Taylor & Francis
Taylor & Francis Group

Journal of Computational and Graphical Statistics

ISSN: 1061-8600 (Print) 1537-2715 (Online) Journal homepage: www.tandfonline.com/journals/ucgs20

Discrete Autoregressive Switching Processes with Cumulative Shrinkage Priors for Graphical Modeling of Time Series Data

Beniamino Hadj-Amar, Aaron M. Bornstein, Michele Guindani & Marina Vannucci

To cite this article: Beniamino Hadj-Amar, Aaron M. Bornstein, Michele Guindani & Marina Vannucci (15 Oct 2025): Discrete Autoregressive Switching Processes with Cumulative Shrinkage Priors for Graphical Modeling of Time Series Data, Journal of Computational and Graphical Statistics, DOI: 10.1080/10618600.2025.2551271

To link to this article: https://doi.org/10.1080/10618600.2025.2551271

JOURNAL OF COMPUTATIONAL AND GRAPHICAL STATISTICS
2025, VOL. 00, NO. 0, 1–15
https://doi.org/10.1080/10618600.2025.2551271

Discrete Autoregressive Switching Processes with Cumulative Shrinkage Priors for Graphical Modeling of Time Series Data

Beniamino Hadj-Amar[a], Aaron M. Bornstein[b], Michele Guindani[c], and Marina Vannucci[d]

[a]Department of Epidemiology and Biostatistics, University of South Carolina, Columbia, SC; [b]Department of Cognitive Sciences, University of California, Irvine, CA; [c]Department of Biostatistics, UCLA Fielding School of Public Health, Los Angeles,CA; [d]Department of Statistics, Rice University, Houston, TX

ABSTRACT

We propose a flexible Bayesian approach for sparse Gaussian graphical modeling of multivariate time series. We account for temporal correlation in the data by assuming that observations are characterized by an underlying and unobserved hidden discrete autoregressive process. We assume multivariate Gaussian emission distributions and capture spatial dependencies by modeling the state-specific precision matrices via graphical horseshoe priors. We characterize the mixing probabilities of the hidden process via a cumulative shrinkage prior that accommodates zero-inflated parameters for non-active components, and further incorporate a sparsity-inducing Dirichlet prior to estimate the effective number of states from the data. For posterior inference, we develop a sampling procedure that allows estimation of the number of discrete autoregressive lags and the number of states, and that cleverly avoids having to deal with the changing dimensions of the parameter space. We thoroughly investigate performance of our proposed methodology through several simulation studies. We further illustrate the use of our approach for the estimation of dynamic brain connectivity based on fMRI data collected on a subject performing a task-based experiment on latent learning. Supplementary materials for this article are available online.

ARTICLE HISTORY
Received May 2024
Accepted August 2025

KEYWORDS
Brain connectivity; Cumulative shrinkage prior; Discrete autoregressive process; fMRI data; Graphical models; Horseshoe prior

1. Introduction

In this article we consider the problem of estimating sparse Gaussian graphical models based on time series data. Time-changing dependencies and sparse structures are often encountered when investigating multi-dimensional physiological signals (Safikhani and Shojaie 2022), environmental and sensor data (Lam and Yao 2012), as well as macroeconomic and financial systems (Kastner and Huber 2020). Among existing approaches, Song, Kolar, and Xing (2009) introduced a time-varying dynamic Bayesian network for modeling the fluctuating network structures underlying nonstationary biological time series. Kolar et al. (2010) proposed a method for estimating time-varying networks based on temporally smoothed l_1-regularized logistic regression. Danaher, Wang, and Witten (2014) and Qiu et al. (2016) addressed the challenge of estimating multiple related Gaussian graphical models when observations belong to distinct classes, and Warnick et al. (2018) and Quinn et al. (2018) employed Hidden Markov Models (HMMs) for the estimation of recurrent brain connectivity networks during a neuroimaging experiment. Other procedures for modeling the temporal evolution of dynamic networks include change-point detection methods (Cribben, Wager, and Lindquist 2013; Xu and Lindquist 2015) and time-varying parameter models (Lindquist et al. 2014; Zhang et al. 2021). Change-point techniques provide a data-driven approach for the temporal partitioning of the network structure into segments of adaptable length. However, these methods do not provide a system for identifying potentially recurring network patterns over time. Time-varying parametric methods offer a principled way of modeling dynamic correlations but are computationally intensive.

We propose a flexible Bayesian approach for sparse Gaussian graphical modeling of multivariate time series. In order to represent switching dynamics, we assume an unobserved hidden process, underlying the time series data, which at each time point exists in one of a finite number of states. We account for the temporal structure of this hidden process by assuming a Discrete Autoregressive (DAR) process of order P (Biswas and Song 2009), which flexibly incorporates long-term dependencies by considering the P previous lags of the process. Given the state of the latent process, we model the observations as conditionally independent of the observations and states at previous times and generated from state-specific multivariate Gaussian emission distributions. Under the multivariate Gaussian assumption, networks can be estimated by the graphical models induced by the state-specific inverse covariance matrices. We capture these spatial dependencies by modeling the state-specific precision matrices via graphical horseshoe priors.

The DAR hidden process construction we adopt is reminiscent of higher-order HMMs, where the present state depends not only on the immediately preceding state but also on prior states

CONTACT Beniamino Hadj-Amar *HADJAMAR@mailbox.sc.edu* Department of Epidemiology and Biostatistics, University of South Carolina, 915 Greene Street, Columbia, SC 29201.

Supplementary materials for this article are available online. Please go to *www.tandfonline.com/r/JCGS*.

further back in time. First-order HMMs, which constrain the temporal dynamics of the hidden state sequence to be Markovian, have been successfully applied in many scientific fields, including neuroimaging (Warnick et al. 2018; Quinn et al. 2018), climate (Holsclaw et al. 2017) and animal behavior (DeRuiter et al. 2017), to cite a few. While higher-order HMMs have been suggested (Cappé, Moulines, and Rydén 2005), they require the estimation of transition probability matrices that grow exponentially in size as the order increases, making their estimation challenging (see, for a discussion, Sarkar and Dunson 2019). In our proposed model, the state-switching behavior of the process is captured by the time-varying mixing probabilities of the DAR process. To model these probabilities, we propose a nonparametric zero-inducing cumulative shrinkage prior. Building upon the construction of the finite Dirichlet process (DP; see Ishwaran and James 2001), the proposed prior accommodates zero-inflated parameters, to account for non-active components, and employs cumulative shrinkage (Legramanti, Durante, and Dunson 2020) to handle increasing model complexity. This construction ensures that if a parameter in the DAR model is zero, then all subsequent lag parameters are also zero. This results in a flexible and computationally efficient framework for learning the time-varying mixing probabilities and the effective order of the process, as opposed to learning the entire transition matrix, as required in HMM modeling. Such reduction in the number of parameters leads to a substantial computational advantage. It also allows to learn the number of lags in a data-driven fashion. Related sparsity-inducing prior constructions have been developed by Heiner, Kottas, and Munch (2019) for the simplex model, and by Tang and Chen (2019) for zero-inflated generalized Dirichlet multinomial regression models. These constructions are specific to those models and less flexible than our approach, which models the ordering of the lags as the process evolves in time while promoting lower-order complexity. We complete our modeling framework with a sparsity-inducing Dirichlet prior that allows estimation of the effective number of hidden states in a data-driven manner. Drawing inspiration from the literature on overfitted finite mixture models (Rousseau and Mengersen 2011; Malsiner-Walli, Frühwirth-Schnatter, and Grün 2016), we consider more states than strictly necessary, while employing a prior that effectively constrains the model's complexity. This promotes sparsity while leading to more interpretable inferences.

