Noname manuscript No.
(will be inserted by the editor)

Estimating Time-varying Directed Neural Networks

Haixu Wang · Jiguo Cao

Received: date / Accepted: date

Abstract Reconstructing the functional network of a neuron cluster is a fundamental step to reveal the complex interactions among neural systems of the brain. Current approaches to reconstruct a network of neurons or neural systems focus on establishing a static network by assuming the neural network structure does not change over time. To the best of our knowledge, this is the first attempt to build a time-varying directed network of neurons by using an ordinary differential equation model, which allows us to describe the underlying dynamical mechanism of network connections. The proposed method is demonstrated by estimating a network of wide dynamic range (WDR) neurons located in the dorsal horn of the rats' spinal cord in response to pain stimuli applied to the Zusanli acupoint on the right leg. The finite sample performance of the proposed method is also investigated with a simulation study.

Keywords Differential Equation · Neural Data Analysis · Network Reconstruction · Poisson Process · Spike Sequence.

1 Introduction

Reconstructing the functional network of a neural cluster is a fundamental step to reveal the complex interactions among neural systems in the brain. Establishing a reliable neural network can facilitate early detection of dementia and neurological disorders. Researchers can examine how the neural network of a patient is disrupted in comparison to a normally functioning network (de Haan et al., 2012). For example, examination can be searching for loss of connectivities or identifying disturbance in the network topology. Subsequently, treatments that either maintain or reverse disorders can be found by investigating whether they can control or restore the neural network.

Current approaches to reconstruct a neural network focus on establishing a static network by assuming the network structure does not change over time. One popular method is to calculate pair-wise correlations between individual neurons in the network, then it establishes a binary adjacency matrix for identifying edges of the network by thresholding or checking statistical significance of correlations (Siggiridou et al., 2014; Horvath, 2014; Bullmore and Sporns, 2009; de Haan et al., 2009). Correlation-based methods only provide a naive approximation to the complete network based on the pair-wise connections (Gerhard et al., 2011). In addition, connections estimated with the correlation-based methods are undirectional since they are obtained from a symmetric adjacency matrix.

A generalized linear model is proposed to estimate directional connections while conditioning on the population activities (Paninski, 2004; Truccolo et al., 2005). However, This approach ignores the dynamic nature of neural networks, and is not suitable for capturing highly variable and unknown connections. Connectivities between neurons are not necessarily always visible through entire observation period. Because the brain is a highly dynamical entity composed of various neural systems, the constructed neural network should be able to preserve any temporal changes in its structure. Hence, we need a reconstruction method to capture any time-varying neural networks.

The objectives of this article is to infer a time-varying directed neural network based on the neuron's spiking intensity. To the best of our knowledge, this is the first

Department of Statistics and Actuarial Science, Simon Fraser University, Burnaby, BC, Canada, V5A1S6. E-mail: jiguo_cao@sfu.ca

attempt to build a time-varying directed neural network by using a nonlinear ordinary differential equation (ODE) model.

The foremost contribution of this paper is to model the dynamical feature of the directed connectivities between neurons with a nonlinear ODE model. Instead of relying on a binary adjacency matrix to determine undirected connections, our model employs regulation functions to incorporate directional effects for each target neuron. In addition, regulation functions are exactly zero in certain time period when no regulation effects or connections are present. As a result, our model allows the topology of the constructed network to change over time. We propose a carefully-designed shrinkage technique called the functional smoothly clipped absolute deviation (fSCAD) to do three tasks simultaneously: detecting regulatory neurons for any target neuron, identifying the time period when regulation effects or connections are present, and estimating the nonlinear regulation function without any parametric assumption.

The paper is structured as follows: Section 2 includes the methodologies for inferring spiking intensities and construction of an ODE model. It also explains how to define regulation functions and the subsequent parameter selection and estimation of the ODE model. Section 3 presents a real data example for applying the ODE model framework. A simulation study is used to examine the finite-sample performance of the method and summarized in section 4, and conclusions are given in section 5.

2 Time-varying Directed Neural Networks

Neurons in our brains constantly fire electrochemical signals in response to external stimulations, for interactions with others, or simply under a resting state. Each spike results from a fluctuation in the membrane potentials once the potential reaches a certain threshold. The duration of spikes only lasts for a few milliseconds, and the shape of changes in membrane potential is usually identical for all spikes from a neuron. Hence, it is convenient to assume that information conveyed through neurons are coded in the timing of spikes rather than the shape of spikes. After converting a recording of membrane potential into a time sequence of identical spike events, we obtain a more common type of data. That is, a neuron's spike sequence can be viewed as a point process and investigated in this framework.

2.1 Estimating Spike Intensity Funcitons

The neuon's spike sequence, as a point process, can be completely characterized by its intensity function (Daley and Vere-Jones, 2003). Therefore, modeling spike sequences can be achived through establishing a model of the intensity function. Let $\mu_g(t)$, $t \in (0, T]$, be the intensity function of the spike sequence of the g-th neuron, $g = 1, \ldots, G$. We propose to use the smoothing spline method combined with penalization to estimate the spike intensity function given its flexibility in estimation. As a result of smoothing, the derivative of the spike intensity function is immediately available.

Suppose the g-th neuron, $g = 1, \ldots, G$, has m_g spikes in the time interval $[0, T]$. Let $s_{gj}, j = 1, ..., m_g$, represent the spike times for the g-th neuron. In addition, we take the ease to assume that the spike sequence behaves as an inhomogeneous Poisson process. Essential to constructing the likelihood function of a spike sequence, the probability distribution of waiting time w_j for next spike needs to be defined where $w_j = s_{g,j+1} - s_{gj}$. The distribution of w_j given the current spike time s_{gj} follows an exponential distribution with pdf (Kass et al., 2014)

$$\begin{aligned} & f(s_{g,j+1} - s_{gj} = w_j | s_{gj}, \mu_g(t)) \\ = & \mu_g(s_{g,j+1}) \exp[-\int_{s_{gj}}^{s_{g,j+1}} \mu_g(t) dt] \end{aligned} \tag{1}$$

Let $\boldsymbol{s}_g = (s_{g1}, s_{g2}, \ldots, s_{gm_g})^T$ be the vector containing all the spike times of the g-th neuron. The likelihood function $L(\mu_g(t)|\boldsymbol{s}_g)$ of a spike sequence $\boldsymbol{s}_g$ can be established based on (1):

$$\begin{aligned} & L(\mu_g(t)|\boldsymbol{s}_g) \\ = & f(s_{g1}, s_{g2}, \ldots, s_{gm_g}|\mu_g(t)) \\ = & f(s_{g1}|\mu_g(t)) f(s_{g2} - s_{g1} = w_1|\mu_g(t)) \\ & \cdots f(T - s_{gm_g} < w_{m_g}|\mu_g(t)) \\ = & \prod_{j=1}^{m_g} \mu(s_{gj}) \exp[-\int_{s_{g,j-1}}^{s_{gj}} \mu_g(t) dt] \exp[-\int_{s_{gm_g}}^{T} \mu_g(t) dt] \\ = & \prod_{j=1}^{m_g} \mu(s_{gj}) \exp[-\int_{0}^{T} \mu_g(t) dt] \end{aligned} \tag{2}$$

where the product term in (2) is the probability of the spike happening at s_{gj}, $j = 1, \ldots, m_i$, and the remaining term is the probability of no spike happening in the time period from s_{gm_g} to T.

