Dynamic Graphical Models with Variable Selection for Effective Connectivity[*]

Rebecca Souza[†], Lilia Costa[‡], Marina Paez[§] João Sato[¶], and Candida Barreto[‖]

Abstract. This paper proposes a novel approach that combines dynamic linear models applied to graph data and variable selection through spike-and-slab priors. The new class of models, called *Dynamic Graphical Variable Selection*, is used to infer effective connectivity in certain brain regions allowing both connectivity weights and graphical structure to vary over time. One advantage of our method is that as the graphical structure is estimated inferentially, the computational cost is reduced. That way our methodology can accommodate high-dimensional data, such as large networks observed through long periods of time. We illustrate our methodology via numerical experiments with simulated and synthetic data, and then applied to fNIRS real data. The obtained results showed that the static version of our model is competitive against previous methodologies and demands a lower computational cost. Our model is more flexible than the previous methodologies by allowing the graphical structure to vary over time.

Keywords: dynamic variable selection, dynamic graphical models, effective connectivity, brain network, dynamic spike-and-slab priors, time series, network data.

1 Introduction

The neuroscience community has been developing studies about brain mapping techniques. Medical imaging methods such as functional magnetic resonance imaging (fMRI) (Glover, 2011) and functional near-infrared spectroscopy (fNIRS) (Ferrari and Quaresima, 2012) are widely employed in cognition research. Although these non-invasive neuroimaging techniques are notably widespread and have important clinical applicability, there is still much to learn about anatomical and functional relations between brain regions, known as brain connectivity. The identification of different types of brain connectivity, anatomical (structural), functional (correlational), and effective (causal) connections, is crucial to understand the complexity of the human brain (Rykhlevskaia et al., 2008). For this purpose, two recurring topics in the neuroscience literature are the occurrence of brain activity in a resting state (non-task condition), and the occurrence of brain activity when receiving stimulus (task condition). Fluctuations in

[†]Department of Statistical Methods, Federal University of Rio de Janeiro, BR, rebecca@dme.ufrj.br
[‡]Department of Statistics, Federal University of Bahia, BR, liliacosta@ufba.br
[§]Department of Statistical Methods, Federal University of Rio de Janeiro, BR, marina@dme.ufrj.br
[¶]Center for Mathematics, Computing and Cognition, Federal University of ABC, BR, joao.sato@ufabc.edu.br
[‖]South Africa National Research Foundation Research Chair at the University of Johannesburg, Faculty of Education, ZA

non-task conditions are not random (Damoiseaux et al., 2006) and abnormalities in the resting state connections may be associated with pathologies. Recent researches demonstrate alterations in functional connectivity associated with diseases, for example, chronic schizophrenia (Yanagi et al., 2020), attention-deficit and autism spectrum disorders (Ray et al., 2014), dementia or Alzheimer (Chabran et al., 2020; Filippi et al., 2020). On the other hand, there is an interest in discovering which brain regions are more activated or connected when performing a specific task such as continuous finger-tapping (Vergotte et al., 2017), solving the tangram puzzle (Hu et al., 2019), and a group of people beating drum simultaneously (Liu et al., 2021).

The study of graphs (networks) is strongly used to represent effective connectivity. Connections or links between two or more brain regions (nodes) can be estimated regarding strength and directionality, i.e. which region influences (parent node), which is influenced (child node), and the connection strength. For full details on graph theory see Diestel et al. (2010). Previous works are based mainly on Granger causality (GC) (Granger, 1969; Nakajima and West, 2015; Hu et al., 2019). Granger represents the causal influence as being the connectivity between the past with lag l and current observation. However Smith et al. (2011) show that lag-based approaches perform very poorly when estimating effective connections. This is because the sampling time of the measured signal is usually much slower than the underlying neuronal process. The authors indicate that successful lag-based causality estimations for brain activity data add to their simulations long lags for the majority of neural connections, with unrealistic lags of tens of seconds. Studies suggest that lag-based methods can yield reasonable results in relatively rare situations where neural connections have an effective lag greater than 100 ms, absence of hemodynamic response function (HRF) variability, high field strength and low repetition time (eg, 250 ms) (Rogers et al., 2010; Smith et al., 2011). Therefore, although these methods are regularly used in brain mapping analysis, their applications are very restricted and are usually accompanied by inappropriate simulations.

Instantaneous connectivity occurs at faster timescales than image processing can capture. In light of this, authors propose the application of *Dynamic Linear Models* (DLM) considering instantaneous connections. Besides not considering lagged relationships to estimate connectivity, dynamic networks allow time-varying connection strength. For instance, Costa et al. (2015) propose the *Multiregression Dynamic Model* (MDM) which builds a DLM for each child node associating a pre-established set of parent nodes as covariates. Furthermore, MDM is restricted to directed acyclic graphs (DAG). Such restriction is relaxed in the *Dynamic Graphical Models* (DGM) proposed by Schwab et al. (2018). MDM and DGM proposals present good estimates in realistic simulations of fMRI. Both have high sensitivity rates to identify active connections and perform well in identifying the directionality of connections, showing better results when compared to other lag-based methods. Nevertheless these proposals estimate the network topology through an exhaustive search, which can be challenging in very large networks, as the number of possible graphics increases exponentially as a function of the number of nodes. These proposals are unfeasible even with modern computational resources.

In the present study, we use *Dynamic Graphical Variable Selection* (DGVS) to estimate effective connectivity (considering instantaneous connections) in certain brain

regions. We propose a novel approach that combines DLM applied to graph data and variable selection through spike-and-slab priors. The main idea of DGVS is to assume a prior that all nodes belong to the set of parent nodes for each child node (except itself) and to impose sparsity for that set over time. Variable shrinkage strategy for the context of dynamic neural networks brings two major advantages: (1) the graphical structure is estimated inferentially, without a search engine, which reduces the computational cost even for large networks and long sampling times; (2) it allows for a time-varying graphical structure, which can be attractive when applied to task data. It should also be emphasized that DGVS accommodates both resting state and task networks.

We validate our methodology from simulated data. Furthermore, we investigate the accuracy of DGVS to identify the true graphical structure in synthetic data that have similar properties to real fMRI data. For the synthetic data, we use the data available in Schwab et al. (2018) and compare the DGVS's performance to DGM's. Finally, we applied our methodology to fNIRS data in which participants listened to musical stimuli of positive and negative emotional valences. We aimed to evaluate the effects of the stimulus on brain networks and compare the two styles of music regarding the differences caused in the graphic structures. Before discussing our proposal, we briefly expose some general characteristics of the data that the proposed model can handle.

1.1 Data Typology

There are some types of techniques to measure brain activity present in neuroscience. We pay special attention to two: fMRI and fNIRS. Both are non-invasive techniques that can dynamically measure brain activity in different brain regions, that is, they provide a time series for each volume element or voxel (cube of a few millimeters).

FMRI is a safe technique for measuring and tracing brain activities during resting state and task design. It measures the changes in blood circulation in the brain that happen with brain activity. Increased neural activity raises the blood oxygen level due to an increased energy demand that subsequently calls for more oxygen. The oxygen-binding hemoglobin concentration in the blood has different magnetic properties for oxygen-rich blood and oxygen-deficient blood. When the blood is more oxygenated, the signal is stronger, and vice versa. This phenomenon gives rise to the blood oxygenation level-dependent signal (BOLD), the time series variable measured by fMRI.

fNIRS is a tool composed of a light source that is coupled to the participant's head via either light-emitting diodes (LEDs) or through fiber-optical bundles (the optode) and with a light detector that receives the light after it has interacted with the tissue. FNIRS indirectly measures cortical activation based on changes in concentrations of oxygenated (oxyHb) and deoxygenated (deoxy-Hb) hemoglobin. The most commonly used method measures changes in the ratio of oxy-Hb to blood volume. fNIRS has advantages and disadvantages compared to fMRI. While fMRI has become the gold standard for human brain imaging, fNIRS stands out for its portability, ease to handle, robustness to noise, cost-effectiveness compared to other neuroimaging modalities, and bring functional imaging to more realistic environments. However, it is limited by its spatial resolution and depth of range.

Several statistical methods have been proposed to infer causal relationships from brain activity data. Applications for fMRI signals include commonly used methods such as *Dynamic Causal Modeling* (DCM; Friston et al. (2003); Li et al. (2011)) and *Bayesian Networks* (BN; Mumford and Ramsey (2014); Li et al. (2008)). The first is committed to constructing a realistic neuronal model and the second uses the concept of conditional independence among variables to define causal relationships. The article by Tak and Ye (2014) contains a detailed review of some statistical methods for the analysis of the fNIRS signal, which include signal processing methods: such as correlation-based methods (Cooper et al., 2012) and *Principal Component Analysis* (PCA)/*Independent Component Analysis* (ICA) (Wilcox et al., 2008; Patel et al., 2011); as well as statistical analysis methods: such as *Analysis Of Variance* (ANOVA) (Okamoto et al., 2004) and *Statistical Parameter Mapping* (SPM) (Friston et al., 1994). Recent applications use regression models to measure human social interaction, as in Barreto et al. (2021) where the authors propose a student fNIRS signal prediction model using the teachers' signal as predictors. Another example is the study developed in Nguyen et al. (2021) that analyzes the interpersonal neural synchrony between mothers and children in the problem-solving task from *Generalized Linear Mixed Models* (GLMM).

The remainder of the paper is organized as follows. We describe the DGVS models in Section 2 and review the Dynamic Spike-and-Slab priors in Section 3. The Inferential Procedure is developed in Section 4. Section 5 contains the numerical results for the simulated, synthetic, and real data. Finally, the discussion is drawn in Section 6.

2 Dynamic Graphical Variable Selection

In this section, we present the *Dynamic Graphical Variable Selection* models (DGVS) which can be regarded as an extension of the DGM of Schwab et al. (2018), with the inclusion of variable selection. The DGVS models node activity in a specific region of interest (ROI[1]) by considering time-varying connectivity weights. Such weights are the regression coefficients that reflect the effect of all other regions as covariates in a child node. The use of a shrinkage prior allows the selection of parent nodes for each child node at every period of time, specifying a sparse time-varying graphical structure.