For posterior inference, we take a fully Bayesian approach and develop a sampling procedure that accommodates the multiple model selection problems, namely the number of DAR lags and the number of states, while cleverly avoiding having to deal with the changing dimensions of the parameter space. Specifically, we implement a Gibbs sampler that alternates between updating the DAR parameters, the sparse emission parameters, and the latent state sequence. To update the DAR probabilities, we leverage the stick-breaking construction of the DP by augmenting the space with auxiliary indicator variables and design a joint sampling scheme that alternates between adding or removing the sticks of the zero-inducing DP formulation. Our zero-inducing cumulative shrinkage prior significantly accelerates the proposed sampler, particularly in regard to the forward-backward algorithm for updating the latent state sequence. Estimates of the number of hidden states and DAR order are determined based on the most frequently occurring number of active states and DAR order observed during MCMC sampling, respectively.

We thoroughly investigate performance of our proposed methodology through several simulation studies. We further illustrate the use of our proposed approach to estimate dynamic brain connectivity networks based on functional Magnetic Resonance Imaging (fMRI) data. Identifying the dynamic nature of brain connectivity is critical for understanding our current knowledge about human brain functioning. In our application, we consider data collected on a subject performing an experiment aimed at understanding neural representations that are formed during latent learning. Inferred networks by our method identify distinct regimes of functional connectivity, that can be mapped onto cognitive interpretation.

The rest of the article is organized as follows. Section 2 introduces the proposed model, including the DAR process and the proposed prior structures, and the MCMC algorithm for posterior inference. Section 3 contains results from the simulation studies and Section 4 illustrates the application to fMRI data on latent learning. Section 5 provides concluding remarks. The Julia software `sggmDAR`, which implements our proposed methodology is available on GitHub at *https://github.com/Beniamino92/sggmDAR*.

2. Sparse Modeling of Multivariate Time Series Data via Cumulative Shrinkage DAR

In this section, we describe the proposed latent variable approach for modeling sparse multi-dimensional time series. Let $\boldsymbol{y} = \{\boldsymbol{y_t}\}_{t=1}^{T}, \boldsymbol{y}_t = (y_{t1}, \ldots, y_{tD}) \in \mathbb{R}^D$, be the observed D-dimensional time series data, with T indicating the number of time points. We envision an unobserved, latent hidden process underlying the observations and assume that, at each time point, the process assumes one of a finite number of states, represented as $\boldsymbol{\gamma} = \{\gamma_t\}_{t=1}^{T}$, with $\gamma_t \in \{1, \ldots, M\}$ and M denoting the (unknown) finite number of latent states. Given the value of γ_t, the observations $\boldsymbol{y}_t$ are assumed to be independent of both the observations and states at previous time points. We further assume that the state-specific emissions follow a D-variate Gaussian distribution

$$\boldsymbol{y}_t \mid \gamma_t, \boldsymbol{\mu}, \boldsymbol{\Omega} \sim \sum_{j=1}^{M} \mathbb{1}_{\{j\}}(\gamma_t)\, \mathcal{N}_D(\boldsymbol{y}_t \mid \boldsymbol{\mu}_j, \boldsymbol{\Omega}_j^{-1}), \tag{1}$$

with state-specific means $\boldsymbol{\mu}_j$ and precision matrices $\boldsymbol{\Omega}_j$, $j = 1, \ldots, M$, $t = 1, \ldots, T$. Here, $\mathbb{1}_{\{j\}}(\gamma_t)$ denotes the indicator function, which equals 1 if $\gamma_t = j$, and 0 otherwise. Conditional dependencies can be inferred from the off-diagonal entries of the precision matrices. Specifically, for a given state j, if the entry $\omega_{j,il}$ is zero, the corresponding variables y_{ti} and y_{tl} are conditionally independent given the other variables.

2.1. State Dynamics via Discrete Autoregressive Processes

In order to learn the dependence structure between time points, represented by the sequence $\boldsymbol{\gamma}$, we design an approach that employs a discrete autoregressive process, with a cumulative shrinkage prior that enables a computationally efficient estimation of the order of the process. More specifically, we assume

that the evolution of the hidden state sequence γ_t follows a Discrete Autoregressive (DAR) process of order P (Biswas and Song 2009), a framework originally introduced in the context of multi-lag finite-state Markov chains by Pegram (1980), and subsequently adapted to categorical time series modeling by Biswas and Song (2009). The DAR process allows the hidden sequence to incorporate long-term dependencies by considering the previous P lags. Formally, the conditional distribution of γ_t given the past values $\gamma_{t-1:t-P}$ is expressed as

$$p(\gamma_t|\gamma_{t-1:t-P}, \boldsymbol{\phi}, \boldsymbol{\pi}) = \phi_1 \mathbb{1}_{\{\gamma_{t-1}\}}(\gamma_t) + \phi_2 \mathbb{1}_{\{\gamma_{t-2}\}}(\gamma_t) + \cdots + \phi_P \mathbb{1}_{\{\gamma_{t-P}\}}(\gamma_t) + \phi_0 \pi_{\gamma_t}, \quad (2)$$

where $\boldsymbol{\phi} = (\phi_0, \ldots, \phi_P)$ and $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_M)$. We denote with $\{\phi_j\}_{j=0}^P$ the *autoregressive probabilities*, with $\phi_0 = 1 - \sum_{j=1}^P \phi_j$, while the *state innovation probabilities* $\{\pi_i\}_{i=1}^M$ are defined as $\pi_i := p(\gamma_t = i)$, for $i = 1, \ldots, M$, and allow the process to transition to any of the M states, including those not observed in the previous P lags. Here, $\mathbb{1}_{\{j\}}(i)$ is an indicator function equal to 1 if $i = j$ and 0 otherwise. According to (2), at each time point t, the model selects a latent state $\gamma_t \in \{1, \ldots, M\}$, and assigns probabilities based on how this state relates to the past P values of the sequence. Specifically, the probability of choosing a particular state $i \in \{1, \ldots, M\}$ is given by the sum of two components: an autoregressive component and an innovation component. The autoregressive probabilities $\{\phi_j\}_{j=1}^P$ govern the recurrence structure, so that if $\gamma_t = \gamma_{t-j}$ for any $j \in \{1, \ldots, P\}$, then state i receives mass ϕ_j. That is, the model assigns probability ϕ_j to each state i that matches γ_{t-j}. In this way, the latent sequence favors persistence or recurrence of recently visited states. Complementing this, the innovation probability ϕ_0 allows for exploration beyond the observed history. With probability ϕ_0, the model selects a state independently of $\gamma_{t-1}, \ldots, \gamma_{t-P}$, according to the innovation distribution $\boldsymbol{\pi}$, which assigns nonzero probability to each of the M possible states.

The transition probabilities in the DAR process can be represented by a multi-dimensional array. The dimensions of this array are determined by the number of autoregressive lags, P, and the number of hidden states, M. As an illustration, when $P = 2$, the transition probabilities are described by an $[M \times M \times M]$ array, say $\boldsymbol{\eta}$. The individual components of this array, denoted as $\eta_{l,i,j}$, represent the probability $p(\gamma_t = j|\gamma_{t-1} = i, \gamma_{t-2} = l)$ for $l, i, j \in \{1, \ldots, M\}$, as defined in (2). However, as the number of lags P increases, the dimensionality of this array grows exponentially. Therefore, the DAR characterization simplifies inference by allowing us to focus only on making inferences on the ϕ parameters, as opposed to learning the entire transition matrix, which is the case with HMM models, for example. In fact, when dealing with higher order HMMs, the task involves estimating M^P parameters for the transition arrays. In contrast, our proposed method streamlines this process by estimating $(M + P)$ parameters, resulting in a substantial computational advantage.