Intensity functions are curves of conditional probabilities which should remain postive in estimations. Therefore, we model the logarithm of intensities $ln(\mu_g(t))$ as a linear combination of a set of basis functions. Using basis expansion method for smoothing allevaites the

requirement of any parametric assumptions. That is,

$$\ln(\mu_g(t)) = \sum_{k=1}^{K_g} c_{gk}\phi_{gk}(t) = \boldsymbol{c}_g^T \boldsymbol{\phi}_g(t) \tag{3}$$

where $\boldsymbol{\phi}_g(t) = (\phi_{g1}(t), \phi_{g2}(t), ..., \phi_{gK_g}(t))^T$ is a vector of basis functions and $\boldsymbol{c}_g = (c_{g1}, c_{g2}, ..., c_{gK_g})^T$ is the vector of basis coefficients. In this case, the B-spline functions are applied to compose a basis due to its compact support property (de Boor, 2001). This property will show its importance again in later sections. Each B-spline has a minimal support for which it is only nonzero within a short subintervals of the entire interval $[0, T]$. This feature will not only provide more efficiency in computation but also an estimation of intensity reflecting instantaneous spiking activities of neurons.

The B-splines are defined by both the degree of functions and the number of knots where the latter is more important. One strategy is to use regurization of penalization to allievate the burden of choosing optimal knot locations. Under penalization, we can put a knot on every spike times according to the smoothing spline method expalined by Wahba (1990). Alternatively, knots can be placed in relation to the density of observations, that is, more knots are located within intervals that have a dense cluster of spikes. In fact, there is great flexibity in the placement of knots given the introduction of a roughness penalty term to ensure the smoothness of the fitted curve estimates. The roughness penalty concerns about the curvature of a curve which is determined by the second derivative. Hence, the penalty is defined as

$$\text{PEN}(\mu_g(t)) = \int_0^T \left[\frac{d^2 \ln \mu_g(s)}{ds^2}\right]^2 ds.$$

corresponding to the total curve of an intensity fucntion. The time-varying spike intensity, $\mu_g(t)$, is then estimated from the spike time sequence, $\boldsymbol{s}_g = (s_{g1}, \ldots, s_{gm_g})^T$ by maximizing the penalized log-likelihood function

$$\begin{aligned} H(\boldsymbol{c}_g) &= \ln\left[L(\mu_g(t)|\boldsymbol{s}_g)\right] - \lambda_g \cdot \text{PEN}(\mu_g(t)) \\ &= \sum_{j=1}^{m_g} \ln \mu_g(s_{gj}) - \int_0^T \mu_g(s)ds - \lambda_g \int_0^T \left[\frac{d^2 \ln \mu_g(s)}{ds^2}\right]^2 ds \\ &= \sum_{j=1}^{m_g} \boldsymbol{\phi}_g^T(s_{gj})\boldsymbol{c}_g - \int_0^T \exp[\boldsymbol{\phi}_g^T(s)\boldsymbol{c}_g]ds - \lambda_g \boldsymbol{c}_g^T \boldsymbol{R}\boldsymbol{c}_g \end{aligned} \tag{4}$$

where λ_g is the smoothing parameter which controls the trade-off between fitting the data and smoothness of the fitted function. $\boldsymbol{R}$, a $K_g \times K_g$ is defined as

$$\boldsymbol{R} = \int_0^T \left[\frac{d^2\boldsymbol{\phi}_g^T(s)}{ds^2}\frac{d^2\boldsymbol{\phi}_g(s)}{ds^2}\right] ds.$$

An analytic solution to the penalized log-likelihood $H(\boldsymbol{c}_g)$ is not immediately available given that $H(\boldsymbol{c}_g)$ is not a linear function of the coefficients $\boldsymbol{c}_g$. Hence, maximization of $H(\boldsymbol{c}_g)$ will be achived through the application of numeric optimization, i.e., Newton-Raphson method. Given an initial value $\boldsymbol{c}_g^{(0)}$ of the basis coefficients, $H(\boldsymbol{c}_g)$ is numerically maximized by fiding the argument iteratively as follows

$$\boldsymbol{c}_g^{(\kappa+1)} = \boldsymbol{c}_g^{(\kappa)} - \left(\frac{d^2 H}{d\boldsymbol{c}_g^T d\boldsymbol{c}_g}\bigg|_{\boldsymbol{c}_g^{(\kappa)}}\right)^{-1} \left(\frac{dH}{d\boldsymbol{c}_g}\bigg|_{\boldsymbol{c}_g^{(\kappa)}}\right).$$

until convergence, where $\kappa = 0, 1, \ldots,$ represents the iteration indices. The analytical expressions of the derivatives involved in the Newton-Raphson iteration are given in the supplementary document.

Finding the optimal smoothing parameter λ_g in (4) can be done by cross-validation (Picard and Cook, 1984; Efron and Tibshirani, 1993). For each candidate of λ_g, we can compute the cross-validation score as follows

$$\begin{aligned} CV_1(\lambda_g) &= -\sum_{j=1}^{m_g} \ln p(s_{g,j} - s_{g,(j-1)}|\hat{\mu}_g^{(-j)}(t)) \\ &= -\sum_{j=1}^{m_g} \left[\ln \hat{\mu}_g^{(-j)}(s_{gj}) - \int_{s_{g,(j-1)}}^{s_{gj}} \hat{\mu}_g^{(-j)}(s)ds\right], \end{aligned}$$

where $\hat{\mu}_g^{(-j)}(t)$ is the predicted time-varying spike intensity for the g-th neuron after removing s_{gj} from the spike time sequence. λ_g is then chosen to be the one with smallest cross-validation score.

2.2 Modeling a Directed Time-varying Neural Network

Given that the intensity $\mu_g(t)$ and its derivative $\dot{\mu}_g(t)$ are available per the result of previous section, the time-varying directed neural network of G neurons is constructed via the following ordinary differential equation (ODE) model:

$$\dot{\mu}_l(t) = \alpha_l + \sum_{g=1}^{G} f_{gl}(\mu_g(t)), \;\; l = 1, \ldots, G, \tag{5}$$

where α_l is the intercept term and f_{gl}'s are non-linear functions reflecting regulation effects from g-th neuron to l-th neuron in the network. One great property of the ODE model is the decoupling of the network. Each l-th differential equation has its own set of regulation functions, hence we can analyze the directed effects from all neurons in the group to a specific one, and combine the results of individual dimensions to reconstruct the network. This ODE model was used to model the gene regulation networks in Nie et al. (2017).

In order to identify only the significant connections in the neural network, the network is equipped with a

sparsity assumption such that not all connections are significant, i.e., only some of f_{gl}'s are non-zero. Furthermore, we need to highlight the time-varying nature of the network. For that matter, the introduction of another sparisity assumption in each f_{gl} can ensure that regulation functions are only significant within some sub-intervals of the support $\mu_g(t)$. That is, the regulation effects are present only if the regulator neurons are firing within a certain range of intensities at the corresponding time. The latter sparsity assumption will naturally provide a time-varying network.