We now define effective connectivity in the context of dynamic linear models (West and Harrison, 2006). Our model assumes conditional independence among brain activities of the regions, given the activities of their parent regions for each time step. This allows decomposing the n-dimensional model into n separate conditional models, where for each region we have a univariate DLM. Let $Y_t(r)$ be a scalar response (or the measured BOLD response) from region r at time t that is linked to a set of p known regressors $\mathbf{F}_t(r) = (1, \mathbf{Y}_t'(-r))'$ through the relation

$$Y_t(r) = \mathbf{F}_t'(r)\beta_t(r) + \varepsilon_t(r), \quad \varepsilon_t(r) \sim \mathcal{N}\left(0, v_t(r)\right), \tag{2.1}$$

where we denote X' as the transpose of X, $r = 1, \ldots, n$, $t = 1, \ldots, T$ and p corresponds to the maximum number of parents of the region r (all but itself) plus one

---

[1]The ROI concept is a common practice in neuroimaging studies. It consists of defining, previously to the study, the brain region to be investigated. For instance, if a person receives a visual stimulus it is expected that the occipital cortex should be activated.

for the intercept. The set of all possible parents of $Y_t(r)$ is represented by $\mathbf{Y}_t(-r) = (Y_t(1), \ldots, Y_t(r-1), Y_t(r+1), \ldots, Y_t(n))'$. The time-varying regression coefficients $\beta_t(r) = (\beta_{t1}(r), \ldots, \beta_{tp}(r))'$ measure the effective connectivity strength for each parent of region r, with $\beta_{t1}(r)$ being the intercept. The observational error $\varepsilon_t(r)$ is taken to be independent over t and comes from a normal distribution with mean zero and variance $v_t(r)$. The observational variances $v_t(r)$ are unknown and the precisions $\nu_t(r) = 1/v_t(r)$ are specified according to the following Markov evolution model (West and Harrison, 2006):

$$\nu_t(r) = \tau_t(r)\nu_{t-1}(r)/\delta, \quad \text{where} \quad \tau_t(r) \sim \mathcal{B}\left(\delta\eta_{t-1}(r)/2, (1-\delta)\eta_{t-1}(r)/2\right)$$
$$\text{and} \quad \eta_t(r) = \delta\eta_{t-1}(r) + 1, \tag{2.2}$$

where $\delta \in (0,1]$ is a discount factor, $\mathcal{B}(a,b)$ denotes a beta distribution with shape parameters a and b and the observation precision sequence $\nu_t(r)$ evolves stochastically from independent random results $\tau_t(r)/\delta$. This allows for closed-form updates of the prior and posterior distributions to $\nu_t(r)$ at each time, with minor changes to $\eta_{t-1}(r)$ which is discounted between successive updates. The discount factor δ controls the random fluctuation of observational precision with smaller values leading to greater variation in $\nu_t(r)$, and values near unity representing smoother estimates of the precision. The value $\delta = 1$ implies static variance for all t. Assuming that the observational variance is time-dependent supports the hypothesis that brain activity is subject to randomness over the time interval $t-1$ and t.

For each time we have np regression coefficients, and a set of $T \times np$ coefficients to be estimated. With the number of parameters increasing with the number of nodes, this regression model is vulnerable to overfitting with the increase in the number of regions. This can be particularly problematic here where the nodes are dynamically interconnected providing representations with poor out-of-sample predictive performance. In addition, it is known that not all regions are parents of the others and we need to select which coefficients shrink toward zero. A variable selection allows us to indicate which regions are relevant for the outcome at any given time. This returns the subjacent structure of the graph which is also dynamic since the regressors are allowed to enter and leave the model as time progresses.

In the context of Bayesian inference, there are some attractive alternatives to shrinkage priors. Among the most common are two component discrete mixture priors known as the spike-and-slab and continuous shrinkage priors. A particular example of continuous shrinkage priors is the horseshoe prior, where the coefficients follow a normal distribution with zero mean and variance composed of two global-local hyperparameters. The global hyperparameter allows all parameters to shrink to zero, while the local hyperparameter assumes a heavy-tailed half-Cauchy prior that allows some coefficients not to be shrunk (Piironen and Vehtari, 2017). Recent work expands horseshoe priors to dynamic models (Kowal et al., 2019; Molinari et al., 2022), and this can be easily adapted for estimating effective connections. On the other side, spike-and-slab priors are intuitively attractive, as it has parameters that are easily interpretable in practice. We adopt the Dynamic Spike-and-Slab (DSS) priors introduced by Rockova and McAlinn (2021). These priors use a mixture of stationary time series for each coefficient (as will

be seen in Section 3) driven by a binary latent indicator $\gamma_t(r) = (\gamma_{t1}(r), \ldots, \gamma_{tp}(r))'$ which signals the active and negligible coefficients. This way, Spike-and-Slab priors allow separating the regression coefficients into two groups: active and inactive. This grants flexibility and plausible interpretations in the context of graph models since active coefficients can be interpreted as relevant connectivity strengths. Additionally, experts can add their beliefs to the mixture weights indicating whether a region is naturally more influenced without necessarily interfering with the variance of the prior distributions of the coefficients.

3   Dynamic Spike-and-Slab Prior

In this section, we review the DSS priors for selection/shrinkage developed by Rockova and McAlinn (2021) and suggest some small changes to accommodate dynamic graphs in order to recover a graphical structure. Here we assume that regression coefficients $\{\beta_{tj}(r)\}_{t=1}^{T}$, $j = 1, \ldots, p$, follow independent and identical Gaussian DSS priors.

Given $\beta_{t-1,j}(r)$ and a binary indicator $\gamma_{tj}(r) \in \{0, 1\}$ (which signals the spike/slab membership in region r at time t), then $\beta_{tj}(r)$ assumes a mixture of the form

$$P(\beta_{tj}(r)|\beta_{t-1,j}(r), \gamma_{tj}(r)) = (1 - \gamma_{tj}(r)) \, P_0(\beta_{tj}|c\lambda(r)) + \gamma_{tj}(r) P_1(\beta_{tj}|\mu_{tj}(r), \lambda(r)), \tag{3.1}$$

where

$$\mu_{tj}(r) = \phi(r)\beta_{t-1,j}(r) \quad \text{with} \quad |\phi(r)| < 1 \tag{3.2}$$

and

$$\gamma_{tj}(r)|\beta_{t-1,j}(r) \sim Bernoulli\left(\theta_{tj}(r)\right). \tag{3.3}$$

From now on we suppress the index r for notational simplicity. The readers must keep in mind that all parameters for each region are estimated independently, as mentioned in the latter section. We assume the spike density $P_0(\beta_{tj}|c\lambda) \equiv \mathcal{N}(\beta_{tj}|0, c\lambda)$, i.e. it is equal to a Gaussian density concentrated at zero with low variability, obtained as $c \to 0$, which allows the inert coefficients to shrink to zero. The slab density $P_1(\beta_{tj}|\mu_{tj}, \lambda) \equiv \mathcal{N}(\beta_{tj}|\mu_{tj}, \lambda)$ is a Gaussian density with its mean defined as an autoregressive first-order lag polynomial and sufficiently large variance ($\lambda \gg c\lambda$). The variance of the regression coefficients distribution represents the oscillation of the connectivity strength. For the set of active coefficients, it is expected that the variance is far enough from zero to indicate a given brain activity. Furthermore, the time-varying indicator parameter γ_{tj} follows a Bernoulli distribution with success θ_{tj}, which controls the stability in the change between spike and slab distributions. Parameters ϕ and λ will be considered unknown and will be estimated.

This definition of the DSS induces the separation of the regression coefficients into two groups: the inactive ones centered in zero (when $\gamma_{tj} = 0$ and $P(\beta_{tj}|\beta_{t-1,j}, \gamma_{tj}) = \mathcal{N}(\beta_{tj}|0, c\lambda)$) and the active ones that follow an autoregressive path (when $\gamma_{tj} = 1$ and $P(\beta_{tj}|\beta_{t-1,j}, \gamma_{tj}) = \mathcal{N}(\beta_{tj}|\mu_{tj}, \lambda)$) (Rockova and McAlinn, 2021). In terms of graphs, the set of active coefficients represents the connectivity strength of the parents. As there will always be some independent brain activity, it is natural to assume that the intercept β_{t1} will be under the slab distribution, so that the first position of vector $\gamma_t(r)$ is equal to

one. Thus, the active coefficients and the intercept follow a stationary Gaussian $AR(1)$ process

$$\beta_{tj} = \phi\beta_{t-1,j} + w_{tj}, \quad |\phi| < 1, \quad w_{tj} \overset{iid}{\sim} \mathcal{N}\left(0, \lambda\right) \tag{3.4}$$

whose stationary distribution is characterized by

$$P_1^{ST}\left(\beta_{tj}|\phi,\lambda\right) \equiv \mathcal{N}\left(\beta_{tj}\bigg|0, \frac{\lambda}{1-\phi^2}\right). \tag{3.5}$$

The tractable stationary distribution above is an appeal of the conditional Gaussian slab distribution (Rockova and McAlinn, 2021).

An alternative to the autoregressive process is to consider that the connectivity strength follows a traditional random walk model, $\beta_{tj} = \beta_{t-1,j} + w_{tj}$, as performed in the works of Schwab et al. (2018) and Costa et al. (2015) for computational convenience. Bhattacharya and Maitra (2011), in their studies, argues that the random walk model can be too restrictive, failing to identify its parameters where the model is not so clearly distinguished from nonstationarity ($\phi = 0.95$ or $\phi = 0.999$). From the study of simulations, the authors confirm the effectiveness of the $AR(1)$ model against the random walk model. Authors such as Nakajima and West (2015) and Bhattacharya and Maitra (2011) use $AR(1)$ in applications to fMRI and electroencephalography (EEG) signals dataset respectively and get good fits.