2.1.1. Zero-Inducing Cumulative Shrinkage Prior for Learning Time Dependence

The time-varying mixing probabilities of the DAR model, denoted as ϕ_j, characterize the state-switching behavior of the process. To model these probabilities, we propose a nonparametric zero-inducing cumulative shrinkage prior that accommodates zero-inflated parameters to account for non-active components, and that employs cumulative shrinkage (Legramanti, Durante, and Dunson 2020) to handle increasing model complexity. This prior modifies the stick-breaking construction to allow for an increasing probability of setting $\phi_j = 0$ as j increases. In addition, our formulation enforces that once ϕ_j becomes zero for a specific j, $j = 1, 2, \ldots, P$, subsequent lags obey the condition $p(\phi_{j+k} = 0 \,|\, \phi_j = 0) = 1$, $k = 1, \ldots, P - j$. To formally introduce our prior, we need to define a binary latent process, namely an "active order" latent indicator, denoted as $z_j \in \{0, 1\}$, $j = 1, \ldots, P$. If $z_j = 0$, then ϕ_j is almost surely nonzero. However, when the first j such that $z_j = 1$ occurs, then $\phi_j = 1 - \sum_{l=1}^{j-1} \phi_l$ and $z_l = 1$ almost surely for $l = j + 1, \ldots, P$. More formally, the mixing probabilities ϕ_j are generated via a modified stick-breaking construction,

$$\phi_j = v_j \prod_{l=0}^{j-1} (1 - v_l), \quad \text{for } j = 1, \ldots, P, \quad (3)$$

with $\phi_0 = v_0$, where the stick-breaking weights v_j are mixtures of a Beta distribution and a spike at one,

$$v_j \mid z_j \sim (1 - z_j)\ \text{Beta}\,(a_v, b_v) + z_j\, \delta_1, \quad (4)$$

with δ_x denoting a point mass at $\{x\}$, $j = 1, \ldots, P$, and by specifying $v_0 \sim \text{Beta}(a_0, b_0)$.

For $z_j = 0$, (3)–(4) define the stick-breaking construction typical of the Dirichlet process. If at some point $z_j = 1$ occurs, then $v_j = 1$, and $\phi_j = \prod_{l=0}^{j-1}(1 - v_l) = 1 - \sum_{l=1}^{j-1} \phi_l$. For all remaining lags, our construction ensures $\phi_l = 0$, $l = j+1, \ldots, P$. More specifically, to enforce the desired behavior and promote lower order model complexity, we leverage the increasing shrinkage prior construction of Legramanti, Durante, and Dunson (2020) and assign increasing probability mass to selecting the spike component as the order of the DAR grows. In particular, we assume $z_j \mid \boldsymbol{v}_{0:j-1} \sim \text{Bern}\,(\xi_j)$ with probability $\xi_j = \sum_{i=0}^{j-1} \phi_i$ increasing with the lag j, where $z_1|v_0 \sim \text{Bern}(v_0)$. See also Zhang et al. (2021), where an increasing shrinkage prior is used in a VAR model. Our construction ensures that $p\,(z_l = 1|z_{l-1} = 1) = 1$ and $p\,(\phi_l = 0 \,|\, \phi_{l-1} = 0) = 1$. We define the effective order of the DAR process as the random element $\hat{P} = \inf_{j \in \{1,\ldots,P\}}\{z_j = 1\}$, that is the number of "active" lags of the DAR process. The proposition below demonstrates the aforementioned property.

Proposition 1. Let $\boldsymbol{\phi} = \{\phi_j \in \Delta_\phi : j = 0, \ldots, P\}$ with $\Delta_\phi = \{\phi_l : 0 \le \phi_l \le 1, \sum_{l=0}^{\infty} \phi_l = 1\}$, be constructed according to (3), and $\boldsymbol{v} = \{v_i\}_{i=0}^P$ and $\boldsymbol{z} = \{z_i\}_{i=1}^P$ be specified as in (4). Under these assumptions, the cumulative shrinkage DAR formulation implies that $p(z_{j+1} = 1|z_j = 1) = 1$, for $j = 1, \ldots, P$.

Proof. Recall that $z_j \mid \boldsymbol{v}_{0:j-1} \sim \text{Bern}\,(\xi_j)$ with probability $\xi_j = \sum_{i=0}^{j-1} \phi_i, j = 1, \ldots P$. Therefore, for $j = 1, \ldots, \hat{P}$, we can write

$$p(z_{j+1}=1|z_j=0,\boldsymbol{v}_{0:j}) = \sum_{i=0}^{j}\phi_i = v_0 + v_1(1-v_0) + \ldots + v_j\prod_{l=0}^{j-1}(1-v_l).$$

Thus, $p(z_{j+1}=0|z_j=0,\boldsymbol{v}_{0:j}) = 1 - p(z_{j+1}=1|z_j=0,\boldsymbol{v}_{0:j}) = \prod_{l=0}^{j}(1-v_l)$. For $j=\hat{P}$, since $v_{\hat{P}}=1$ a.s., we have $\sum_{j=0}^{\hat{P}}\phi_j = v_0 + v_1(1-v_0) + \cdots + v_{\hat{P}}\prod_{l=0}^{\hat{P}-1}(1-v_l) = 1$. Thus, for $j=\hat{P}+1,\ldots,P-1$, $p(z_{j+1}=1|v_{0:\hat{P}}) = p(z_{j+1}|z_j=1) = 1$. □

Given the one-to-one relationship between the sequence z_j and $\hat{P}$, the process can be alternatively defined in terms of the random quantity $\hat{P}$, which is computationally convenient, as we explain in Section 2.3. We note that the previous characterization can also be extended to the case of $P=\infty$. However, for computational purposes, it is convenient to consider only a finite number of terms, say, $P_{\max}$, and thus specify the autoregressive coefficients as $\boldsymbol{\phi} = (\phi_0,\ldots,\phi_{P_{\max}})$, where $\phi_{P_{\max}} = 1 - \sum_{l=0}^{P_{\max}-1}\phi_j$. In implementations, this approach offers considerable versatility when $P_{\max}$ is set to a moderately high upper bound, and it is advisable to choose $P_{\max}$ such that it exceeds the expected number of lags.

2.1.2. Sparsity-Inducing Dirichlet Prior to Infer State Transitions and Space Size

As for the innovation probabilities $\boldsymbol{\pi}$, to facilitate a substantial reduction in the effective number of states compared to the maximum number, $M = M_{\max}$, we draw insights from recent literature on overfitted finite mixture models (Rousseau and Mengersen 2011; Malsiner-Walli, Frühwirth-Schnatter, and Grün 2016). Specifically, we assume a symmetric Dirichlet prior $\boldsymbol{\pi} = (\pi_1,\ldots,\pi_M) \sim \text{Dir}(\kappa_0,\ldots,\kappa_0)$, where the concentration parameter κ_0 is set at a very small value, so that the marginal densities of each π_j are spiked around the values zero and one, $j=1,\ldots,M$. This approach results in estimating a reduced number of hidden states, denoted as $\hat{M}$, which is significantly less than M. Thus, unnecessary hidden states are effectively removed from the posterior distribution. The hyperparameter κ_0 plays a crucial role. Here, we set $\kappa_0 = 0.001$ following the recommendation by Malsiner-Walli, Frühwirth-Schnatter, and Grün (2016). In Section 2.4, we propose to estimate the number of hidden states based on the most frequent number of active states during MCMC sampling. By setting a large value for M, our approach provides a simple and automated framework for estimating the number of hidden states, without relying on computations of marginal likelihoods, post-MCMC model selection criteria, or reversible-jump MCMC.