2.3 Estimating Regulation Functions

To exploit the smoothing spline method, the regulation funtions $f_{gl}(\mu_g(t))$ can be estimated without any parametric assuptions. Here, $f_{gl}(\mu_g(t))$ are represented by a linear combination of basis functions, i.e.,

$$f_{gl}(\mu_g(t)) = \sum_{k=1}^{K_{gl}} \beta_{glk}\phi_{glk}(\mu_g(t)) = \boldsymbol{\beta}_{gl}^T\boldsymbol{\phi}_{gl}(\mu_g(t)) \qquad (6)$$

where K_{gl} represents the number of basis functions, $\boldsymbol{\phi}_{gl}(\mu_g(t)) = (\phi_{gl1}(\mu_g(t)), \ldots, \phi_{glK_{gl}}(\mu_g(t)))$ is a vector of basis functions, and $\boldsymbol{\beta}_{gl} = (\beta_{gl1}, \beta_{gl2}, ..., \beta_{glK_{gl}})^T$ is the vector of basis coefficients. The choice of a basis for estimating $f_{gl}(\mu_g(t))$ should comply the second sparsity assumption. Forementioned B-spline basis is the most suitable basis. The compact support property limits basis functiosn to be non-zero only within some sub-intervals. Subsequently, it allows us to find sub-intervals of $\mu_g(t)$ for which $f_{gl}(\mu_g(t))$'s are non-zero once the estimates of basis coefficients are available.

An estimation strategy should be desgined to comply the objective of achieving two sparsity assumptions. The penalization method easily obliges the objective as it conveniently ensures the sparsity in both the network connections and the regulation functions by shrinking the basis coefficients $\boldsymbol{\beta}_{gl}$. If all basis coefficients $\boldsymbol{\beta}_{gl}$ are shrank to zero for g-th regulation function of l-th neuron, then $f_{gl}(\mu_g(t))$ will be entirely zero within $[0, T]$. Hence, g-th neuron does not possess regulation effects on the l-th neuron. If only some of $\boldsymbol{\beta}_{gl}$ are shrank to zero, then the regulation function $f_{gl}(\mu_g(t))$ is only significant in sub-intervals given the corresponding non-zero basis coefficients.

First, a criterion needs to be chosen for assessing the fidelity of regulation functions to the ODE model. For that matter, the least square criterion coveniently takes the role of measuring how well the regulation functions approximate the derivative of the intensity functions, and it is determined by the mean of sum of squared errors as

$$\frac{1}{n}\sum_{i=1}^{n}(\dot{\mu_l}(t_i) - \alpha_l - \sum_{g=1}^{G} f_{gl}(\mu_g(t_i)))^2$$

given a vector of observtion time points $(t_1, ..., t_n)$.

In order to accomplish sparsity in the coefficients, the functional SCAD penalty is chosen to be imposed on the regulation functions $f_{gl}(\mu_g(t))$. The functional SCAD penalty was first proposed by Lin et al. (2017) to estimate the coefficient function in the functional linear model. The nice shrinkage property of the functional SCAD penalty allows to locate region of the support with the regulation functions are non-zero without excessive shrinkage on the significant coefficients. Lin et al. (2017) focused on the estimation of a single function. In this article, we extend their method to estimate a large number of locally sparse regulation functions simultaneously. At last, the identification of significant connections in the network will be also achieved based on the estimation result.

A B-spline basis is defined by the degree of the basis functions and the number of knots where the latter is crucial to the calculation of fSCAD penalty. Let M_{gl} be the number of sub-intervals partitioned by the interior knots of B-splines, and define $\mu_{g,min}$ and $\mu_{g,max}$ to be the lower and upper bound of the intensity $\mu_g(t)$ of the g-th neuron and $\Delta_g = \mu_{g,max} - \mu_{g,min}$. The functional SCAD penalty for the l-th differential equation is defined as

$$\sum_{g=1}^{G} \frac{M_{gl}}{\Delta_g} \int_{\mu_{g,min}}^{\mu_{g,max}} p_\lambda(|f_{gl}(\mu_g)|)d\mu_g, \qquad (7)$$

where $p_\lambda(\cdot)$ is the SCAD penalty function (Fan and Li, 2001):

$$p_\lambda(|f|) = \begin{cases} \lambda|f| & \text{if} \quad 0 \le |f| \le \lambda, \\ -\frac{|f|^2 - 2\alpha\lambda|f| + \lambda^2}{2(\alpha-1)} & \text{if} \quad \lambda < |f| < \alpha\lambda, \\ \frac{(\alpha+1)\lambda^2}{2} & \text{if} \quad |f| \ge \alpha\lambda. \end{cases}$$

The parameter λ controls the sparsity of the regulation function through shrinking the coefficients.

The fSCAD penalty is calculated seperately for the set of regulation functions of each target neuron. Each fSCAD penalty is also a summation of inidividual penalty on each regulation function. Lin et al. (2017) states an approximation to the penalty of a single function, since the direct integration in (7) is not feasible. We can see that penalty defined in (7) is the product of a mean value and the number of sub-intervals partititioned by the knots of the basis functions. Subsequently, the fS-

CAD penalty for $f_{gl}(\mu_g)$ can be approximated as

$$\frac{M_{gl}}{\Delta_g}\int_{\mu_{g,min}}^{\mu_{g,max}} p_\lambda(|f_{gl}(\mu_g)|)d\mu_g \approx \sum_{m=1}^{M_{gl}} p_\lambda\left(\sqrt{\frac{M_{gl}}{\Delta_g}\int_{x_{g,m-1}}^{x_{gm}}[f_{gl}(\mu_g)]^2 d\mu_g}\right).$$

One great property of SCAD penalty is having flexible shrinkage on the estimated coefficients. The fSCAD penalty preserve this property by the approximation based on sub-intervals. That is, the approxiamtion is the summation of penalties on the L_2 norm of the regulation functions over each sub-interval. As a reulst, this partition allows the regulation functions to be shrank to zero where they are insignificant and estimated without shrinkage elsewhere.

In addition to the sparsity penalty, a roughness penalty can ensure that the estimated regulation functions are reasonably smooth and not proning to any particular points. Furthermore, the penalization will allievate the burden of choosing an optimal number of basis functions or equivalently the location of knots. For applications, one can simply choose a saturated number of basis functions with equally spaced knots as long as it is enough for approximating the previous fSCAD penalty. As commonly defined, the roughness penalty is calculated based on the curvature or the second order derivative of regulation functions as follows

$$\int_0^T \left(\frac{d^2 f_{gl}(\mu_g(t))}{dt^2}\right)^2 dt$$

The second order derivative is derived by using the chain rule and defined in the form of

$$\frac{d^2 f_{gl}(\mu_g(t))}{dt^2} = \sum_{k=1}^{K}\beta_{glk}\frac{d^2\phi_k(\mu_g(t))}{dt^2} = \sum_{k=1}^{K}\beta_{glk}\left[\frac{d^2\phi_k}{d\mu_g^2}\left(\frac{d\mu_g}{dt}\right)^2 + \frac{d\phi_k}{d\mu_g}\frac{d^2\mu_g}{dt^2}\right].$$

Since both the intensity and regulation functions are expressed as a linear combination of basis functions, individual derivatives in the above equation can be easily calculated. For each differential equation model, the total roughness penalty is the summation of individual ones times a tuning parameter γ for which a larger value penalizes more on roughness and provide a more smooth estimation of regulation functions.

The decoupling feature allows us to estimate the parameters of the ODE model (5) through the estimation of each differential equation. Let $\boldsymbol{\beta}_l = (\beta_{1l}^T, \beta_{2l}^T, ..., \beta_{gl}^T)^T$ represent the parameters of l-th differential equation. An objective function $Q(\boldsymbol{\beta}_l)$ of $\boldsymbol{\beta}_l$ is constructed by combining the least square criterion, the sparsity penalty, and the roughness penalty. That is,

$$Q(\boldsymbol{\beta}_l) = \frac{1}{n}\sum_{i=1}^{n}\left(\dot{\mu}_l(t_i) - \alpha_l - \sum_{g=1}^{G} f_{gl}(\mu_g(t_i))\right)^2 + \gamma\sum_{g=1}^{G}\int_0^T\left(\frac{d^2 f_{gl}(\mu_g(t))}{dt^2}\right)^2 dt + \sum_{g=1}^{G}\frac{M_{gl}}{\Delta_g}\int_{\mu_{g,min}}^{\mu_{g,max}} p_\lambda(|f_{gl}(\mu_g)|)d\mu_g\,. \quad (8)$$

Estimation and variable selection of $\boldsymbol{\beta}_l$ is achieved simultaneously by minimizing the objective function.