To finish the formulation of DSS, we need to define the mixing weights θ_{tj} in (3.3). It is important that the choice between spike and slab distributions is stable and includes the information of previous values $\theta_{t-1,j}$ and $\beta_{t-1,j}$. The main idea is to obtain a sequence of slab probabilities that evolve smoothly over time without erratic switching. We resort to a stationary distribution to construct θ_{tj}.

Given $(\Theta, c, \lambda, \phi)$, we define the conditional inclusion probability as Rockova and McAlinn (2021)

$$\theta_{tj} = \frac{\Theta P_1^{ST}\left(\beta_{t-1,j}|\phi,\lambda\right)}{\Theta P_1^{ST}\left(\beta_{t-1,j}|\phi,\lambda\right) + (1-\Theta)P_0\left(\beta_{t-1,j}|c,\lambda\right)}, \tag{3.6}$$

where the scalar $0 < \Theta < 1$ is a tuning parameter that balances between the spike and slab distributions. Smaller values of Θ return greater shrinkage of coefficients. Note that if the module of the past value $|\beta_{t-1,j}|$ was large, then θ_{tj} will be close to one and it is more likely that the current value β_{tj} has a slab distribution, otherwise it will be more susceptible to the spike distribution. This way, θ_{tj} can be interpreted as the posterior probability of classifying the past coefficient $\beta_{t-1,j}$ into the stationary slab distribution.

Returning to the non-stationary slab distribution in (3.1), we can obtain the conditional probability

$$P^*(\beta_{tj}) \equiv \frac{\theta_{tj}P_1\left(\beta_{tj}|\mu_{tj},\lambda\right)}{\theta_{tj}P_1\left(\beta_{tj}|\mu_{tj},\lambda\right) + (1-\theta_{tj})P_0\left(\beta_{tj}|c,\lambda\right)}, \tag{3.7}$$

where $P^*(\beta_{tj}) = P(\gamma_{tj} = 1|\beta_{tj}, \beta_{t-1,j}, \theta_{tj})$, and it can be interpreted as the posterior probability of classifying β_{tj} into the slab distribution. Recall that if $|\beta_{t-1,j}|$ is large then θ_{tj} will be close to one and β_{tj} is likely to be a slab. Then, if $|\beta_{tj}|$ is in fact large,

$P^*(\beta_{tj})$ will be close to one and β_{tj} will shrink towards μ_{tj}. On the other hand, if $|\beta_{tj}|$ is small enough, $P^*(\beta_{tj})$ will shrink towards zero.

As an illustration, consider a graph with 3 nodes. The observation equations of the model are written as

$$
\begin{aligned}
Y_t(1) &= \beta_{t1}(1) + Y_t(2)\beta_{t2}(1) + Y_t(3)\beta_{t3}(1) + \varepsilon_t(1); \\
Y_t(2) &= \beta_{t1}(2) + Y_t(1)\beta_{t2}(2) + Y_t(3)\beta_{t3}(2) + \varepsilon_t(2); \\
Y_t(3) &= \beta_{t1}(3) + Y_t(1)\beta_{t2}(3) + Y_t(2)\beta_{t3}(3) + \varepsilon_t(3),
\end{aligned}
$$

where $\varepsilon_t(r) \sim \mathcal{N}(0, v_t(r))$ for $r = 1, 2, 3$. If the binary indicators were estimated as $\gamma_t(1) = (1,0,0)'$, $\gamma_t(2) = (1,1,0)'$ and $\gamma_t(3) = (1,1,1)'$, the evolution equations for region $r = 1$ would be

$$
\begin{aligned}
\beta_{t1}(1) &= \phi(1)\beta_{t-1,1}(1) + w_{t1}(1), \quad w_{t1}(1) \sim \mathcal{N}(0, \lambda(1)); \\
\beta_{t2}(1) &= w_{t2}(1), \quad \beta_{t3}(1) = w_{t3}(1), \quad w_{t2}(1), w_{t3}(1) \sim \mathcal{N}(0, c\lambda(1)),
\end{aligned}
$$

where only the intercept has a slab distribution, indicating that this region has no parents. For region $r = 2$, the evolution equations would be:

$$
\begin{aligned}
\beta_{t1}(2) &= \phi(2)\beta_{t-1,1}(2) + w_{t1}(2), \quad w_{t1}(2) \sim \mathcal{N}(0, \lambda(2)); \\
\beta_{t2}(2) &= \phi(2)\beta_{t-1,2}(2) + w_{t2}(2), \quad w_{t2}(2) \sim \mathcal{N}(0, \lambda(2)); \\
\beta_{t3}(2) &= w_{t3}(2), \quad w_{t3}(2) \sim \mathcal{N}(0, c\lambda(2)),
\end{aligned}
$$

where the intercept and the second coefficient have slab distribution, indicating that region $r = 1$ is a parent of the region $r = 2$. Finally, for region $r = 3$, we would have

$$
\begin{aligned}
\beta_{t1}(3) &= \phi(3)\beta_{t-1,1}(3) + w_{t1}(3), \quad w_{t1}(3) \sim \mathcal{N}(0, \lambda(3)); \\
\beta_{t2}(3) &= \phi(3)\beta_{t-1,2}(3) + w_{t2}(3), \quad w_{t2}(3) \sim \mathcal{N}(0, \lambda(3)); \\
\beta_{t3}(3) &= \phi(3)\beta_{t-1,3}(3) + w_{t3}(3), \quad w_{t3}(3) \sim \mathcal{N}(0, \lambda(3)),
\end{aligned}
$$

where all the coefficients have slab distribution, being regions $r = 1$ and $r = 2$ parents of region $r = 3$. The graphical structure of this example can be represented in Figure 1.
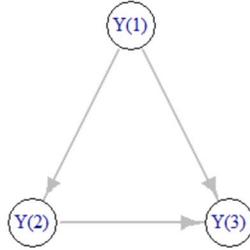

Figure 1: Example of a graphical structure considering 3 nodes, with node $Y(1)$ being the parent of nodes $Y(2)$ and $Y(3)$, and node $Y(2)$ being the parent of node $Y(3)$.

4   The Inferential Procedure

It is essential to develop efficient and fast estimation methods in the context of brain networks, as the number of nodes may increase considerably and make the inference unfeasible. Rather than using tools that generate posterior samples like the Markov chain Monte Carlo (MCMC) method (Kundu et al., 2019; Nakajima and West, 2015), in this paper, we propose estimation based on optimization. We focus on finding the maximum a posteriori probability (MAP) $\hat{\beta}_{1:T} = \arg \max P(\beta_{1:T}|\mathbf{Y}_{1:T})$ based on the dynamic expectation–maximization variable selection (EMVS) procedure of Rockova and McAlinn (2021) and Ročková and George (2014).

As in dynamic EMVS, we treat the binary indicators $\boldsymbol{\gamma_{0:T}} = \left\{\boldsymbol{\gamma_t}\right\}_{t=0}^{T}$ and the precision parameters $\nu_{1:T} = \{\nu_t\}_{t=1}^{T}$ as missing. The initial vector $\beta_0 = \{\beta_{0j}\}_{j=1}^{p}$ is estimated jointly with the remaining coefficients $\beta_{1:T}$. Also, we have introduced two extra estimation steps for the autoregressive parameter ϕ and the slab the variance λ. To simplify the algorithm, we factorize the prior distribution as

$$
\begin{aligned}
P(\beta_{0:T}, \gamma_{0:T}, \nu_{1:T}, \phi, \lambda) &= P(\beta_0|\gamma_0, \phi, \lambda) P(\gamma_0) P(\phi) P(\lambda) \\
&\times \prod_{t=1}^{T} \left[P(\nu_t|\nu_{t-1}) \prod_{j=1}^{p} P(\beta_{tj}|\beta_{t-1j}, \gamma_{tj}, \phi, \lambda) P(\gamma_{tj}|\beta_{t-1j}) \right],
\end{aligned}
$$

where $P(\beta_{tj}|\beta_{t-1j}, \gamma_{tj}, \phi, \lambda)$ and $P(\gamma_{tj}|\beta_{t-1j})$ are defined in (3.1) and (3.3), respectively. Previous works such as Rockova and McAlinn (2021) and Koop and Korobilis (2018) assume that the evolution variances are fixed and known. However, Narisetty et al. (2014) shows that this technique can lead to inconsistency in the selection as T and p grow. In light of this, based on the specifications of Koop and Korobilis (2020), we assume that λ has an non-informative Inverse Gamma prior, i.e.

$$
P(\lambda) \propto \lambda^{-(a_\lambda+1)} exp\left(-\frac{b_\lambda}{\lambda}\right) \qquad \text{with } a_\lambda = 3 \text{ and } b_\lambda = 200 \tag{4.1}
$$

implying a prior mean of $b_\lambda/(a_\lambda - 1) = 100$. For spike variance, we set $c \in \{0.01, 0.001\}$ which minimizes the mean squared one-step-ahead forecast errors (MSFE) to ensure that the spike distribution has a small variance. For the autoregressive parameter, we assume a reparametrized Beta prior (even as Rockova and McAlinn (2021) and Kim et al. (1998))

$$
P(\phi) \propto \left(\frac{1+\phi}{2}\right)^{a_\lambda-1} \left(\frac{1-\phi}{2}\right)^{b_\phi-1} \mathbb{I}(|\phi| < 1) \qquad \text{with } a_\phi = 20 \text{ and } b_\phi = 1.5, \tag{4.2}
$$

with a prior mean of $2a_\phi/(a_\phi + b_\phi) - 1 = 0.86$. This proposal of an informative prior is in agreement with the studies carried out by Phillips (1991) which show that flat priors on the autoregressive coefficients are informative in time series models, and indeed downweight large values and might result in instability in the estimates. Authors such as Nakajima and West (2013, 2015), Lopes et al. (2022) and Rockova and McAlinn (2021) also use a non-informative prior peaked around one. The latter updates ϕ with

a Metropolis step, using an uniform proposal on the interval $[0.8, 1)$. Here, we consider a grid of possible values for ϕ.