2.2. Graphical Horseshoe Priors for the Precision Matrices

To induce prior sparsity in the state-specific precision matrices $\boldsymbol{\Omega}_j$'s, we employ the graphical horseshoe (GHS) prior proposed by Li, Craig, and Bhadra (2019). This prior uses normal scale mixtures with half-Cauchy hyperpriors for the off-diagonal entries of the precision matrix while using uninformative priors for its diagonal elements. Specifically,

$$\begin{aligned}\omega_{j,ii} &\propto 1,\\ \omega_{j,il:i<l} &\sim \mathcal{N}(0,\lambda_{j,il}^2\tau_j^2),\\ \lambda_{j,il:i<l} &\sim C^+(0,1),\\ \tau_j &\sim C^+(0,1),\end{aligned}$$

for $i,l=1,\ldots,D$, and $j=1,\ldots,M$. The global shrinkage parameter τ_j plays a crucial role in promoting sparsity across the entire matrix $\boldsymbol{\Omega}_j$, by shrinking the estimates of all the off-diagonal values towards zero. On the other hand, the local shrinkage parameters $\lambda_{jil:i<l}$ allow to preserve the magnitudes of the nonzero off-diagonal elements, ensuring that the element-wise biases do not become too large. This combination of global and local shrinkage enables the GHS prior to induce sparsity in the precision matrices while capturing the relevant dependencies between the elements.

We complete the prior specification on the emission distributions by assuming Gaussian priors on the state-specific means, that is, $p(\boldsymbol{\mu}_j) \sim \mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{R}_0^{-1})$, for $j=1,\ldots,M$.

2.3. Markov Chain Monte Carlo Algorithm

We now outline the MCMC algorithm we designed for posterior inference. For notational convenience, we collect all parameters except $\boldsymbol{\gamma}$ as the set $\boldsymbol{\theta} = \{\boldsymbol{v},\boldsymbol{z},\boldsymbol{\pi},\boldsymbol{\mu},\boldsymbol{\Omega},\boldsymbol{\tau},\boldsymbol{\Lambda}\}$ with $\boldsymbol{\Lambda} = \{\boldsymbol{\Lambda}_j\}_{j=1}^M$ and $\boldsymbol{\Lambda}_j = \{\lambda_{j,il}^2\}$ the matrices of local shrinkage parameters in the GHS prior, $\boldsymbol{\tau} = (\tau_1,\ldots,\tau_M)$ the global parameters, $\boldsymbol{\mu} = (\boldsymbol{\mu}_1,\ldots,\boldsymbol{\mu}_M)$, and $\boldsymbol{\Omega} = (\boldsymbol{\Omega}_1,\ldots,\boldsymbol{\Omega}_M)$. We then write the posterior distribution of $\boldsymbol{\theta}$ conditional upon the current value of $\boldsymbol{\gamma}$ as

$$p(\boldsymbol{\theta}\,|\,\boldsymbol{y},\boldsymbol{\gamma}) \propto \mathcal{L}(\boldsymbol{\theta};\boldsymbol{y},\boldsymbol{\gamma})\,p(\boldsymbol{v},\boldsymbol{z})\,p(\boldsymbol{\pi})\,p(\boldsymbol{\mu})\,p(\boldsymbol{\Omega},\boldsymbol{\tau},\boldsymbol{\Lambda}), \quad (5)$$

where the conditional likelihood is factorized as

$$\mathcal{L}(\boldsymbol{\theta};\boldsymbol{y},\boldsymbol{\gamma}) = \prod_{t=P+1}^{T} p(\gamma_t\,|\,\gamma_{t-1:t-P},\boldsymbol{v},\boldsymbol{z},\boldsymbol{\pi})\,p(\boldsymbol{y}_t\,|\,\gamma_t,\boldsymbol{\mu},\boldsymbol{\Omega}) \quad (6)$$

and where the joint prior $p(\boldsymbol{v},\boldsymbol{z})$ of the indicator variables and the stick-breaking weights can be expressed as

$$p(\boldsymbol{v},\boldsymbol{z}) = p(v_0)\prod_{j=1}^{\hat{P}-1}p(v_j|z_j)\prod_{j=0}^{\hat{P}}p(z_{j+1}|\boldsymbol{v}_{0:j}), \quad (7)$$

with

$$\begin{aligned}p(v_j|z_j) &\propto \text{Beta}(a_v,b_v)^{1-z_j},\\ p(z_{j+1}|\boldsymbol{v}_{0:j}) &\propto \text{Bern}(1)^{z_j}\,\text{Bern}(\xi_j)^{1-z_j},\end{aligned} \quad (8)$$

and the conditioning on $\boldsymbol{v}_{0:j-1}$ induced by the cumulative shrinkage parameter ξ_j.

Since the posterior distribution is not available in closed form, we develop a Gibbs sampler that alternates between: (i) drawing the stick-breaking weights $\boldsymbol{v}$ and auxiliary indicators $\boldsymbol{z}$. For this, we design a Metropolis-Hastings algorithm similar to Savitsky, Vannucci, and Sha (2011), that cleverly avoids having to deal with the changing dimensions of the parameter space via a joint update of the indicators and the weights; (ii) updating the innovation probabilities $\boldsymbol{\pi}$ related to the sparsity-inducing Dirichlet prior; (iii) sampling the multivariate sparse emission

parameters, that is the mean vectors in $\boldsymbol{\mu}$, the precision matrices in $\boldsymbol{\Omega}$ and the global and local shrinkage parameters $\boldsymbol{\tau}$ and $\boldsymbol{\Lambda}$; (iv) updating the latent state sequence $\boldsymbol{\gamma}$, through a forward-backward algorithm, which is significantly accelerated by the proposed zero-inducing cumulative shrinkage prior formulation. We now describe these updates in full detail.

- **Update $\boldsymbol{z}$ and $\boldsymbol{v}$**: We perform a joint update of the indicators $\boldsymbol{z}$ and weights $\boldsymbol{v}$ by designing a Metropolis-Hastings sampler with *birth* and *death* moves, that increase or decrease the order of the DAR process by one. Formally, let us define the current number of active components $\hat{P}^{curr}$, stick-breaking weights $\boldsymbol{v}^{curr} = (v_0, v_1, \ldots, v_{\hat{P}^{curr}-1}, 1)$, and indicator variables $\boldsymbol{z}^{curr} = (0, 0, \ldots, 0, 1)$, of dimensions $\hat{P}^{curr} + 1$ and $\hat{P}^{curr}$, respectively; note that $\boldsymbol{z}^{curr} = 1$ when $\hat{P}^{curr} = 1$. A new vector of indicators $\boldsymbol{z}$ is drawn by proposing at random one of the following two moves:

  (i) *birth move:* Set $\hat{P}^{prop} = \hat{P}^{curr} + 1$ and construct $\boldsymbol{z}^{prop}$ from $\boldsymbol{z}^{curr}$ by adding a zero entry; for this move, the proposed vector of weights is constructed as $\boldsymbol{v}^{prop} = (v_0, v_1, \ldots, v_{\hat{P}^{curr-1}}, v_{\hat{P}^{prop-1}}, 1)$ with $v_{\hat{P}^{prop}-1}$ drawn from the prior, that is $v_{\hat{P}^{prop}-1} \sim \text{Beta}(a_v, b_v)$, and $v_{\hat{P}^{prop}}$ set equal to one. This move is accepted or rejected with probability

$$\alpha = \min\left\{1, \frac{p(\boldsymbol{v}^{prop}, \boldsymbol{z}^{prop} \,|\boldsymbol{\gamma}, \boldsymbol{y}, \cdot)}{p(\boldsymbol{v}^{curr}, \boldsymbol{z}^{curr} \,|\boldsymbol{\gamma}, \boldsymbol{y}, \cdot)} \frac{1}{\text{Beta}(v_{\hat{P}^{prop}-1}|a_v, b_v)}\right\}, \tag{9}$$

  where the joint posterior distribution $p(\boldsymbol{z}, \boldsymbol{v}|\cdot)$ is easily available by appropriate conditioning of the relevant variables in (5) and (6), that is

$$p(\boldsymbol{v}, \boldsymbol{z} \,|\boldsymbol{\gamma}, \boldsymbol{y}, \cdot) \propto p(\boldsymbol{v}, \boldsymbol{z}) \prod_{t=P+1}^{T} p(\gamma_t \mid \gamma_{t-1:t-P}, \boldsymbol{v}, \boldsymbol{z}, \boldsymbol{\pi}), \tag{10}$$

  with the DAR probabilities $p(\gamma_t|\cdot)$ defined as in (2), recalling that $\boldsymbol{\phi}$ is a by-product of $\boldsymbol{v}$ and $\boldsymbol{z}$ using the formulation presented in (4).

  (ii) *death move:* Set $\hat{P}^{prop} = \hat{P}^{curr} - 1$ and construct $\boldsymbol{z}^{prop}$ from $\boldsymbol{z}^{curr}$ by removing a zero entry; here, $\boldsymbol{v}^{prop}$ is obtained from $\boldsymbol{v}^{curr}$ by replacing the component $v_{\hat{P}^{curr}-1}$ with a one and setting $v_{\hat{P}^{curr}}$ equal to zero, namely $\boldsymbol{v}^{prop} = (v_0, v_1, \ldots, v_{\hat{P}^{curr}-2}, 1)$. This is move is accepted or rejected with probability the inverse of (9) with the appropriate change of labeling.

After each death/birth move, to enhance the mixing efficiency of the MCMC algorithm, we further update each component of the weight vector $\boldsymbol{v}$ using a one-at-a-time slice sampler (Neal 2003). Slice sampling is particularly advantageous for drawing samples from one-dimensional conditional distributions within a Gibbs sampling framework. Here, we focus on multivariate targets by iteratively sampling each variable. In particular, we obtain posterior samples from the target function $p(v_j \,|\, \boldsymbol{v}_{-j}, \cdot)$, for $j = 0, \ldots, \hat{P} - 1$, where $\boldsymbol{v}_{-j} = (v_0, \ldots, v_{j-1}, v_{j+1}, \ldots, v_{\hat{P}-1})$.

We remark here that the order P of the DAR process is not modeled as a random variable, but rather inferred directly from $\boldsymbol{z}$ and $\boldsymbol{v}$, eliminating the necessity of employing a trans-dimensional MCMC sampler (Green 1995).

- **Update $\boldsymbol{\pi}$:** We update the components of $\boldsymbol{\pi}$ with a one-at-a-time slice sampler, drawing samples from the target function $p(\pi_l \,|\, \boldsymbol{\pi}_{-l}, \cdot)$, for $j = 0, \ldots, M_{\max} - 1$, where $\boldsymbol{\pi}_{-l} = (\pi_0, \ldots, \pi_{l-1}, \pi_{l+1}, \ldots, \pi_{M_{\max}})$. Note that $\pi_{M_{\max}}$ is automatically obtained from its simplex, that is $\pi_{M_{\max}} = 1 - \sum_{l=0}^{M_{\max}-1} \pi_l$.
- **Update $\boldsymbol{\Omega}_j$, $\boldsymbol{\Lambda}$, and $\boldsymbol{\tau}$**: We use the augmented block Gibbs sampler method proposed by Li, Craig, and Bhadra (2019). We center the observations belonging to each state to its current value of the emission mean, $\boldsymbol{\mu}_j$, and consider a modified set of observations denoted as $\tilde{\boldsymbol{Y}}_j = \{\boldsymbol{y}_t - \boldsymbol{\mu}_j : \gamma_t = j\}$. By doing this, we can closely follow the scheme proposed by Li, Craig, and Bhadra (2019), which assumes zero-mean multivariate normal distributions. We apply the Gibbs sampler independently for each state j, from$j = 1$ to $M_{\max}$ and subsequently update the global shrinkage parameter τ_j and its corresponding augmented parameter ξ_j. We refer to the reader to Algorithm 1 of Li, Craig, and Bhadra (2019), for the details of the GHS sampler.
- **Update $\boldsymbol{\mu}_j$**: We sample the mean vectors $\boldsymbol{\mu}_j$ from the corresponding full conditional, as is typical in the context of Gaussian Bayesian regression settings (see e.g., Gelman et al. 1995). The posterior distribution is given by $\boldsymbol{\mu}_j|_{\boldsymbol{\Omega}_j, \boldsymbol{y}, \cdot} \sim \mathcal{N}(\boldsymbol{\mu}_j^{\star}, \boldsymbol{\Omega}_j^{\star})$, where

$$\boldsymbol{\Omega}_j^{\star\,-1} = \boldsymbol{R_0} + N_j \boldsymbol{\Omega}_j, \quad \text{and} \quad \boldsymbol{\mu}_j^{\star} = \boldsymbol{\Omega}_j^{\star}(\boldsymbol{R_0}\boldsymbol{\mu}_0 + N_j \boldsymbol{\Omega}_j \boldsymbol{Y}_j), \tag{11}$$

  and $\boldsymbol{Y}_j$ denotes the $(N_j \times D)$-dimensional matrix of observations assigned to state j, with N_j the corresponding number of observations belonging to that regime.
- **Update $\boldsymbol{\gamma}$**: We update the sequence of latent states $\boldsymbol{\gamma}$ with a block-wise approach that adapts the forward-backward procedure employed by Fox et al. (2011) and Hadj-Amar et al. (2021) to take into account temporal dynamics that extend beyond a simple Markovian structure. Conditional upon $\boldsymbol{\phi}$, $\boldsymbol{\pi}$, $\boldsymbol{\mu}$, and $\boldsymbol{\Omega}$, we harness the dependence structure of the DAR and develop an iterative sampling scheme based on the following representation of the posterior distribution of the hidden states

$$\begin{aligned} p(\boldsymbol{\gamma} \,|\boldsymbol{y}, \cdot) = p(\gamma_1|\boldsymbol{y}, \cdot)\, p(\gamma_2|\gamma_1, \boldsymbol{y}, \cdot) \ldots p(\gamma_{\hat{P}}|\gamma_{1:\hat{P}-1}, \boldsymbol{y}, \cdot) \\ \times \prod_{t=\hat{P}+1}^{T} p(\gamma_t|\gamma_{t-1:t-\hat{P}}, \boldsymbol{y}, \cdot). \end{aligned} \tag{12}$$