Identifiability issue arises when non-linear regulation functions are added together. That is, the prediction of each $\dot{\mu}_l(t)$ remains unchaged if a constant is added to a regulation function as long as it is beining subtraced from another function. Equivalently, the constant term α_l is not identifiable. This problem is also encountered for general additive models. Hence, a constraint is imposed on regulation functions for addressing the identifiability issue based on Wood (2017). The constraint states that the summation of reuglation functios for each l-th neuron should be zero, i.e.,

$$\sum_{g=1}^{G}\int_0^T f_{gl}(\mu_g(t))dt = 0 \quad (9)$$

which implies that $\hat{\alpha}_l = E(\dot{\mu}_l(t))$. Given that regulation functions are expressed as a linear combination of basis functions, the above constraint can be written in the form of

$$\sum_{g=1}^{G}\left(\sum_{k=1}^{K}\left[\beta_{glk}\sum_{i=1}^{n}\phi_k(\mu_g(t_i))\right]^2\right) = 0$$

where $(t_1, ..., t_n)$ are time points of observations of intensities. This constraint can be expressed in matrix notation as follows

$$\sum_{g=1}^{G}\left(\sum_{k=1}^{K}\left[\beta_{glk}\sum_{i=1}^{n}\phi_k(\mu_g(t_i))\right]^2\right) = \boldsymbol{\beta_l}^T\Psi\boldsymbol{\beta_l} \quad (10)$$

where Ψ is a $GK_{gl} * GK_{gl}$ matrix filled with block matrix $(\Psi_1, ..., \Psi_g, ..., \Psi_G)$ on its diagonals. Each block is a $K_{gl} * K_{gl}$ matrix with elements:

$$\Psi_{g[j,k]} = \left[\sum_{i=1}^{n}\phi_j(\mu_g(t_i))\right]\left[\sum_{i=1}^{n}\phi_k(\mu_g(t_i))\right]$$

To comply the identifiability constraint in (9), we can add (10) as a penalty term in the objective function by multiplying it with a large positive tuning parameter λ_I.

The range of intensity function $\mu_g(t)$, for $g = 1, ..., G$, varies from neuron to neuron, and so is the support of

$f_{gl}(\mu_g(t))$. In some occasions, the intensity is too close to zero which may cause problems for computation. For convenience, the intensity functions $\mu_g(t)$'s are normalized within a uniform range of $[0,1]$ by $\frac{\mu_g(t)-\mu_{g,min}}{\mu_{g,max}-\mu_{g,min}}$, so the regulation functions will have a uniform support. The derivative of intensity $\dot{\mu}_l(t)$ is standardized by its mean and standard deviation before fitting the model, and the estimation of $\hat{\alpha}_l$ is then omitted in the estimation process.

Fan and Li (2001) provides a unifed algorithm for minimizing the penalized least square functions through local quadrature aproximation. In order to execute the algorithm, the objective function should be expressed in matrix notations. The mean squared errors in (8) has a matrix representation after centering the response:

$$\frac{1}{n}\sum_{i=1}^{n}(\dot{\mu}_l(t_i)-\sum_{g=1}^{G}f_{gl}(\mu_g(t_i)))^2 = \frac{1}{n}(\dot{\boldsymbol{\mu}}_l-\boldsymbol{\beta}_l^T\boldsymbol{\phi}_l)^T(\dot{\boldsymbol{\mu}}_l-\boldsymbol{\beta}_l^T\boldsymbol{\phi}_l)$$

where $\dot{\boldsymbol{\mu}}_l=(\dot{\mu}_l(t_1),\dot{\mu}_l(t_2),...,\dot{\mu}_l(t_n))$ corresponds to the vector that contains derivatives at sample time points and $\boldsymbol{\phi}_l=(\Phi(\mu_1(t),...,\Phi(\mu_G(t)))^T$ is a $LK*n$ matrix with each $\Phi(\mu_g(t))$ contains K basis functions evaluated at n sample time points. Since all intensities are normalized to range $[0,1]$, sets of basis functions are the same across neurons, and K_{gl} reduces to K.

The fSCAD penalty for each regulation function can be approximated by the sum of penalty over equally spaced sub-intervals:

$$\frac{M_{gl}}{\Delta_g}\int_{\mu_{g,min}}^{\mu_{g,max}}p_\lambda(|f_{gl}(\mu_g)|)d\mu_g = \sum_{j=1}^{M_{gl}}p_\lambda\left(\sqrt{\frac{M_{gl}}{\Delta_g}\int_{x_{g,(j-1)}}^{x_{gj}}[f_{gl}(\mu_g)]^2d\mu_g}\right)$$

Denote the L_2 norm of function over j-th sub-interval as follows

$$||f_{gl,j}||_2=\sqrt{\int_{x_{g,(j-1)}}^{x_{g,j}}f_{gl}(\mu_g)^2d\mu_g}$$

Let $\boldsymbol{H}_{glj}$ be a $K*K$ matrix with elements $h_{glj[u,v]}$, which are defined as

$$h_{glj[u,v]}=\begin{cases}\int_{x_{g,(j-1)}}^{x_{g,j}}\phi_u(\mu_g)\phi_v(\mu_g)d\mu_g\,, & \text{if} \quad j\le u,v\le j+d\,,\\ 0\,, & \text{otherwise}\,,\end{cases}$$

then $||f_{gl,j}||_2=\sqrt{\boldsymbol{\beta}_{gl}^T\boldsymbol{H}_{glj}\boldsymbol{\beta}_{gl}}$. This is due to the fact that ϕ_u or ϕ_u are only non-zero in $d+1$ intervals based on their indices where d is the degress of basis fucntions. The fSCAD penalty can be then approximated by a quadratic function mentioned in Fan and Li (2001). Given an initial estimate of $\boldsymbol{\beta}_{gl}^{(0)}$, assumed to be close to the minimizer, with corresponding function estimate $f_{gl}^{(0)}$, the fSCAD penalty for any new parameter estimates $\hat{\boldsymbol{\beta}_{gl}}$ can be computed as

$$p_\lambda(|\hat{f_{gl}}|)\approx p_\lambda(|f_{gl}^{(0)}|)+\frac{1}{2}\frac{p'_\lambda(|f_{gl}^{(0)}|)}{|f_{gl}^{(0)}|}(|\hat{f_{gl}}|^2-|f_{gl}^{(0)}|^2)$$