We consider a stationary distribution for β_0 (as in Rockova and McAlinn (2021))

$$P(\beta_0|\gamma_0, \phi, \lambda) = \prod_{j=1}^{p} \left[\gamma_{0j} P_1^{ST}\left(\beta_{0j}|\phi, \lambda\right) + (1 - \gamma_{0j})P_0\left(\beta_{0j}|c\lambda\right) \right], \qquad (4.3)$$

whereas $P(\gamma_{0j} = 1) = \Theta$, for each $1 \leq j \leq p$, being $\gamma_0 = (\gamma_{01}, \dots, \gamma_{0p})'$ independent. The optimization method adopted here starts the estimation of the parameters from pre-established initial values that are updated at each iteration. The choice of these values is fundamental for the proper functioning of the algorithm. We assume for the initial values of $\boldsymbol{\beta}_{0:T}$ random samples taken from an uniform distribution on the interval $[0, 2]$ and we admit a sequence of decaying values $\Theta \in \{0.9, 0.5, 0.1\}$, similarly to what was suggested by Rockova and McAlinn (2021). Specifically, the initialization of the parameters is systematized as follows: (1) we assume values for $\boldsymbol{\beta}_{0:T}$ randomly sampled from an uniform distribution and update these values considering the highest value of Θ in the sequence, $\Theta = 0.9$; (2) the results obtained considering $\Theta = 0.9$ are used as initial values for the estimations considering the next value of Θ in the sequence, which is $\Theta = 0.5$; (3) analogously to the previous step, the estimates obtained with $\Theta = 0.5$ are updated considering $\Theta = 0.1$. In so doing, in addition to decreasing the number of iterations needed for convergence, we also increase the chances of reaching appropriate modes.

Then, we have that the logarithm of the posterior distribution is as follows

$$\log P\left(\beta_{0:T}, \gamma_{0:T}, \nu_{1:T}, \phi, \lambda | Y_{1:T}\right) = C(v_{1:T}) + \sum_{t=1}^{T} \left\{ \log P(\nu_t|\nu_{t-1}) + \frac{(Y_t - F_t'\beta_t)^2}{2v_t} \right\}$$

$$+(a_\phi - 1)\log(1 + \phi) + (b_\phi - 1)\log(1 - \phi) - (a_\lambda + 1)\log\lambda - \frac{b_\lambda}{\lambda}$$

$$-\sum_{j=1}^{p} \left\{ \gamma_{0j} \left[\frac{1}{2}\log\left(\frac{\lambda}{1-\phi^2}\right) + \frac{(1-\phi^2)\beta_{0j}^2}{2\lambda} \right] + (1 - \gamma_{0j})\left(\frac{\log(c\lambda)}{2} + \frac{\beta_{0j}^2}{2c\lambda} \right) \right\} \quad (4.4)$$

$$-\sum_{t=1}^{T}\sum_{j=1}^{p} \left\{ \gamma_{tj} \left[\frac{1}{2}\log\lambda + \frac{(\beta_{tj} - \phi\beta_{t-1j})^2}{2\lambda} \right] + (1 - \gamma_{tj})\left(\frac{\log(c\lambda)}{2} + \frac{\beta_{tj}^2}{2c\lambda} \right) \right\}$$

$$+\sum_{j=1}^{p} \left[\gamma_{0j}\log\Theta + (1 - \gamma_{0j})\log(1 - \Theta) \right] + \sum_{t=1}^{T}\sum_{j=1}^{p} \left[\gamma_{tj}\log\theta_{tj} + (1 - \gamma_{tj})\log(1 - \theta_{tj}) \right].$$

The proposed expectation maximization (EM) algorithm is implemented following two steps. In the E-step, we calculate the expectation of the logarithm of the posterior distribution in (4.4) with respect to the current conditional distribution of $[\gamma_{0:T}, \nu_{1:T}]$ given $Y_{1:T}$ and the current estimates of the parameters $\beta_{0:T}^{(m)}$, i.e,

$$E_{\gamma_{0:T}, \nu_{1:T}|Y_{1:T}, \beta_{0:T}^{(m)}} \left[\log P\left(\beta_{0:T}, \gamma_{0:T}, \nu_{1:T}, \phi, \lambda | Y_{1:T}\right) \right].$$

This expectation can be obtained from the posterior defined in (4.4) by replacing each γ_{tj} with $P_{tj}^* = P^*(\beta_{tj})$ from (3.7), when $t > 0$, and $P_{0j}^* = \theta_{0j}$ from (3.6). Besides that, we exchange $1/v_t$ by the conditional expectation $E(\nu_t|\beta_{0:T}^{(m)}, Y_{1:T})$ using a discount factor δ (established in West and Harrison (2006)). We have that:

$$E(\nu_t|\beta_{0:T}^{(m)}, Y_{1:T}) = (1 - \delta)\eta_t/d_t + \delta E(\nu_{t+1}|\beta_{0:T}^{(m)}, Y_{1:T}) \text{ for } 1 \leq t < T,$$

where $E(\nu_T|\beta_{0:T}^{(m)}, Y_{1:T}) = \eta_T/d_T$, $\eta_t = \delta\eta_{t-1} + 1$ and $d_t = \delta d_{t-1} + (Y_t - F_t'\beta_t)^2$. In the M-step, we maximize the conditional expectation of the E-step with respect to $\beta_{0:T}$. We assume the discount stochastic volatility model with $\delta = 0.9$ and $\eta_0 = d_0 = 1$. In Section 5.2, we briefly discuss the sensitivity of the results for different choices of the discount factor. Additionally, we can update the autoregressive parameter ϕ and the evolution variance λ at each iteration taking the values that maximize the expected log-complete posterior. For parameter ϕ the maximum value was computed from the grid of values $\phi \in \{0.5, 0.8, 0.85, 0.90, 0.95, 0.99\}$. The choice of these values meets the expectations of a more informative prior for ϕ that ensures the stability of the estimation, avoiding abrupt changes in the estimates of $\beta_{0:T}$. The algorithm was implemented in the statistical software R (R Core Team, 2020) using C++ via the "Rcpp" (Eddelbuettel et al., 2011) and "RcppArmadillo" (Eddelbuettel and Sanderson, 2014) packages. The computation of estimation and simulation results are available in open source on Github.

5   Results

We applied the proposed methodology to a simulated dataset, a synthetic dataset that emulates fMRI signals, and a real dataset collected with fNIRS. To evaluate the efficiency of the model in recovering the true structure of the network in the simulated and synthetic datasets, we used the concepts of accuracy, sensitivity, and specificity. For this, we measured the number of true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN), where accuracy is the hit rate $((TP+TN)/(TP+TN+FP+FN))$, sensitivity is the rate of true positives $(TP/(TP+FN))$ and specificity is the true negative rate $(TN/(TN+FP))$.

Before presenting the results, we analyzed the computational time according to the number of nodes in the network and compare the performance of our proposal (DGVS) to the performance of the DGM. The clear advantage of DGVS is that it does not use the exhaustive search system to estimate the graphic structure, requiring at least 3 adjustments for each node (for the evaluation of the parameter $\Theta = \{0.9, 0.5, 0.1\}$ and considering a fixed c value). Furthermore, the selection of a set of parent nodes for each child node is done independently, allowing the models to be fitted in parallel and further reducing computational time (see Figure 2).

As already pointed out, one important demand in the literature is for improving methods that seek to optimize computational time when mapping numerous brain regions through a long time series. In that vein, DGVS can get results in 25 minutes (AMD Ryzen 5 3400G) for a 25-node network with 1000 periods of time, while the
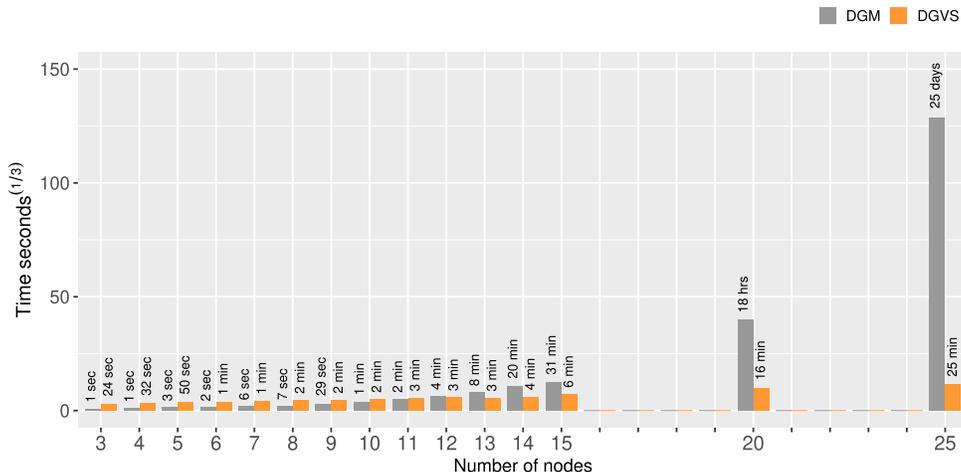
Figure 2: Average computational time for a single node considering a different number of nodes from DGM model (gray bars) and DGVS model (orange bars), for a time series of length 1000 using an AMD Ryzen 5 3400G processor.

DGM takes 25 days for the same assignment. The DGVS computational time grows slowly and linearly with the number of nodes, which is an advantage compared to the DGM whose computational time grows exponentially. As it can be seen by Figure 2, the larger the number of nodes the more computationally efficient the DGVS is compared to the DGM, with the average computational time for a single node, considering a series of length 1000, being higher under the DGM for networks with more than 11 nodes.

