–796.

Reiss, P. and Ogden, R. (2007). Functional principal component regression and functional partial least squares. *Journal of the American Statistical Association*, **102**, 984–996.

Reiss, P. and Ogden, R. (2009). Smoothing parameter selection for a class of semiparametric linear models. *Journal of the Royal Statistical Society, Ser. B*, **71**, 505–523.

Rubin, D. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, **66**(1), 688–701.

Rubin, D. (1978). Bayesian inference for causal effects: The role of randomization. *The Annals of Statistics*, **6**(1), 34–58.

Shao, J. (1998). Convergence rates of the generalized information criterion. *Nonparametric Statistics*, **9**, 217–225.

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. *Clinical Neurophysiology*, **120**, 140–149.

Sporns, O. (2011). *Networks of the Brain*. The MIT Press, Cambridge, Massachusetts.

Sporns, O. (2013). Network attributes for segregation and integration in the human brain. *Current Opinion in Neurobiology*, **23**, 162–171.

Stephan, K., Harrison, L., Kiebel, S., David, O., Penny, W., and Friston, K. (2007). Dynamic causal models of neural system dynamics: current state and future extensions. *Journal of Biosciences*, **32**, 129–144.

Steven, B., Martinez, M., , and Parsons, L. (2006). Music and language side by side in the brain: a pet study of the generation of melodies and sentences. *European Journal of Neuroscience*, **23**, 2791–2803.

Stuss, D. and Levine, B. (2002). Aduct clinical neuropsychology: lessons from studies of the frontal lobes. *Annual Review of Psychology*, **53**, 401–33.

Swanson, L. (2003). *Brain Architecture: Understanding the Basic Plan*. Oxford University Press, New York.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Ser. B.*, **58**, 267–288.

Tononi, G., Sporns, O., and Edelman, G. (1994). A measure for brain complexity: relating functional segregation and integration in the nervous system. *Proceedings of the National Academy of Sciences of USA*, **91**, 5033–5037.

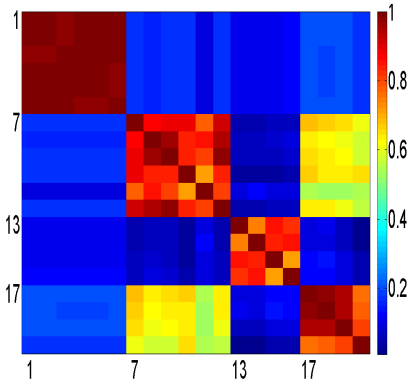
Truccolo, W., Ding, M., Knuth, K., Nakamura, R., and SL, B. (2002). Trial-to-trial variability of cortical evoked responses: implications for the analysis of functional connectivity. *Clinical Neurophysiology*, **113**(2), 206–226.

Turetsky, B., Raz, J., and Fein, G. (1988). Noise and signal power and their effects on evoked potential estimation. *Electroencephalogr Clin Neurophysiol*, **71**(4), 310–8.

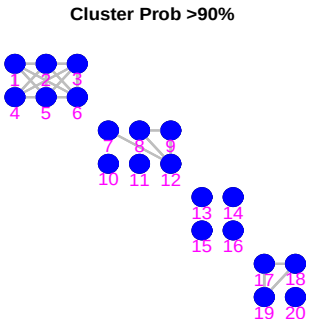
Varah, J. (1982). A spline least squares method for numerical parameter estimation in differential equations. *SIAM Journal of Scientific Computing*, **3**, 28–46.

Wahba, G. (1990). *Spline Models for Observational Data*. SIAM, Philadelphia.

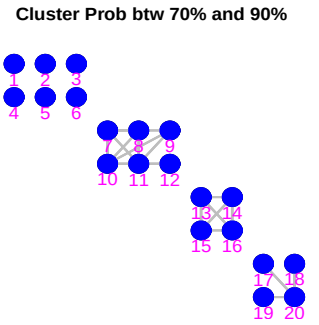
Wood, S. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society, Ser. B*, **73**, 3–36.



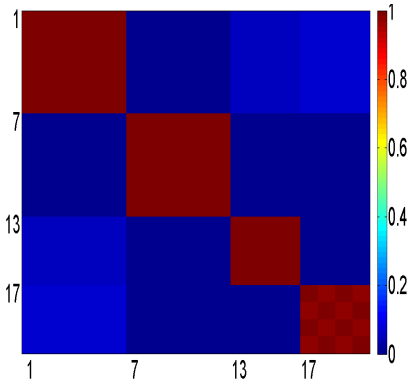
(a) Clustering matrix for $T = 100$



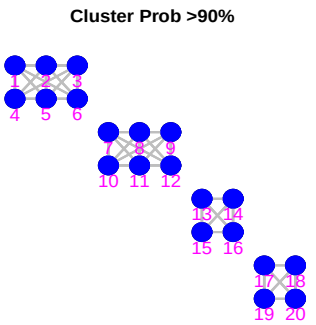
(b) Edge denotes clustering prob. larger than 90%



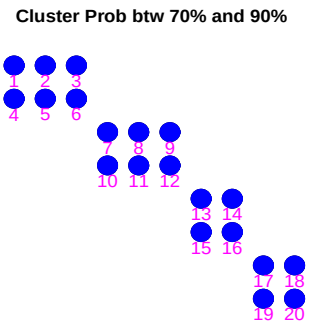
(c) Edge denotes clustering prob. between 70% and 90%



(d) Clustering matrix for $T = 500$



(e) Edge denotes clustering prob. larger than 90%



(f) Edge denotes clustering prob. between 70% and 90%

Figure 8: The simulated data use the same model parameters as Figure 7 of different lengths. The upper panels are based on the data with $T = 100$, and the lower use $T = 500$. The (i, j)th (same as the (j, i)th) element of the 20×20 symmetric matrix in (a)/(d) represents the percentage of channels i and j clustered into the same module by P-iPDA among 100 simulations. Figures (b)/(e) and (c)/(f) are networks constructed based on the clustering matrix (a)/(d) with different thresholds: Each node represents one recording channel and each edge in (b)/(e) and (c)/(f) respectively indicates that the effect, exerted by one region on another, is estimated non-zero by P-iPDA in more than 90% and between 70% and 90% of the 100 simulations.