Then, the penalty of each regulation function is calculated using LQA:

$$\begin{aligned}&\frac{M_{gl}}{\Delta_g}\int_{\mu_{g,min}}^{\mu_{g,max}}p_\lambda(|f_{gl}(\mu_g)|)d\mu_g\\ &=\sum_{j=1}^{M_{gl}}p_\lambda\left(\sqrt{\frac{M_{gl}}{\Delta_g}\int_{x_{g,(j-1)}}^{x_{gj}}[f_{gl}(\mu_g)]^2d\mu_g}\right)\\ &=\boldsymbol{\beta}_{gl}^t\boldsymbol{W}_{\boldsymbol{gl}}^{(0)}\boldsymbol{\beta}_{gl}+G(\boldsymbol{\beta}_{gl}^{(0)})\end{aligned}$$

where

$$\boldsymbol{W}_{\boldsymbol{gl}}^{(0)}=\frac{1}{2}\sum_{j=1}^{M_{gl}}\left(\frac{\dot{p}_\lambda(\left|\left|f_{gl,j}^{(0)}\right|\right|_2\sqrt{\frac{M_{gl}}{\Delta_g}})}{\left|\left|f_{gl,j}^{(0)}\right|\right|_2\sqrt{\frac{\Delta_g}{M_{gl}}}}\boldsymbol{H}_{glj}\right)$$

and

$$G(\boldsymbol{\beta}_{gl}^{(0)})=\sum_{j=1}^{M_{gl}}p_\lambda\left(\frac{||f_{gl,j}^{(0)}||_2}{\sqrt{\frac{\Delta_g}{M_{gl}}}}\right)-\frac{1}{2}\sum_{j=1}^{M_{gl}}\dot{p}_\lambda\left(\frac{||f_{gl,j}^{(0)}||_2}{\sqrt{\frac{\Delta_g}{M_{gl}}}}\right)\frac{||f_{gl,j}^{(0)}||_2}{\sqrt{\frac{\Delta_g}{M_{gl}}}}$$

The fSCAD penalty for all regulation function of l-th differential equation is the summation of individual penalties, that is,

$$\sum_{g=1}^{G}\frac{M_{gl}}{\Delta_g}\int_{\mu_{g,min}}^{\mu_{g,max}}p_\lambda(|f_{gl}(\mu_g)|)d\mu_g=\boldsymbol{\beta}_{\boldsymbol{l}}^T\boldsymbol{W}_{\boldsymbol{l}}^{(0)}\boldsymbol{\beta}_{\boldsymbol{l}}+\sum_{g=1}^{G}G(\boldsymbol{\beta}_{gl}^{(0)})$$

where $\boldsymbol{W}_{\boldsymbol{l}}^{(0)}=\text{diag}\{\boldsymbol{W}_{\boldsymbol{gl}}\}_{g=1}^{G}$.

Second order derivative of the regulation functiosn are derived by the chain rule and taking the form of

$$d_{glk}=\frac{d^2\phi_k}{d\mu_g^2}(\frac{d\mu_g}{dt})^2+\frac{d\phi_k}{d\mu_g}\frac{d^2\mu_g}{dt^2}\,,$$

Then, the roughness penalty for g-th regulation function is

$$\int_0^T\left(\frac{d^2f_{gl}(\mu_g(t))}{dt^2}\right)^2dt=\int_0^T(\sum_{k=1}^{K}\beta_{glk}d_{glk})^2dt$$

Similar to the case of estimating the intensity function, a $K*K$ matrix $\boldsymbol{V}_{\boldsymbol{gl}}$ is defined with elements $v_{ij}=\int_0^T d_{gli}*d_{glj}dt$. It follows that the roughness penalty can be evaluated using matrix multiplications as

$$\int_0^T\left(\frac{d^2f_{gl}(\mu_g(t))}{dt^2}\right)^2dt=\int_0^T(\sum_{k=1}^{K}\beta_{glk}d_{glk})^2dt=\boldsymbol{\beta}_{gl}^T\boldsymbol{V}_{\boldsymbol{gl}}\boldsymbol{\beta}_{gl}$$

Then, the summation of all smoothness penalties can be calculated as

$$\sum_{g=1}^{G}\int_0^T\left(\frac{d^2 f_{gl}(\mu_g(t))}{dt^2}\right)^2 dt = \boldsymbol{\beta}_l^T \boldsymbol{V}_l \boldsymbol{\beta}_l$$

where $\boldsymbol{V}_l = \text{diag}\{\boldsymbol{V}_{gl}\}_{g=1}^G$.

Adding the identifiability constraint defined in (10) into the penalized least square function (8) with smoothness and sparsity penalties, the objective function $\tilde{Q}(\boldsymbol{\beta_l})$ becomes

$$\frac{1}{n}(\dot{\boldsymbol{\mu}}_l - \boldsymbol{\beta}_l^T\boldsymbol{\phi}_l)^T(\dot{\boldsymbol{\mu}}_l - \boldsymbol{\beta}_l^T\boldsymbol{\phi}_l) + \lambda_I \boldsymbol{\beta_l}^T\boldsymbol{\Psi_l}\boldsymbol{\beta_l} + \gamma\boldsymbol{\beta_l}^T\boldsymbol{V_l}\boldsymbol{\beta_l} + \boldsymbol{\beta_l}^T\boldsymbol{W_l}^{(0)}\boldsymbol{\beta_l} + \sum_{g=1}^{G} G(\boldsymbol{\beta}_{gl}^{(0)}) \tag{11}$$

The one-step algorithm for estimating $\boldsymbol{\beta}_l$ is described as follows:

Step 1: Initialize with $\boldsymbol{\beta_l}^{(0)} = \frac{1}{n}\left(\frac{1}{n}\boldsymbol{\phi}_l^T\boldsymbol{\phi}_l + \lambda_I\boldsymbol{\Psi}_l\right)^{-1}$

Step 2: At each iteration i, compute $\boldsymbol{W}_l^{(i-1)}$ with $\boldsymbol{\beta}_l^{(i-1)}$. Then, the current estimate becomes

$$\boldsymbol{\beta}_l^{(i)} = \frac{1}{n}\left(\frac{1}{n}\boldsymbol{\phi}_l^T\boldsymbol{\phi}_l + \gamma\boldsymbol{V}_l + \boldsymbol{W}_l^{(i-1)} + \lambda_I\boldsymbol{\Psi}_l\right)^{-1}.$$

Step 3: Repeat Step 2 until the convergence of the estimation.

The initial value in the above algorithm is the minimizer of the least square function with addition of the identifiability constraint. If any coefficient is too small, then we are inverting a nearly singular matrix which yields unstable estimates. Hence, coefficients with magnitudes smaller than some prespecified threshold will be manually shrank to zero, and corresponding dimension will be removed from parameter estimation.

2.4 Tuning Parameter Selection

There are five tuning parameters to be considered: K the number of basis functions for expressing all regulation functions, λ_I the identifiability parameter, γ the smoothness parameter of regulation functions, and λ and α which control sparsity in coefficients through the fSCAD penalty. It would be extremely demanding to find an optimal K for each individual regulation functions. As a practical solution, we uniformly adjust the number of basis functions of all regulation functions to a single tuning parameter K. Intensity functions are estimted with smoothing procedure, and the estimates of intensity functions are equally dense in the observation interval and will be normalized within the interval $[0, 1]$. As a result, the interior knots for defining basis function for ϕ_{gl}'s can be located at an equal distance to each other. More basis functions will produce a more wiggly estimate of the regulation function, and a roughness penalty is commonly introduced to restrain the estimate from beining too complex. Also, the penalization appraoch will spare the effort of choosing optimal K. Thus, K is a lesser concern compared to the smoothness parameter γ in the tuning process. In our appraoch, K is fixed so that we can have enough sub-intervals to calculate the fSCAD penalty as a summation of penalty over sub-intervals. Also, K should be large enough for identifying entire zero intervals of the regulation functions. For that matter, K is left out of the automatic tuning process but becomes a subjective choice. In the following application and simulation, any K larger than 10 suffices for a satisfactory estimation.