5.1 Simulated Data

We aim to show the performance of the DGVS to detect the true structure of brain connections analyzing the impact of sample size. To do this, we simulated three different networks of 8-nodes ($p = 8$) and $T = 220$ periods of time generated from the model (2.1) with $\phi = 0.9$, $\lambda = 1$, $c = 0.001$ and $v_t = 0.25$. The first simulation (Sim 1) has a constant network structure over time, the second (Sim 2) has a change in structure at time 120 and the third (Sim 3) has two changes in times 75 and 150. For each simulation, we generate samples of varying sizes $N = (10, 30, 50, 100, 150, 200)$ and evaluate the effect of the sample size on the proportion of active connections. We emphasize that the sample size is an important element for the identification of active connections in the network structure. For the following analysis, we are interested in the sample results that shape the joint behavior of the structures. It is important to point out that we could estimate different network structures for each brain (sample unit) even under the influence of the same stimulus. For example, two people listening to the same music and their brains responding in different ways. However, the goal here is to identify which brain connections are most commonly activated in the population under study. All figures

for the analysis of the simulated data can be found in Section 1 of the Supplementary Material (Souza et al., 2023).

We classify a connection as active at time t when $P^*(\beta_{tj}) \geq 0.5$ as suggested by Rockova and McAlinn (2021). Furthermore, we analyze the proportion of active connections at each time t and establish a graphical structure based on this proportion. Plots of the estimated time-varying structure for all simulations are in Section 2 of the Supplementary Material (Souza et al., 2023). To simplify the interpretations of the plots, we summarize these results in Figures S1(B), S2(B, D), and S3(B, D, F), which show the average of proportions over time according to the sample size. We measure accuracy, sensitivity, and specificity based on relevant connections over time. We consider that a connection is relevant at a certain time if at least 50% of the sample presented an active connection at that time. These measurements are shown in Figures S1(C), S2(E), and S3(G).

The results are satisfactory in recovering the true graphical structure, in simpler cases as in Sim 1, even for a small sample of $N = 10$ subjects. As the graphical structure becomes more complex, with an increasing number of changes over time, a larger sample size may be required. This can be explained by the stationarity property imposed on the coefficients of the dynamic model. Abrupt changes in the graphic structure are not quickly identified, with a few time steps ahead being necessary to provide good estimates. This becomes evident in the trajectory of time sensitivity in Figures S2(E) and S3(G), where there is a drop close to the times when there is a change in the graphical structure. On the other hand, we found that accuracy and specificity remained above 75% for smaller samples ($N \leq 30$) and above 85% for larger samples ($N \geq 50$) in all simulations. In addition, the sensitivity remains at 100% most of the time for almost all sample sizes.

5.2   Synthetic Data

We evaluated the efficiency of DGVS to detect true connections in synthetic fMRI datasets, that simulate BOLD signals. We chose to evaluate our model on fMRI datasets because this technique is widely disseminated and consolidated by the neuroscience community. Furthermore, the results found here can be extended to data from other brain mapping techniques capable of measuring changes in oxy and deoxyhemoglobin concentration, such as fNIRS. Synthetic data is available on the paper of Schwab et al. (2018) and are simulated following the methodology of Smith et al. (2011), using a forward model (Friston et al., 2003) and a nonlinear balloon model (Buxton et al., 1998). The data corresponds to "Sim 22" (Figure 4(A)) from Smith et al. (2011) and consists of a 5-nodes non-stationary dynamic network (N = 50 samples). In other words, the connection strengths are time-dependent and modulated by random processes. The sampling time was 10 minutes with a repetition time (TR) of 3 seconds, i.e. 600 seconds of sampling with the interval of time between consecutive observations of 3 seconds. For further details see Smith et al. (2011).

In order to assess the sensitivity of the volatility model, we verified the accuracy, sensitivity, and specificity for the estimation of synthetic data at different values for

the discount factor. Figure 3 shows the average of these measurements over time. Note that there is a smooth curvature in the three lines and when the value of δ becomes 0.90 an inflection point occurs. This indicates that by assuming this value for δ we guarantee sufficient sensitivity to identify the true connections without considerable loss of specificity.
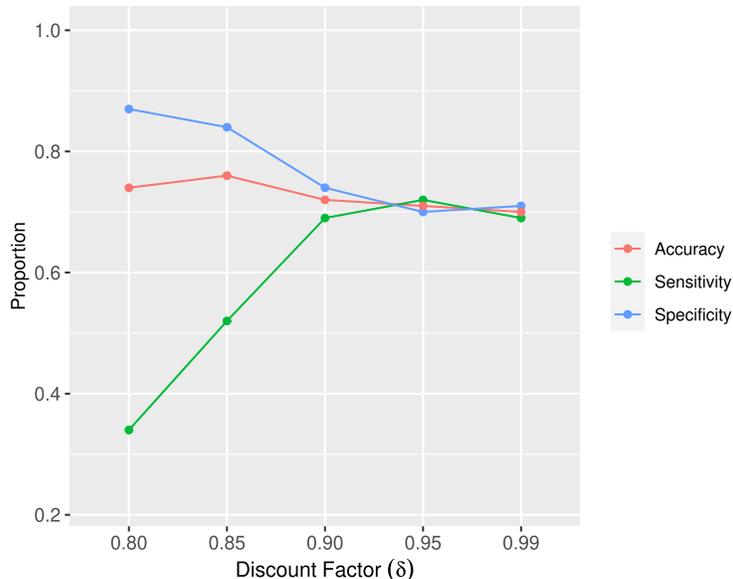

Figure 3: Average over time of accuracy, sensitivity, and specificity for Sim 22 considering different values for δ.

We compare DGVS performance for synthetic data with the results of Dynamic Graphical Models (DGM) from Schwab et al. (2018). DGM is also a dynamic graphical model in which the connectivity strengths are the regression coefficients. This method consists of applying a dynamic model to all possible sets of parent nodes for each child node. Then the structure is selected from the joint log predictive likelihood (LPL) of the best set of parent nodes for each child node. After estimating the parameters, the authors suggest an optional pruning process to avoid spurious bidirectional connections. The disadvantage of this method is the number of possible graphs that increases exponentially with the number of nodes, consequently increasing computational time. In addition, DGM does not allow a time-varying graphical structure.

Figure 4(B) illustrates the average proportion over time of estimated active connections for DGVS and the proportion of estimated active connections for DGM. The results presented for the DGM model are the proportions after the pruning process. We chose not to use this same procedure in the DGVS results because it did not lead to significant changes under this model. Plots of the estimated time-varying structure via the DGVS are provided in Section 2 of the Supplementary Material (Souza et al., 2023).

Again we consider a connection to be active when $P^*(\beta_{tj}) \geq 0.5$ for DGVS and for measuring accuracy, sensitivity, and specificity we consider a connection to be relevant when at least 50% of the sample had an active connection at each time step. DVGS has a sensitivity of 100% at 14% of the time, with a median of 80%, against 100% sensitivity of the DGM. Regarding specificity and accuracy, the DGVS has 73% and 74% (in median) against 93% and 95% of the DGM, see Figure 4(C). It is evident that one of the advantages of DGVS is the inferential estimation of the graphical structure. In contrast, resorting to exhaustive search methods, as in DGM, allows the exploration and evaluation of the entire universe of possibilities through algorithms with simpler implementations. The DGVS, by incorporating the treatment of graphical structures over time, demands a more complex estimation method that adds model calibration steps, which at the same time decrease the computational cost, and increase the estimation cost.
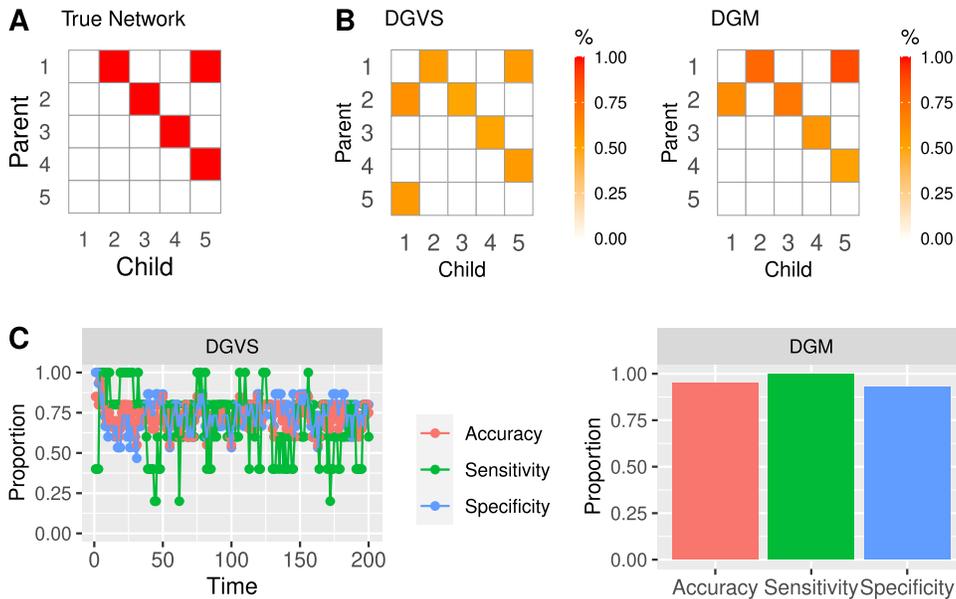

Figure 4: (A) The true adjacency matrix for Sim 22. (B) Average proportion of estimated active connections by method. (C) Accuracy, sensitivity, and specificity.

5.3   Real Data

This study concerns the analysis of network connectivities among recordings of fNIRS signals. FNIRS indirectly measures cortical activation based on fluctuations of oxygenated (oxy-Hb) and deoxygenated (deoxy-Hb) hemoglobin at various scalp locations of a human subject (León-Carrión and León-Domínguez, 2012). The most commonly used method measures changes in the ratio of oxy-Hb to blood volume. This technique is widely used to assess brain activity in subjects who are performing a task or receiving

a stimulus (Balardin et al., 2017). Barreto et al. (2020) conducted an experiment on the emotions conveyed by listening to music, in which they collected fNIRS data. Our main statistical interest here is to explore relations across fNIRS channels (regions) obtained in part of the experiment of Barreto et al. (2020) via DGVS modeling.