Under this factorization, we first sample $\gamma_1 \sim p(\gamma_1|\boldsymbol{y}, \cdot)$, then, conditioning on the value of γ_1, we draw $\gamma_2 \sim p(\gamma_2|\gamma_1, \boldsymbol{y}, \cdot)$, and so on, where we update $\gamma_t \sim p(\gamma_t|\gamma_{t-1:t-\hat{P}}, \boldsymbol{y}, \cdot)$, given the previous sampled states $\gamma_{t-1:t-\hat{P}}$. Assuming $M = M_{\max}$, the general form for the conditional posterior distribution of the states in (12) is given by

$$\begin{aligned} p(\gamma_t = j_0|\gamma_{t-1} = j_1, \ldots, \gamma_{t-\hat{P}} = j_{\hat{P}}, \boldsymbol{y}, \cdot) \\ \propto \eta_{\{j_{\hat{P}}, \ldots, j_1, j_0\}}\, p(\boldsymbol{y}_t|\gamma_t = j_0, \boldsymbol{\mu}, \boldsymbol{\Omega})\, \beta_{t+1}(j_0), \end{aligned} \tag{13}$$

for $t = \hat{P} + 1, \ldots, T$, and $j_l \in \{1, \ldots, M\}$, $l = 1, \ldots \hat{P}$, where $\eta_{\{j_{\hat{P}}, \ldots, j_1, j_0\}}$ are the DAR probabilities of selecting state

j_0, given previous values $j_1, \ldots, j_{\hat{P}}$, as defined in (2), and $p(\boldsymbol{y}_t \,|\cdot)$ are the multivariate spiked Gaussian emission densities specified in (1). Here, we define the *backward messages* $\beta_t(j_1) = p(\boldsymbol{y}_{t:T}|\gamma_{t-1} = j_1, \cdot)$, as the probability of the partial observation sequence from time t to T given the state j_1 at time $t-1$, conditioned on all the other parameters. These messages can be recursively expressed as follows (see Proposition 2, supplementary material)

$$\beta_t(j_1) = \underbrace{\sum_{j_{\hat{P}}=1}^{M} \cdots \sum_{j_2=1}^{M} \sum_{j_0=1}^{M}}_{\hat{P} \text{ times}} \eta_{\{j_{\hat{P}}, \ldots, j_2, j_1, j_0\}} \times p(\boldsymbol{y}_t|\gamma_t = j_0, \boldsymbol{\mu}, \boldsymbol{\Omega}) \beta_{t+1}(j_0), \quad t \leq T, \tag{14}$$

with $\beta_{T+1}(\cdot) = 1$. Our zero-inducing formulation for the DAR probabilities, described in Section 2.1.1, allows a significant speed-up of the proposed sampler, since in (14) we restrict summations to the active DAR terms only, rather than using the entire multi-dimensional array $\boldsymbol{\eta}$. Additionally, we specify the initial DAR probabilities $\eta_{\{\cdot\}}$ in (12) and (13) to be uniformly distributed.

Following similar practices as in Fox et al. (2011) and Hadj-Amar et al. (2021), we update only emission parameters for those states that have at least 1% of the assignments, while for those states that do not satisfy this condition we draw the corresponding emission parameters from their priors. For the GHS prior, we draw the diagonal entries of the precision matrix using a diffuse prior $\omega_{ii} \sim U(0, 100)$.

We acknowledge that the proposed Bayesian procedure may be susceptible to the *label switching* problem (Jasra, Holmes, and Stephens 2005) due to the invariance of the likelihood (6) under permutations of the mixture components' labeling. To mitigate this issue, we adopt a post-processing approach using the Equivalence Classes Representatives (ECR) algorithm, initially introduced by Papastamoulis and Iliopoulos (2013) and later improved by Rodríguez and Walker (2014). The core idea of the ECR algorithm is to categorize analogous allocation vectors as mutually exclusive solutions to the label switching problem. In this context, two allocation vectors are considered analogous if one can be obtained from the other merely by permuting its labels. The ECR procedure divides the allocation vectors into analogous categories and identifies a representative from each category. Consequently, during post-processing, the ECR algorithm identifies the permutation corresponding to each MCMC iteration. This permutation is then applied to reorder the matching allocation with the aim of aligning it with the representative of its category.

2.4. Posterior Inference

After obtaining the (possibly relabeled) MCMC output, we first estimate the number of active DAR components by computing the posterior probabilities $p(\hat{P} = p \,|\cdot)$, $p = 1, \ldots, P_{\max}$ and then identify the posterior mode as the value of $\hat{P}$ that maximizes such posterior probabilities. Similarly, to estimate the number of hidden states, we first calculate the posterior probabilities $p(M = m \,|\cdot)$ for $m = 1, \ldots, M_{\max}$ as

$$P(M = m|\cdot) = \frac{1}{S} \sum_{s=1}^{S} \mathbb{1}(\hat{M}^{(s)} = m), \qquad \text{where}$$
$$\hat{M}^{(s)} = \sum_{j=1}^{M_{\max}} \mathbb{1}\big(N_j^{(s)} \neq 0\big), \tag{15}$$

with N_j the number of observations assigned to state j, and where the superscript (s) indicates the MCMC iteration for $s = 1, \ldots, S$. We then calculate the posterior mode to obtain the final estimate of the number of hidden states, $\hat{M}$. Next, conditional upon these estimates, we perform posterior inference on the model parameters $\hat{\boldsymbol{\phi}}$, $\hat{\boldsymbol{\pi}}$, $\hat{\boldsymbol{\mu}}$, and $\hat{\boldsymbol{\Omega}}$ by averaging their sampled values across the MCMC iterations with number of hidden states $\hat{M}$ and DAR order $\hat{P}$.

As for inference on the sequence of latent states, we perform both global and local decoding. *Global decoding* refers to the determination of the most likely sequence of the entire vector of latent states $\hat{\boldsymbol{\gamma}} = (\hat{\gamma}_1, \ldots, \hat{\gamma}_T)$. We obtain such a maximum a posteriori (MAP) estimate by using a variant of the scheme described in (12). Given the estimated parameters $\hat{\boldsymbol{\phi}}$, $\hat{\boldsymbol{\pi}}$, $\hat{\boldsymbol{\mu}}$, and $\hat{\boldsymbol{\Omega}}$, we iteratively maximize the posterior distribution of the states, where at each time step t, we compute $\hat{\gamma}_t = \arg\max_{j=1,\ldots,\hat{M}} p(\gamma_t = j|\hat{\gamma}_{t-1:t-\hat{P}}, \boldsymbol{y}, \cdot)$. In contrast, *local decoding* of the hidden state at time t, $p(\gamma_t = j \,|\boldsymbol{y}, \cdot)$ refers to the determination of that state which is most likely at that time. This is achieved using

$$p(\gamma_t = j \,|\boldsymbol{y}, \cdot) \propto p(\gamma_t = j, \boldsymbol{y}_{1:t}| \cdot) p(\boldsymbol{y}_{t+1:T}| \gamma_t = j, \cdot) = \alpha_{t+1}(j) \beta_{t+1}(j),$$

where the backward messages are defined as $\beta_t(j) = p(\boldsymbol{y}_{t:T}|\gamma_{t-1} = j, \cdot)$ and the forward messages are expressed as $\alpha_t(j) = p(\boldsymbol{y}_{1:t-1}, \gamma_{t-1} = j \,|\cdot)$. In order to leverage the recursive nature of these messages, we also defined the DAR-forward messages $\alpha_t(j_1, \ldots, j_{\hat{P}}) = p(\boldsymbol{y}_{1:t-1}, \gamma_{t-1} = j_1, \ldots, \gamma_{t-\hat{P}} = j_{\hat{P}} \,|\cdot)$. Further details and the validity of these expressions are provided in the supplementary material.

For inference on the graphs, since the GHS approach is a shrinkage procedure, and thus it does not estimate the entries as exact zeros, we use posterior credible intervals to perform variable selection, as suggested by Li, Craig, and Bhadra (2019). Specifically, we use a 95% interval from the estimated posterior distribution of each off-diagonal element of the precision matrices, so that if the interval corresponding to an entry includes zero, that entry is assessed as non active. Note that Li, Craig, and Bhadra (2019) employed a 50% symmetric credible interval, arguing that such a procedure would have conservative properties, and would reduce false negatives while controlling for false positive. However, in our experiments, a 95% interval seemed to outperform the choice suggested by Li, Craig, and Bhadra (2019).