The identifiability parameter λ_I is not crucial but only chosen to satisfy the identifiability constraint defined previously. The identifiability, smoothness, and sparsity parameters are selected based on a grid search strategy. The optimal set of tuning parameters is then determined based on the information criterion (AIC,BIC, or AICc). However, all information criteria are always improved with smaller sparsity parameter (less sparse coefficients) in our real data application. Instead of choosing parameter based on the minimum, we are selecting tuning parameters where the information criterion is leveling off where the criterion improves most dramatically.

The strategy is then to explore a single tuning parameter at a time and find where the criterion is improved significantly. Starting with the most important parameter λ, we collect the curves of AICc as a function of λ given every combination of other tuning parameters. A much finer grid will be designed for λ given that the parameter estiamtes are much more sensitive to λ than others. Then, we search for the point in the grid of λ where the the most improvement in AICc occurs among all curves. Subsequently, we employ the same strategy to the remaining parameters at a time while fixing the tuned parameters at their optimal.

3 Real Data Application

Experimental units were adult male Sprague-Dawley rats with weights of around 200 grams. Neurons of interest are wide dynamic range (WDR) neurons located in the dorsal horn of the spinal cord. They are believed to be responding to pain stimuli applied to the Zusanli acupoint on the right leg. After anesthetizing the

rat and exposing its lumbosacral spinal cord, microelectrodes were then vertically inserted into spine for recording electric signals from multiple WDR neurons (Qin et al., 2015). Pain stimulation was done by lifting and thrusting stainless steel needles to the acupoint for 120 times per minute. The entire experiment lasted for 16 minutes: (i) 2 minutes without any stimulation for rat to be in a resting state, (ii) 2 minutes for which the needle was only kept in the acupoint, (iii) 2 minutes for which pain stimulation was applied to the acupoint, (iv) 5 minutes after stopping with needled kept in the acupoint, and (v) 5 minutes with no stimulation for rats to return to a resting state. During the third phase of the experiment, mechanical stimulations are applied at a rate of 120 times per minute. We are able to observe obvious fluctuations in spiking activities for the majority of neurons compared to other experiment phases. Hence, we take a 20000ms segment, from the beginning of the third phase, of the whole observation for further anaylsis.

After sorting signals and detecting neuron spikes as done in (Qin et al., 2015), 12 neurons are identified along with their spiking sequences. Figure 1 displays the spike sequence data of 12 neurons for an observation period of 20000 ms. Neuron 3, 5, 8, and 12 only present a few spikes under stimulation while others have relatively dense observations. In addition, we can recognize a steady increase in rate of firing from most of neurons. Neuron 1, 7, 9, and 11 have a cluster of firings within a time period from 12000 ms to 15000 ms. Neuron 4,6, and 10 appear to have a delayed synchronization where cluster of spiking occurs after the previous group.

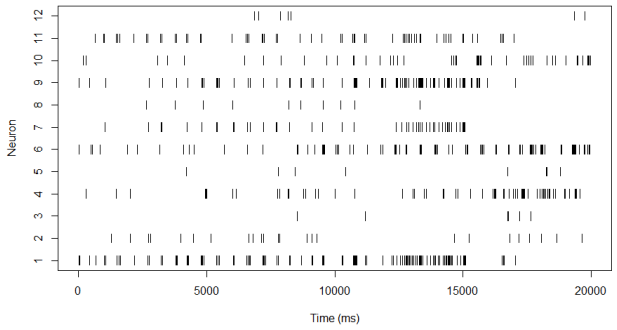

Fig. 1: Spike sequence data of 12 neurons. Each row corresponds to one individual neuron. Each short verticle line indicates one spike event.

First, we use the smoothing splines method explained in section 2.1 to estimate the spike intensity function and its derivative for each neuron. The logarithm of the

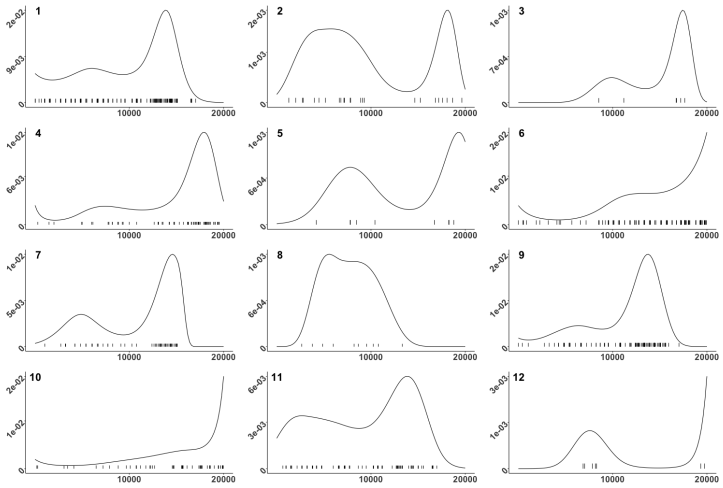

Fig. 2: Estimated spike intensity functions of 12 neurons. Bold numbers on the top left of each graph indicates the index number of neurons. The vertical line segments under the intensity curves represent spikes of neurons.

smoothing parameter γ_g with base 10 is chosen from $\{-5, -4, ..., 4, 5\}$. The consideration of number of cubic basis functions K ranges from 5 to 25, and the location of knots will be equally spaced along the observation interval. The optimal γ_g and K_g are chosen based on the CV score. Figure 2 displays the estimated spike intensity functions for 12 neurons. From the estimated intensity functions, we are able to identify two major groups of neurons in the network. The intensity functions of Neuron 1, 7, 9 and 11 have one major peak around time 15000ms. Their intensity function reflect the flucuations of spiking activities as shown in Figure 1. The delayed synchronization of neuron 4, 6, 10, 12 can be also observed from their intensity fuctions where the function either has a peak near the end of observation or a rising tail. Neuron 2 exihibits an opposite spiking behavior since it is more activie in the first half of observation. It is hard to interpret firing pattern of neuron 3 and 5 from the spiking train plot in Figure 1, but we can group them with neurons in delayed synchronization based on the peak of intensity functions.

We then estimate the regulation functions $f_{gl}(\mu_g(t))$ for the ODE model (5). The regulation functions are expressed as a linear combination of cubic B-spline basis functions with 9 equally-spaced interior knots. The basis coefficients of regulation functions are estimated by minimizing the penalized loss function (11). The search grid will consists of $10^{\{0,...,5\}}$ for λ_I, $10^{\{0,...,5\}}$ for γ, $\{0, ..., 5\}$ for α, and $\{0, 0.01, ..., 1\}$ for λ. Instead of searching for the minimum AICc among the expanded grid of combinations, a dramatic improvement in AICc

is more desirable as we search in the direction of each tuning parameter.