The fNIRS data consists of time series collected from 20 brain regions/channels in 40 participants (40% females; mean age of 25 ± 5.1 years). Eight participants were excluded from the data analysis due to the use of continuous medications that could affect hemodynamic responses or signal quality, resulting in 32 subjects. The subjects listened to instrumental musical segments of Richard Wagner's operas classified with positive and negative valences. Valence is defined as a positive/negative feature of the emotion generated by the music. We selected two of these excerpts for our analysis: (1) Siegfried's Funeral March (FM) with a duration of the 62s and negative valence; (2) The Rhine Gold Entrance of the Gods into Valhalla (RG) with a duration of the 60s and positive valence.

The data are recorded from 8 sources and 7 detectors, located around and over the participants' scalp, specifically in the prefrontal cortex due to its relationship to emotional processes (Moghimi et al., 2012). The data sampling rate was 7.81 Hz. To record the hemodynamic response, eight near-infrared LED sources were positioned at (F3, AF7, AF3, Fz, Fpz, AF4, F4, AF8) and seven light detectors were positioned at (F5, F1, Fp1, AFz, F2, Fp2, F6), following the EEG 10–10 international system (see Figure 5). It gives 20 channels positioned at the pre-frontal cortex (see Figure 5). Data pre-processing was performed to ensure signal quality and minimize the effects of systemic artifacts, such as cardiac and respiratory cycles; for full details see Barreto et al. (2020).

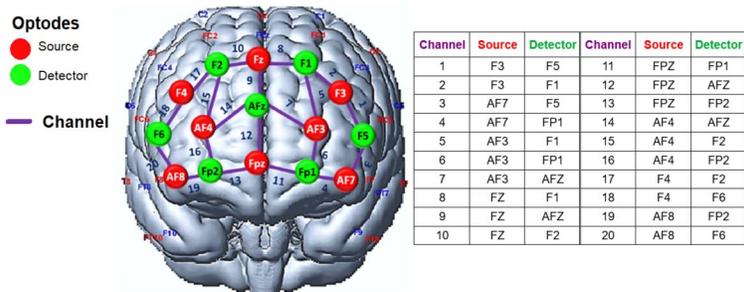| Channel	Source	Detector	Channel	Source	Detector
1	F3	F5	11	FPZ	FP1
2	F3	F1	12	FPZ	AFZ
3	AF7	F5	13	FPZ	FP2
4	AF7	FP1	14	AF4	AFZ
5	AF3	F1	15	AF4	F2
6	AF3	FP1	16	AF4	FP2
7	AF3	AFZ	17	F4	F2
8	FZ	F1	18	F4	F6
9	FZ	AFZ	19	AF8	FP2
10	FZ	F2	20	AF8	F6

Figure 5: Placement of the fNIRS optodes. Letters and numbers inside the circles represent equivalences on the EEG 10–20 system. Red and green circles represent sources and detectors, respectively. The purple lines represent the fNIRS channels. (Modified from Barreto et al. (2020).)

The analysis uses the DGVS model specified in Section 2 with the priors and inferential procedure defined in Section 4. On average for each region, the computational time for the excerpt FM took 172 seconds for $\Theta = 0.9$, 201.8 seconds for $\Theta = 0.5$, and 226 seconds for $\Theta = 0.1$. While for the RG stretch, on average the computational time

for each region took 206.3 seconds for $\Theta = 0.9$, 203.5 seconds for $\Theta = 0.5$, and 211.8 seconds for $\Theta = 0.1$. As an illustration, we show in Figure 6 the proportion of active connections estimated at three distinct periods of time under both excerpts of music. The plots for the other times are found in Section 2 of the Supplementary Material (Souza et al., 2023). As in the analyzes for simulated and synthetic data, we consider an active connection when $P^*(\beta_{tj}) \geq 0.5$ and a relevant connection if at least 50% of the sample had such an active connection. It can be noticed by Figure 6 that the active connections are more concentrated in certain regions for the excerpt with negative valence. On the other hand, the connections related to the positive excerpt presented a more spread pattern. It is evident that the model estimated a more complex graphical structure for the RG excerpt, indicating that different stimuli provoke different organizations in the effective connections of the brain.
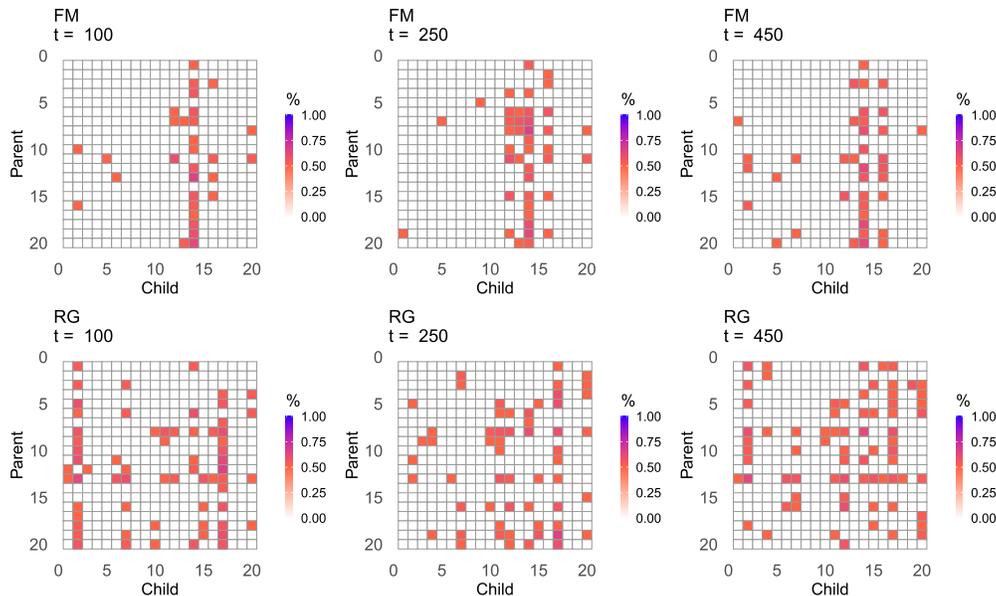

Figure 6: The proportion of active connections estimated at time 100, 250, and 450 for the excerpt Siegfried's Funeral March (FM – first line) with negative valence, and the excerpt Rhine Gold Entrance of the Gods into Valhalla (RG – second line) with positive valence.

In addition to computing the adjacency matrix of the proportions of active connections, other useful measures to obtain are the *degree in* and *degree out*. The *degree in* is the number of parents each node has, and the *degree out* is the number of children of each node. We also measure the *weighted degree in* and *out* that we consider being the mean of the absolute weights ($|\beta_{tj}|$) of the parents and children respectively of each node. Figure 7 shows the *degree in* over time, the mean and standard deviation of the *weighted degree in* for the sample of participants. In Figure 8, we present these same measures for the *degree out*.
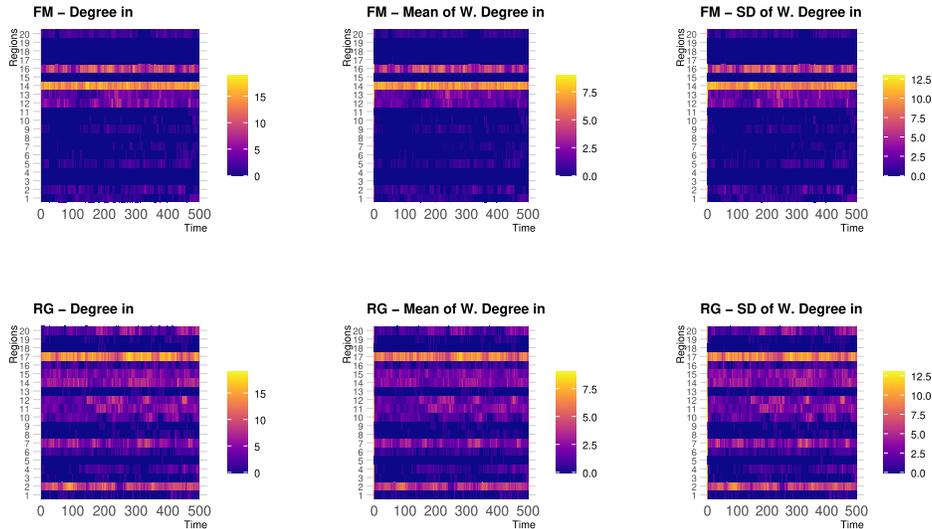
Figure 7: (Left) *Degree in* of each region over time. (Middle) Mean of *weighted degree in* among participants. (Right) The standard deviation of *weighted degree in* among participants. In the first line are the statistics for Siegfried's Funeral March (FM) with negative valence and in the second line are the Rhine Gold Entrance of the Gods into Valhalla (RG) with positive valence.

We observed that both *degree in* and *weighted degree in* are higher for regions 12, 13, 14, and 16 (especially regions 14 and 16), for the negative valence. This indicates that these regions are the most influenced by others. These results are in line with the findings of Balconi et al. (2015) that point out that negative valences induce a greater oxy-Hb increase, compared to positive stimuli, in the right hemisphere of the prefrontal cortex. Similar results can also be seen in Barreto et al. (2020). The authors show that oxy-Hb signals from regions 9, 10, and 12 may be better contributing to the emotional valence associations. Such regions with significant results are located in the right hemisphere close to the center of the brain and are relatively close to region 14 which stands out in our findings.

For the positive valence, the most influenced regions are more disseminated, highlighting regions 17 and 2 that are lateralized. On the other hand, we did not notice a specific pattern in the *degree out* for both valences, that is, it seems that most regions act as influencing regions. It is also possible to observe that influenced regions are not very influential.