3. Simulation Studies

We investigate the performance of our proposed methodology using simulated data, aiming to assess its ability to recover the true DAR probabilities, emission parameters, and the correct number of autoregressive lags and hidden states. We further evaluate the robustness of the approach under model misspecification and its effectiveness on synthetic fMRI data.

3.1. Data Generation

We first consider a simulation framework where the data were generated with underlying time-varying means and structured precision matrices. We generated 30 distinct datasets from model (1)–(2), each consisting of $D = 15$-dimensional time series of length $T = 2000$, and assumed $M = 5$ latent states and a DAR of order $P = 2$. The autoregressive probabilities were set to $\boldsymbol{\phi} = (0.1, 0.75, 0.15)$ and the innovations to $\boldsymbol{\pi} = (0.6, 0.1, 0.1, 0.1, 0.1)$. For each state j, the emission means $\boldsymbol{\mu}_j$ were independently simulated from a multivariate Gaussian distribution with mean vector $\boldsymbol{b}_0 = (-\frac{5}{D}, \ldots, -\frac{1}{D}, 0, \frac{1}{D}, \ldots, \frac{5}{D})$ and identity matrix as the covariance matrix, that is $\boldsymbol{\mu}_j \sim \mathcal{N}(\boldsymbol{b}_0, \boldsymbol{I}_D)$, and where the simulated components of these vectors were randomly shuffled. The state-specific precision matrices $\boldsymbol{\Omega}_j$ were assumed to be sparse with diagonal elements fixed to one and off diagonal elements constructed using the following five structures:

(i) *Identity graph*: this structure assumes that the components are independent, that is the off-diagonal elements are all set to zero.
(ii) *Star graph*: a configuration similar to the identity matrix, except for the first row and first column, whose elements are set to $\omega_{il} = -\frac{1}{D}$ if $i = 1$ or $l = 1$, and 0 otherwise.
(iii) *Hub graph*: this structure is organized into five blocks (hubs) of the same size. For any $l \neq i$ in the same block as i we specify $\omega_{il} = \omega_{li} = -\frac{2}{\sqrt{D}}$, and 0 otherwise.
(iv) *AR(2) graph*: in this structure the precision matrix displays an autoregressive pattern of order two over the main diagonal. The entries are specified as $\omega_{il} = \frac{1}{2}$ if $l = i-1, i+1$, $\omega_{il} = \frac{1}{4}$ if $l = i-2, i+2$, and 0 otherwise.
(v) *Random sparse graph*: for this setting, the precision matrix is generated by randomly selecting $\lfloor \frac{3}{2}D \rfloor$ off-diagonal entries, and drawing each ω_{jl} uniformly from the interval $[-1.0, -0.4] \cup [0.4, 1.0]$, while the diagonal elements are fixed at 1, and the other entries at 0. Each of the off-diagonal element is then divided by the sum of the off-diagonal elements in its row, and then the matrix is averaged with its transpose, to produce a symmetric, positive definite, matrix.

Partial correlation matrices corresponding to these five scenario are displayed in Figure 1 (top row). A single realization from the simulation setting described in this section is shown in Figure 2 (top panel), with vertical bands representing the true underlying state sequence. Here, we have further scaled the time series realization, independently for each dimension $d = 1, \ldots, D$, in such a way that the corresponding standard deviation of those observations $\{y_{td}\}_{t=1}^{T}$ is equal to one. We note that the partial correlations are invariant under a change of scale and origin, allowing a meaningful comparison between true and estimated values of these matrices.

3.2. Parameter Settings

Results reported below were obtained by fixing the maximum number of states to $M_{\max} = 10$ and the maximum DAR order to $P_{\max} = 5$. The DAR hyperparameters were chosen as $a_0 = 1, b_0 = 10$, and $a_v = 10, b_v = 1$, so that the prior probabilities of innovation and autoregression were driven toward zero and one, respectively. The hyperparameters for the emission vector mean were specified as $\boldsymbol{\mu}_0 = \mathbf{0}$ and $\boldsymbol{R}_0 = (1/10)\,\boldsymbol{I}_D$, so that the mean components were a priori independent across different dimensions and with fairly large variance, hence, reflecting weakly informative beliefs.

Initial values of the MCMC sampler were chosen as follows: the DAR parameters were sampled from the prior; the Gaussian emission means were fixed to the centers of a k-mean clustering and the covariance matrices were set to the identity. The GHS parameters were set to one. MCMC chains were run for 4000 iterations, with 1200 iterations discarded as burn-in. The algorithm took approximately 10 min to run, for each simulated time series, with a program written in Julia 1.6 on an Intel® Core™ i5 2GHz Processor 16GB RAM. We verified convergence of the MCMC sampler by: (i) analyzing the trace plots of the parameters, for example the mean of the multivariate spiked Gaussian emissions, observing no pathological behavior; (ii) storing the values of the overall likelihood (6) and plotting the corresponding trace, noting that it reached a stable regime; (iii)

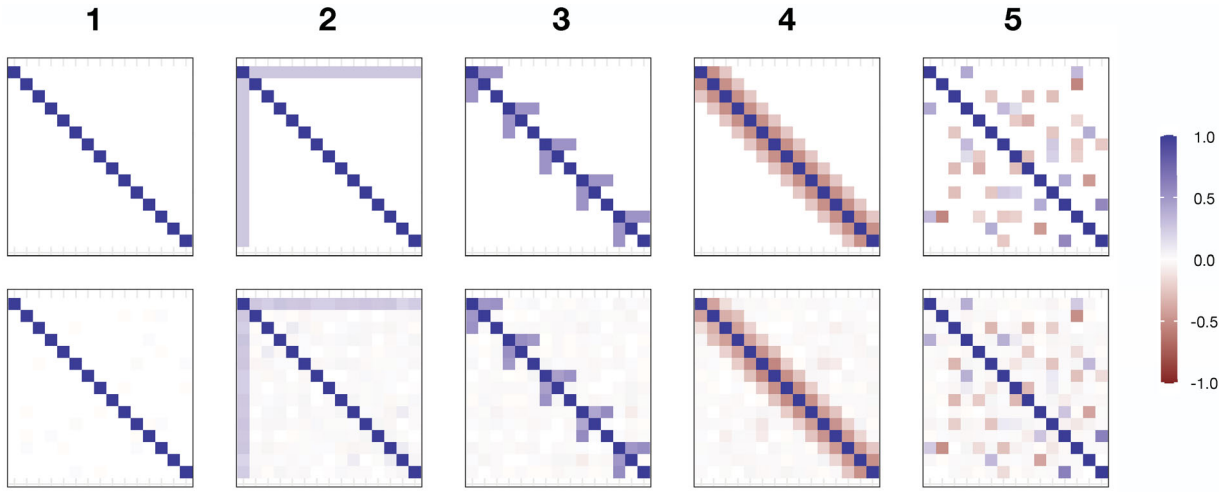

Figure 1. Simulation study. (top) true state-specific partial correlation matrices; (bottom) estimated state-specific partial correlation matrices. These results are conditioned upon the estimated modal number of states and autoregressive order.