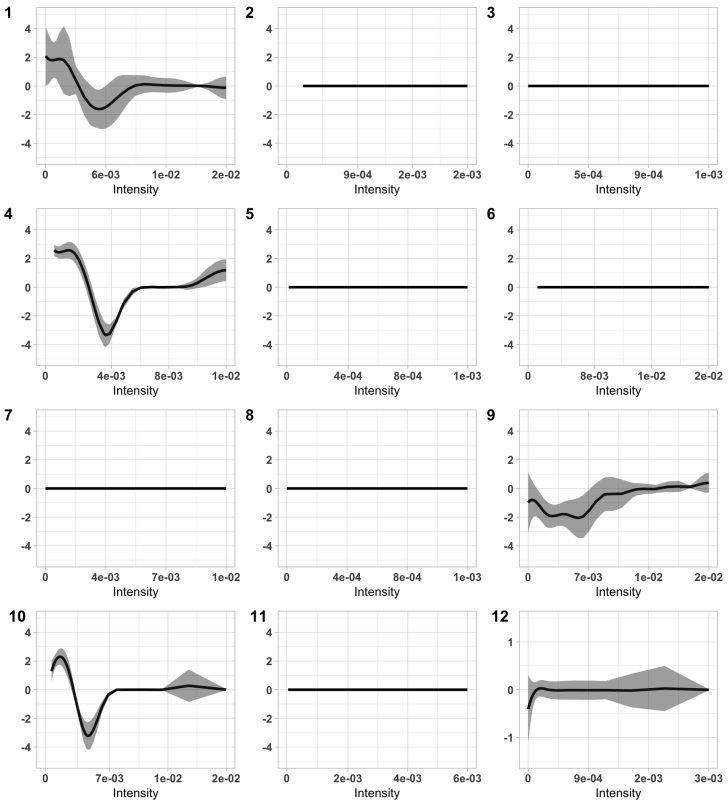

Fig. 3: Regulation functions for the target neuron 9. Bold numbers on the top-left corner indicate the number of regulation neuron. Shaded bands correspond to the pointwise 95% confidence intervals.

Figure 3 shows the estimated 12 regulation functions for the target neuron 9. It shows that 7 out of 12 regulation functions are entirely zero across the entire observation period, hence neuron 9 is only regulated by 5 neurons (neuron 1, 4, 10, and 12) including itself. In addition, some significant regulation functions are not entirely non-zero but only within a certain interval. When the regulation functions are zero, the corresponding network connection disappears. Therefore, the corresponding network connection is time-varying. The pointwise 95% confidence intervals are obtained with the boostrap method. A bootstrap sample is not taken directly from the intensity functions to estimate the regulation functions. Instead, we simulate a neural spiking sequence as a bootstrap sample based on the original intensity estimation. The estimated regulation functions for the other 11 target neurons are displayed in the supplementary file.

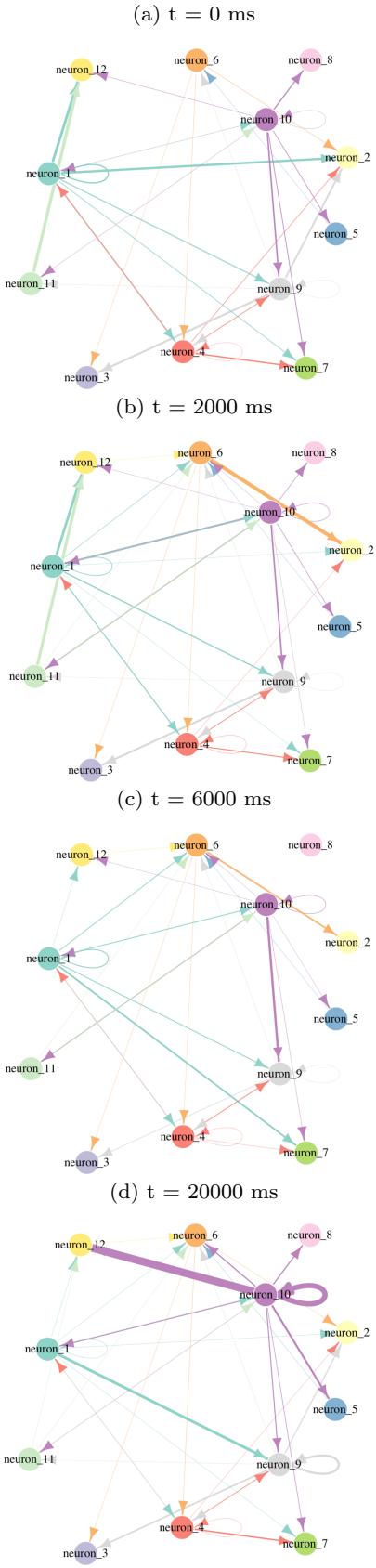

Fig. 4: The time-varying neuronal networks at the time 0, 2000 ms, 6000 ms, and 20000 ms. The arrow lines represent the regulation direction. The thickness of the lines represents the regulation strength defined by the magnitudes of the estimated regulation functions i.e, $|\hat{f}_{gl}(\mu_g(t))|$.

The time-varying neuronal network is displayed in Figure 4. The connectivity strength from the g-th to the l-th neuron is quantified by the magnitudes of the estimated regulation functions, i.e, $|\hat{f}_{gl}(\mu_g(t))|$ at any time t, where

$$\hat{f}_{gl}(\mu_g(t)) = \sum_{k=1}^{K_{gl}} \hat{\beta}_{glk}\phi_{glk}(\mu_g(t)).$$

It shows that neuron 1, 4, and 10 are constantly regulating other neurons, and these three neurons seem to be the core of networks in terms of coordinating the firing of others. Neuron 4 always has influences on neuron 1 and 10, whereas the influence from neuron 1 to 4 is periodical in both presence and magnitude. Neuron 7 is not regulating any other neurons but under the command of others. The animated time-varying network and snap shots of the network at different times are available in the supplementary materials.

4 Simulation Study

We assess the finite sample performance of the estimation procedure for recovering the edges of the neuronal network with a simulation study. The regulation functions in the ODE model (5) estimated from the real data are set as the true regulated functions. The ODE model (5) is then solved numerically to obtain the intensity functions of all 12 neurons. Then we generate synthetic neural spike sequences for all 12 neurons based on the obtained intensity functions. The algorithm of generating spike sequences are explained in the following paragraph. In this way, we can ensure that the simulated data are close to real observations.

For each neuron, synthetic spike sequences are generated by the thinning algorithm described in Lewis and Shedler (1979) based on its estimated spike intensity function by smoothing splines. The thinning algorithm generates a non-homogeneous Poisson process without the need of integrating the spike intensity function and can be used for any intensity function. The thinning algorithm generates the spikes as follows. Let A_g be the maximum for a spike intensity function $\mu_g(t), g = 1, \ldots, G$. First, we generate a homogeneous Poisson process according to the constant intensity A_g. Each spike in the simulated process is then accepted with the probability $\mu_g(t)/A_g$. The remaining spikes will resemble a non-homogeneous Poisson process with the intensity function $\mu_g(t)$.

We then estimate the spike intensity functions from the synthetic spike sequence with the smoothing spline method outlined in Section 2.1. The regulation functions in the ODE model (5) are estimated with the functional SCAD method proposed in Section 2.3. The basis functions and the searching grids for the tuning parameters are the same as the real data analysis. The simulaited data consists of 100 replicates.

Table 1: Selection accuracy for 12 target neurons in the simulation study. The numbers in the table represent how many times out of 100 simulations that a neuron is selected to be connected to the target neuron. Each column corresponds to a target neuron. Gray color indicates a true connection.