The results show that the signals from the medial prefrontal regions are more related to negative valence stimuli and lateral regions are related to positive valence stimuli. The recent study of Ozawa et al. (2019) corroborates these findings. The authors argue that the anterior ventral medial prefrontal cortex (vmPFC), which is responsible for the responses to emotional stimuli, is related to exposure to negative images, while the lateral regions are responsible for the cognitive control of emotions.
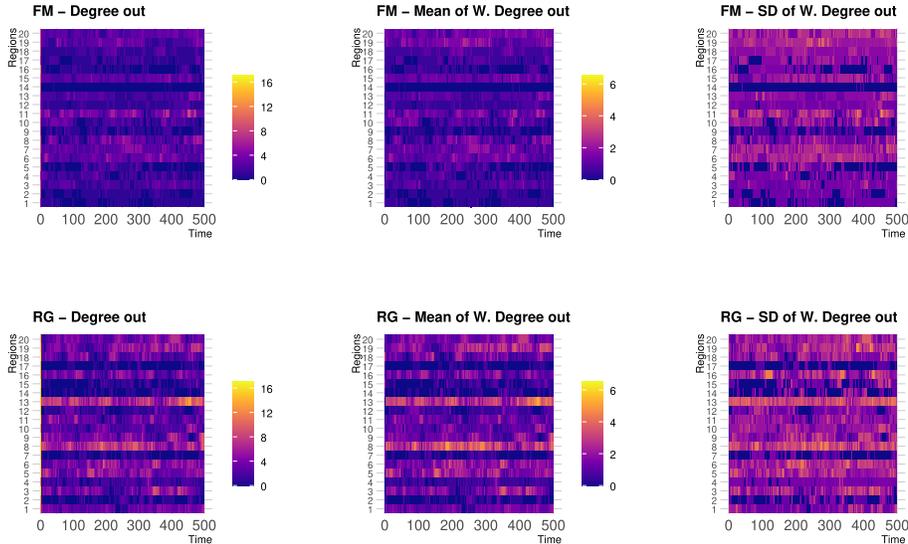
Figure 8: (Left) *Degree out* of each region over time. (Middle) Mean of *weighted degree out* among participants. (Right) The standard deviation of *weighted degree out* among participants. In the first line are the statistics for Siegfried's Funeral March (FM) with negative valence and in the second line are the Rhine Gold Entrance of the Gods into Valhalla (RG) with positive valence.

6 Discussion

In this article, we illustrate the use of shrinkage priors in studies of dynamic networks — the underlying connectivity structure emerges from time series models with dynamically evolving sparsity. The spike and slab priors class makes the graphical model very flexible and efficient and the indicator parameter enables adding sparsity in contemporaneous connections among variables, besides allowing the graphic structure to adapt to time and data. The presented approach is competitive against previous methodologies and demands a lower computational cost. The analysis based on optimization methods is essentially suitable for high-dimensional data. Furthermore, the assumption of independence of nodes given its parents allows parallel univariate analysis.

DGVS can be applied to several data sources, such as fMRI, fNIRS, electroencephalography, among others. It can also be used to estimate effective connectivity on data in a task and resting state. We emphasize that the proposal has a greater appeal in contributions for research purposes, in particular, with the objective of brain/cortical mapping related to cognitive issues. We believe that the new approach can collaborate with the interpretation of data from these techniques allowing an analysis of the temporal evolution of the connections among different brain areas. For example, in a behavioral study in which individuals are encouraged to press a button when viewing a specific image or when hearing a certain sound (Fong et al., 2019; Costa et al., 2019),

it would be interesting to compare the different structures formed with the evolution of these tasks. In the same way, the DGVS can accommodate data from different contexts, e.g. in the field of economics and financial markets.

Note that we suggested and used a sequence of values for the scalar Θ and showed that the fitted models performed well for simulated, synthetic, and real data. However, the user can benefit from the choice of other values and sequences, bearing in mind that we can interpret it as a tuning parameter. That is, values of Θ closer to 1 return denser graphs and values of Θ closer to 0 return more sparse graphs.

The study has some limitations that could be improved in future research. First, the estimation method via the EM algorithm does not allow inference of the complete posterior distribution, only point estimates are found. Strong inference techniques from the perspective of optimization can be applied, such as the Variational Bayes (VB) method which is an optimization method for the approximation of distributions. Second, it is recommended to calibrate the model, regarding the initial points of the parameters and hyperparameters of the priors, before its application to the real data. This point specifically refers to the already known particularities of the optimization methods that can be sensitive to initial points, and also the sensitivity of the prior spike and slab to changes in parameterization. This procedure, when poorly performed, can compromise the accuracy of the model, resulting in spurious connections. Lastly, no restrictions on the graphical structure are imposed. A possible extension of the DGVS models is to impose the DAG restriction at each time step. We are able to verify the presence of cycles in the network at each time step and select the best DAG from the conditional probability of classifying β_{tj}. This approach is challenging as the models associated with each node will no longer update independently. However, this enables the construction of joint data distribution which is often of great interest in multivariate data analysis. The results of these models are promising and should be explored in further studies.

Software Availability

The R code is available at https://github.com/RbeccaSouza/DGVS.

Supplementary Material

Supplementary material for paper Dynamic Graphical Models with Variable Selection for Effective Connectivity (DOI: 10.1214/23-BA1377SUPP; .pdf). Supplementary material available online includes the figures of the analysis of the simulated data and the figures of the adjacency matrices estimated over time to simulated, synthetic, and real datasets, as mentioned in Section 5.

References

Balardin, J. B., Zimeo Morais, G. A., Furucho, R. A., Trambaiolli, L., Vanzella, P., Biazoli Jr, C., and Sato, J. R. (2017). "Imaging brain function with functional near-

infrared spectroscopy in unconstrained environments." *Frontiers in human neuroscience*, 11: 258. 16

Balconi, M., Grippa, E., and Vanutelli, M. E. (2015). "What hemodynamic (fNIRS), electrophysiological (EEG) and autonomic integrated measures can tell us about emotional processing." *Brain and cognition*, 95: 67–76. 18

Barreto, C., Bruneri, G. d. A., Brockington, G., Ayaz, H., and Sato, J. R. (2021). "A New Statistical Approach for fNIRS Hyperscanning to Predict Brain Activity of Preschoolers' Using Teacher's." *Frontiers in human neuroscience*, 15: 181. 4

Barreto, C. D. S. F., Morais, G. A. Z., Vanzella, P., and Sato, J. R. (2020). "Combining the intersubject correlation analysis and the multivariate distance matrix regression to evaluate associations between fNIRS signals and behavioral data from ecological experiments." *Experimental Brain Research*, 238(10): 2399–2408. 16, 18

Bhattacharya, S. and Maitra, R. (2011). "A nonstationary nonparametric Bayesian approach to dynamically modeling effective connectivity in functional magnetic resonance imaging experiments." *The Annals of Applied Statistics*, 1183–1206. MR2849771. doi: https://doi.org/10.1214/11-AOAS470. 7

Buxton, R. B., Wong, E. C., and Frank, L. R. (1998). "Dynamics of blood flow and oxygenation changes during brain activation: the balloon model." *Magnetic resonance in medicine*, 39(6): 855–864. 13

Chabran, E., Noblet, V., de Sousa, P. L., Demuynck, C., Philippi, N., Mutter, C., Anthony, P., Martin-Hunyadi, C., Cretin, B., and Blanc, F. (2020). "Changes in gray matter volume and functional connectivity in dementia with Lewy bodies compared to Alzheimer's disease and normal aging: implications for fluctuations." *Alzheimer's research & therapy*, 12(1): 1–13. 2

Cooper, R. J., Selb, J., Gagnon, L., Phillip, D., Schytz, H. W., Iversen, H. K., Ashina, M., and Boas, D. A. (2012). "A systematic comparison of motion artifact correction techniques for functional near-infrared spectroscopy." *Frontiers in neuroscience*, 6: 147. 4

Costa, L., Smith, J., Nichols, T., Cussens, J., Duff, E. P., Makin, T. R., et al. (2015). "Searching multiregression dynamic models of resting-state fMRI networks using integer programming." *Bayesian Analysis*, 10(2): 441–478. MR3420889. doi: https://doi.org/10.1214/14-BA913. 2, 7

Costa, L., Smith, J. Q., and Nichols, T. (2019). "A group analysis using the Multiregression Dynamic Models for fMRI networked time series." *Journal of statistical planning and inference*, 198: 43–61. MR3809395. doi: https://doi.org/10.1016/j.jspi.2018.03.004. 19

Damoiseaux, J. S., Rombouts, S. A. R. B., Barkhof, F., Scheltens, P., Stam, C. J., Smith, S. M., and Beckmann, C. F. (2006). "Consistent resting-state networks across healthy subjects." *Proceedings of the National Academy of Sciences*, 103(37): 13848–13853. URL https://www.pnas.org/content/103/37/13848. 2

Diestel, R., Schrijver, A., and Seymour, P. (2010). "Graph theory." *Oberwolfach Reports*, 7(1): 521–580. MR2759675. doi: https://doi.org/10.4171/OWR/2010/11.   2

Eddelbuettel, D., François, R., Allaire, J., Ushey, K., Kou, Q., Russel, N., Chambers, J., and Bates, D. (2011). "Rcpp: Seamless R and C++ integration." *Journal of Statistical Software*, 40(8): 1–18.   11

Eddelbuettel, D. and Sanderson, C. (2014). "RcppArmadillo: Accelerating R with high-performance C++ linear algebra." *Computational Statistics & Data Analysis*, 71: 1054–1063. MR3132026. doi: https://doi.org/10.1016/j.csda.2013.02.005.   11

Ferrari, M. and Quaresima, V. (2012). "A brief review on the history of human functional near-infrared spectroscopy (fNIRS) development and fields of application." *Neuroimage*, 63(2): 921–935.   1

Filippi, M., Basaia, S., Canu, E., Imperiale, F., Magnani, G., Falautano, M., Comi, G., Falini, A., and Agosta, F. (2020). "Changes in functional and structural brain connectome along the Alzheimer's disease continuum." *Molecular psychiatry*, 25(1): 230–239.   2