Neuron	1	2	3	4	5	6	7	8	9	10	11	12
1	87	55	99	100	98	100	97	0	73	100	98	100
2	0	34	0	0	0	38	6	0	0	0	0	0
3	0	0	0	0	0	68	0	0	0	0	0	0
4	99	100	99	34	40	63	100	52	100	3	100	100
5	0	23	1	0	1	99	0	0	0	0	2	1
6	42	70	100	100	99	99	75	100	0	100	97	100
7	46	85	13	34	39	44	33	0	9	0	96	2
8	28	29	2	3	99	3	32	0	1	0	0	0
9	27	84	100	98	40	93	49	52	28	100	100	56
10	100	100	100	100	40	10	100	52	100	100	100	100
11	24	54	87	34	0	100	37	0	0	3	0	100
12	73	25	16	0	0	69	31	0	100	100	93	100

Table 1 summarizes the selection accuracy for each of 12 target neurons in this simulation study. The neural network is time-varying, hence the target neuron l is not regulated by the g-th neuron only if the regulation functions $f_{gl}(\mu_g(t))$ is zero across the whole observation period. Alternatively, we conclude a connection between the l-th and the g-th neuron if there is any time that the regulation function is non-zero. By comparing the established connections from simulated and original data, we can assess the reliability of our proposed method. The selection result from simulations is considerably consistent with the data generating model. Spike sequences with few spikes are less likely to have regulation effects on other neurons, for example, neuron 2, 3, 5, and 8. However, we are able to identify where these neurons receive regulation. On the other hand, the esitmation is more consistency for spike sequences with denser observations.

Figure 5 displays the point-wise biases, standard deviations (SDs), and root mean squared errors (RMSEs) for regulation functions from neuron 1 and 10 to the target neuron 9. The bias of estimation appears to decrease as the intensity increases. This is reasonable because having more spikes in the sequence provides more reliable estimation of the spike intensity function and subsequently regulation functions. The point-wise variance of regulation functions also diminishes given a higher intensity. However, we can observe a great jump in the variance for the 10-th regulation function. This

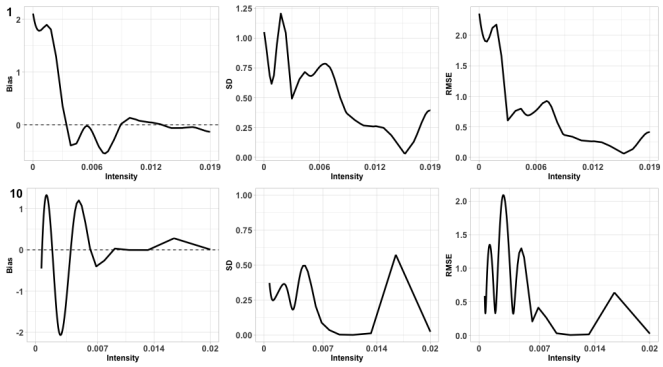

Fig. 5: The point-wise biases, standard deviations (SDs), and root mean squared errors (RMSEs) for regulation functions from neuron 1 (the first row) and 10 (the second row) to the target neuron 9.

is due to the distinct shape of the intensity function of neuron 10, which has a rapid increasing trend near the end. If simulated spike sequences of neuron 9 have more spikes scattered around the end of observation, then the regulation effects from neuron 10 will deviate from zero. Combining the bias and variance, we can observe that RMSE for both regulation functions decreases given the higher spike intensity.

5 Conclusions

This is the first attempt to construct a time-varying directed neural network through a sparse ODE model based on the intensity function of the observed neural spike sequence. First, the intensity function of the neural spike sequence can be estimated by smoothing splines. Furthermore, we express the influence from other neurons to a target neuron as regulation functions and assume an additive model to describe their relationship. Recovering the edges in the neural network can be accomplished by identifying the significant non-zero regulation functions for each target neuron. For that matter, the fSCAD penalty, accompanied with a smoothness penalty, is added to the objective function to estimate the sparse regulation functions, which have no parametric assumption. The fSCAD penalty can not only select significant regulation functions but also identify specific time intervals in which regulation functions are zero. Subsequently, we can construct a time-varying directed neural network by using the estimated sparse regulation functions.

The proposed method is demonstrated by estimating a network of wide dynamic range (WDR) neurons located in the dorsal horn of the rats' spinal cord in response to pain stimuli applied to the Zusanli acupoint on the right leg. We identify significant connections to each target neuron and visualize regulation effects through regulation functions. Subsequently, a time-varying directed neural network can be depicted at any points or animated during any intervals. Simulation studies show that our method can capture the majority of connections with satisfactory accuracy.

Supplementary Material

A supplementary document includes forementioned derivatives and addition real data application result. The computing codes for the application and simulation studies can be downloaded at
`https://github.com/caojiguo/NeuralNetwork`.

References

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

Daley, D. and D. Vere-Jones (2003). *An Introduction to the Theory of Point Processes*. New York: Springer-Verlag.

de Boor, C. (2001). *A Practical Guide to Splines*. New York: Springer.

de Haan, W., Y. A. Pijnenburg, R. L. Strijers, Y. van der Made, W. M. van der Flier, P. Scheltens, and C. J. Stam (2009). Functional neural network analysis in frontotemporal dementia and alzheimers disease using eeg and graph theory. *BMC Neuroscience 10*, 101.

de Haan, W., W. van der Flier, H. Wang, P. F. Van Mieghem, P. Scheltens, and C. J Stam (2012). Disruption of functional brain networks in alzheimer's disease: What can we learn from graph spectral analysis of resting-state magnetoencephalography? *Brain connectivity 2*, 45–55.

Efron, B. and R. J. Tibshirani (1993). *An Introduction to the Bootstrap*. New York: Chapman & Hall.

Fan, J. and R. Li (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association 96*(456), 1348–1360.

Gerhard, F., G. Pipa, B. Lima, S. Neuenschwander, and W. Gerstner (2011). Extraction of network topology from multi-electrode recordings: Is there a small-world effect? *Frontiers in Computational Neuroscience 5*, 4.

Horvath, S. (2014). *Weighted Network Analysis: Applications in Genomics and Systems Biology*. New York: Springer-Verlag.

Kass, R. E., U. Eden, and E. Brown (2014). *Analysis of Neural Data*. New York: Springer-Verlag.

Lewis, P. A. W. and G. S. Shedler (1979). Simulation of nonhomogeneous poisson processes by thinning. *Naval Research Logistics Quarterly 26*, 403–413.

Lin, Z., J. Cao, L. Wang, and H. Wang (2017). Locally sparse estimator for functional linear regression models. *Journal of Computational and Graphical Statistics 26*, 306–318.

Nie, Y., L. Wang, and J. Cao (2017). Estimating time-varying directed gene regulation networks. *Biometrics 73*, 1231–1242.

Paninski, L. (2004). Maximum likelihood estimation of cascade point-process neural encoding models. *Network: Computation in Neural Systems 15*(4), 243–262.

Picard, R. and D. Cook (1984). Cross-validation of regression models. *Journal of the American Statistical Association 79*, 575–583.

Qin, Qing an Wang, J., M. Xue, B. Deng, and X. Wei (2015). Charactering neural spiking activity evoked by acupuncture through state-space model. *Applied Mathematical Modelling 39*, 1400–1408.

Quiroga, R. Q., Z. Nadasdy, and Y. Ben-Shaul (2004). Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering. *Neural Computation 16*, 1661–1687.

Siggiridou, E., D. Kugiumtzis, and V. K. Kimishkidis (2014). Correlation networks for identifying changes in brain connectivity during epileptiform discharges and transcranial magnetic stimulation. *Sensors 14*, 12585–12597.

Truccolo, W., U. Eden, M. R. Fellows, J. P. Donoghue, and E. N. Brown (2005). A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects. *Journal of neurophysiology 93*, 1074–1089.

Wahba, G. (1990). *Spline Models for Observational Data*. Philadelphia: Society for Industrial and Applied Mathematics.

Wood, S. N. (2017). *Generalized Additive Models: An Introduction with R* (2nd ed.). London: Chapman and Hall/CRC.