Fong, A. H. C., Yoo, K., Rosenberg, M. D., Zhang, S., Li, C.-S. R., Scheinost, D., Constable, R. T., and Chun, M. M. (2019). "Dynamic functional connectivity during task performance and rest predicts individual differences in attention across studies." *NeuroImage*, 188: 14–25.   19

Friston, K. J., Harrison, L., and Penny, W. (2003). "Dynamic causal modelling." *Neuroimage*, 19(4): 1273–1302.   4, 13

Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J.-P., Frith, C. D., and Frackowiak, R. S. (1994). "Statistical parametric maps in functional imaging: a general linear approach." *Human brain mapping*, 2(4): 189–210.   4

Glover, G. H. (2011). "Overview of functional magnetic resonance imaging." *Neurosurgery Clinics*, 22(2): 133–139. MR3114749. doi: https://doi.org/10.1007/978-94-007-7302-8.   1

Granger, C. W. (1969). "Investigating causal relations by econometric models and cross-spectral methods." *Econometrica: journal of the Econometric Society*, 424–438.   2

Hu, Z., Lam, K.-F., and Yuan, Z. (2019). "Effective connectivity of the fronto-parietal network during the tangram task in a natural environment." *Neuroscience*, 422: 202–211.   2

Kim, S., Shephard, N., and Chib, S. (1998). "Stochastic volatility: likelihood inference and comparison with ARCH models." *The review of economic studies*, 65(3): 361–393.   9

Koop, G. and Korobilis, D. (2018). "Variational Bayes inference in high-dimensional time-varying parameter models." *Technical report, Munich Personal RePEc Archive*. URL https://mpra.ub.uni-muenchen.de/87972/.   9

Koop, G. and Korobilis, D. (2020). "Bayesian dynamic variable selection in high di-

mensions." *International Economic Review*. doi: https://doi.org/10.1111/iere.12623. 9

Kowal, D. R., Matteson, D. S., and Ruppert, D. (2019). "Dynamic shrinkage processes." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 81(4): 781–804. MR3997101. doi: https://doi.org/10.1111/rssb.12325. 5

Kundu, S., Lukemire, J., Wang, Y., and Guo, Y. (2019). "A novel joint brain network analysis using longitudinal Alzheimer's disease data." *Scientific reports*, 9(1): 1–18. 9

León-Carrión, J. and León-Domínguez, U. (2012). "Functional near-infrared spectroscopy (fNIRS): principles and neuroscientific applications." *Neuroimaging methods*, 48–74. 15

Li, B., Daunizeau, J., Stephan, K. E., Penny, W., Hu, D., and Friston, K. (2011). "Generalised filtering and stochastic DCM for fMRI." *Neuroimage*, 58(2): 442–457. MR2660062. doi: https://doi.org/10.1155/2010/621670. 4

Li, J., Wang, Z. J., Palmer, S. J., and McKeown, M. J. (2008). "Dynamic Bayesian network modeling of fMRI: a comparison of group-analysis methods." *Neuroimage*, 41(2): 398–407. 4

Liu, T., Duan, L., Dai, R., Pelowski, M., and Zhu, C. (2021). "Team-work, Team-brain: Exploring synchrony and team interdependence in a nine-person drumming task via multiparticipant hyperscanning and inter-brain network topology with fNIRS." *NeuroImage*, 237: 118147. 2

Lopes, H. F., McCulloch, R. E., and Tsay, R. S. (2022). "Parsimony inducing priors for large scale state–space models." *Journal of Econometrics*, 230(1): 39–61. MR4436489. doi: https://doi.org/10.1016/j.jeconom.2021.11.005. 9

Moghimi, S., Kushki, A., Guerguerian, A. M., and Chau, T. (2012). "Characterizing emotional response to music in the prefrontal cortex using near infrared spectroscopy." *Neuroscience Letters*, 525(1): 7–11. 16

Molinari, M., Cremaschi, A., De Iorio, M., Chaturvedi, N., Hughes, A., and Tillin, T. (2022). "Bayesian dynamic network modelling: an application to metabolic associations in cardiovascular diseases." *Journal of Applied Statistics*, 1–25. 5

Mumford, J. A. and Ramsey, J. D. (2014). "Bayesian networks for fMRI: a primer." *Neuroimage*, 86: 573–582. 4

Nakajima, J. and West, M. (2013). "Bayesian analysis of latent threshold dynamic models." *Journal of Business & Economic Statistics*, 31(2): 151–164. MR3055329. doi: https://doi.org/10.1080/07350015.2012.747847. 9

Nakajima, J. and West, M. (2015). "Dynamic network signal processing using latent threshold models." *Digital Signal Processing*, 47: 5–16. MR3425313. doi: https://doi.org/10.1016/j.dsp.2015.04.008. 2, 7, 9

Narisetty, N. N., He, X., et al. (2014). "Bayesian variable selection with shrinking and

diffusing priors." *Annals of Statistics*, 42(2): 789–817. MR3210987. doi: https://doi.org/10.1214/14-AOS1207. 9

Nguyen, T., Hoehl, S., and Vrtička, P. (2021). "A guide to parent-child fNIRS hyperscanning data processing and analysis." *Sensors*, 21(12): 4075. 4

Okamoto, M., Dan, H., Shimizu, K., Takeo, K., Amita, T., Oda, I., Konishi, I., Sakamoto, K., Isobe, S., Suzuki, T., et al. (2004). "Multimodal assessment of cortical activation during apple peeling by NIRS and fMRI." *Neuroimage*, 21(4): 1275–1288. 4

Ozawa, S., Kanayama, N., and Hiraki, K. (2019). "Emotion-related cerebral blood flow changes in the ventral medial prefrontal cortex: an NIRS study." *Brain and cognition*, 134: 21–28. 18

Patel, S., Katura, T., Maki, A., and Tachtsidis, I. (2011). "Quantification of systemic interference in optical topography data during frontal lobe and motor cortex activation: an independent component analysis." In *Oxygen Transport to Tissue XXXII*, 45–51. Springer. 4

Phillips, P. C. (1991). "To criticize the critics: An objective Bayesian analysis of stochastic trends." *Journal of Applied Econometrics*, 6(4): 333–364. 9

Piironen, J. and Vehtari, A. (2017). "Sparsity information and regularization in the horseshoe and other shrinkage priors." *Electronic Journal of Statistics*, 11(2): 5018–5051. MR3738204. doi: https://doi.org/10.1214/17-EJS1337SI. 5

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/. 11

Ray, S., Miller, M., Karalunas, S., Robertson, C., Grayson, D. S., Cary, R. P., Hawkey, E., Painter, J. G., Kriz, D., Fombonne, E., et al. (2014). "Structural and functional connectivity of the human brain in autism spectrum disorders and attention-deficit/hyperactivity disorder: A rich club-organization study." *Human brain mapping*, 35(12): 6032–6048. 2

Ročková, V. and George, E. I. (2014). "EMVS: The EM approach to Bayesian variable selection." *Journal of the American Statistical Association*, 109(506): 828–846. MR3223753. doi: https://doi.org/10.1080/01621459.2013.869223. 9

Rockova, V. and McAlinn, K. (2021). "Dynamic variable selection with spike-and-slab process priors." *Bayesian Analysis*, 16(1): 233–269. MR4194280. doi: https://doi.org/10.1214/20-BA1199. 5, 6, 7, 9, 10, 13

Rogers, B. P., Katwal, S. B., Morgan, C. L., V. L.and Asplund, and Gore, J. C. (2010). "Functional MRI and multivariate autoregressive models." *Magnetic resonance imaging*, 28(8): 1058–1065. 2

Rykhlevskaia, E., Gratton, G., and Fabiani, M. (2008). "Combining structural and functional neuroimaging data for studying brain connectivity: a review." *Psychophysiology*, 45(2): 173–187. 1

Schwab, S., Harbord, R., Zerbi, V., Elliott, L., Afyouni, S., Smith, J. Q., Woolrich, M. W., Smith, S. M., and Nichols, T. E. (2018). "Directed functional connectivity using dynamic graphical models." *NeuroImage*, 175: 340–353.   2, 3, 4, 7, 13, 14

Smith, S. M., Miller, K. L., Salimi-Khorshidi, G., Webster, M., Beckmann, C. F., Nichols, T. E., Ramsey, J. D., and Woolrich, M. W. (2011). "Network modelling methods for FMRI." *Neuroimage*, 54(2): 875–891.   2, 13

Souza, R., Costa, L., Paez, M., Sato, J., and Barreto, C. (2023). "Supplement to Dynamic Graphical Models with Variable Selection for Effective Connectivity." *Bayesian Analysis*. doi: https://doi.org/10.1214/23-BA1377SUPP.   13, 14, 17

Tak, S. and Ye, J. C. (2014). "Statistical analysis of fNIRS data: a comprehensive review." *Neuroimage*, 85: 72–91.   4

Vergotte, G., Torre, K., Chirumamilla, V. C., Anwar, A. R., Groppa, S., Perrey, S., and Muthuraman, M. (2017). "Dynamics of the human brain network revealed by time-frequency effective connectivity in fNIRS." *Biomedical optics express*, 8(11): 5326–5341.   2

West, M. and Harrison, J. (2006). *Bayesian forecasting and dynamic models*. Springer Science & Business Media. MR1020301. doi: https://doi.org/10.1007/978-1-4757-9365-9.   4, 5, 11

Wilcox, T., Bortfeld, H., Woods, R., Wruck, E., and Boas, D. A. (2008). "Hemodynamic response to featural changes in the occipital and inferior temporal cortex in infants: a preliminary methodological exploration." *Developmental science*, 11(3): 361–370.   4

Yanagi, M., Hosomi, F., Kawakubo, Y., Tsuchiya, A., Ozaki, S., and Shirakawa, O. (2020). "A decrease in spontaneous activity in medial prefrontal cortex is associated with sustained hallucinations in chronic schizophrenia: An NIRS study." *Scientific Reports*, 10(1): 1–7.   2