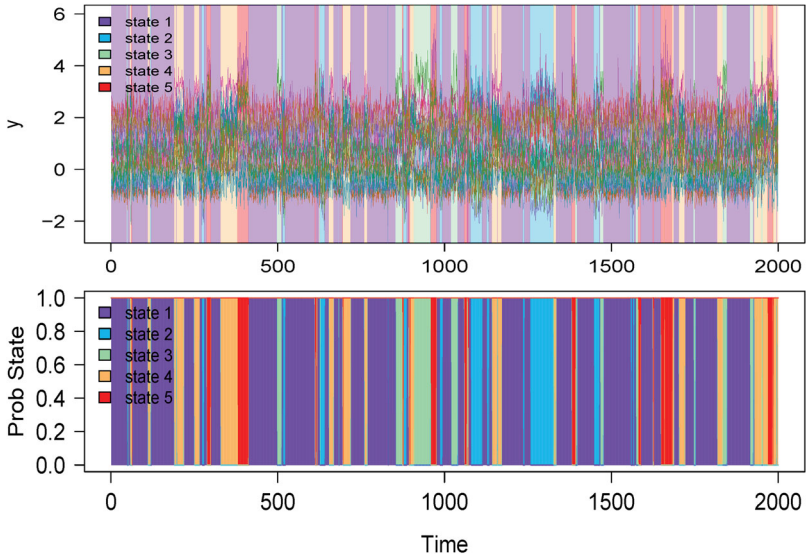

Figure 2. Simulation study. (top) time series realization (lines), with each dimension represented by a different line; vertical bands represent the true underlying state sequence; (bottom) estimated time-varying probability plot.

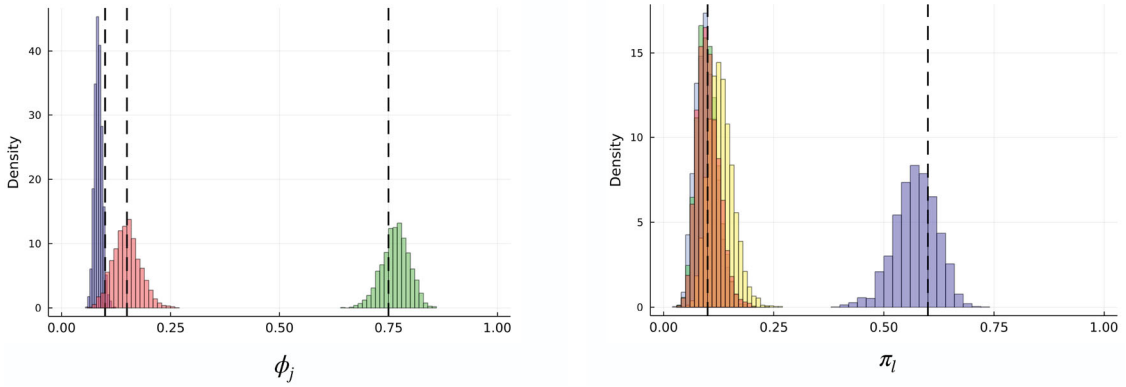

Figure 3. Simulation study. Posterior histograms of the DAR parameters. (left) autoregressive probabilities $\phi_l, l = 0, \ldots \hat{P}$; (right) innovation probabilities $\pi_j, j = 1, \ldots, \hat{M}$. Dotted vertical lines denote true parameters. These results are conditioned upon the modal number of states and autoregressive order.

verifying the Heidelberger and Welch's convergence diagnostic (Heidelberger and Welch 1981) for the likelihood trace. We report some of the results in the supplementary material.

3.3. Results

Our approach consistently estimated the correct number of states $\hat{M} = 5$ as the mode of the posterior distribution and the number of active DAR probabilities as $\hat{P} = 2$ with high posterior probability, on all simulated replicates. For a single replicate, in Figure 1 (bottom row) we show the estimates of the state-specific partial correlation matrices, conditioned upon the modal number of states and modal number of DAR parameters. Our approach successfully retrieves the distinct patterns of the true graphs. Figure 2 (bottom panel) displays a time-varying probability plot, namely the local decoding of the hidden state at time t, $p(\gamma_t = j \,|\, \boldsymbol{y}, \cdot)$, $j = 1, \ldots, \hat{M}$, as described in Section 2.4; these plots are constructed by plotting the local probabilities (which add to 1) cumulatively for each t. The proposed approach appears to correctly retrieve the true latent state sequence. Additionally, Figure 3 displays the posterior histograms of autoregressive and innovation probabilities, with dotted vertical lines denoting the true generating values. Our proposed method appears to provide a good match between true and estimated values for the DAR parameters. In the supplementary material (Section C.1), we further analyze state-specific mean and variance values across different dimensions, revealing distinct patterns in both location and dispersion across states. These findings provide additional support for the model's ability to capture meaningful heterogeneity in the underlying data distribution.

Next, we investigated the performance of our proposed approach over the 30 replicated datasets and performed comparisons with alternative methods. We focused on the recovery of the state-specific precision matrices and compare the proposed methodology, which will be referred to as `sggmDAR`, to two alternative approaches. For the first approach, which we call `mvHMM`, we fit a Bayesian multivariate HMM with Gaussian emissions, with a Normal inverse-Wishart prior on the state-specific emission parameters $(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j) \sim NIW(\boldsymbol{\mu}_0, \boldsymbol{S}_0/\kappa_0; \nu_0, \boldsymbol{S}_0)$, where the hyper-parameters were specified in a weakly informative manner, that is $\boldsymbol{\mu}_0 = \boldsymbol{0}, \kappa_0 = 0.1, \nu_0 = D + 2, \boldsymbol{S}_0 = \boldsymbol{I}_D$. The number of states was set to five (i.e., the truth). The transition probabilities were assumed

Table 1. Simulation study.

		Identity	Star	Hub	AR(2)	Random
Acc	sggmDAR	1.0 (0.0)	0.995 (0.007)	1.0 (0.002)	0.969 (0.021)	0.958 (0.027)
	mvHMM	0.969 (0.032)	0.933(0.039)	0.933 (0.088)	0.822 (0.138)	0.883 (0.110)
	glassoSlide	0.993 (0.018)	0.835 (0.047)	0.805 (0.053)	0.690 (0.031)	0.714 (0.031)
Spec	sggmDAR	1.0 (0.0)	0.995 (0.007)	1.0 (0.0)	0.995 (0.007)	0.997 (0.005)
	mvHMM	0.969 (0.032)	0.942 (0.022)	0.933 (0.093)	0.840 (0.158)	0.924 (0.100)
	glassoSlide	0.993 (0.018)	0.937 (0.052)	0.879 (0.065)	0.899 (0.040)	0.869 (0.044)
MCC	sggmDAR	–	0.979 (0.028)	0.998 (0.011)	0.919 (0.055)	0.868 (0.087)
	mvHMM	–	0.734 (0.212)	0.765 (0.210)	0.591 (0.297)	0.630 (0.365)
	glassoSlide	–	0.140 (0.198)	−0.019 (0.040)	−0.017 (0.101)	−0.004 (0.073)
F1	sggmDAR	–	0.981 (0.025)	0.998 (0.011)	0.936 (0.045)	0.883 (0.083)
	mvHMM	–	0.761 (0.194)	0.747 (0.243)	0.683 (0.254)	0.683 (0.332)
	glassoSlide	–	0.193 (0.189)	0.081 (0.051)	0.125 (0.078)	0.152 (0.062)
Sens	sggmDAR	–	0.994 (0.020)	0.996 (0.020)	0.895 (0.071)	0.807 (0.124)
	mvHMM	–	0.874 (0.235)	0.927 (0.245)	0.772 (0.267)	0.732 (0.356)
	glassoSlide	–	0.171 (0.174)	0.103 (0.072)	0.089 (0.063)	0.126 (0.059)
RMSE	sggmDAR	0.002 (0.001)	0.034 (0.008)	0.021 (0.007)	0.049 (0.011)	0.042 (0.011)
	mvHMM	0.019 (0.012)	0.073 (0.013)	0.080 (0.035)	0.116 (0.057)	0.087 (0.041)
	glassoSlide	0.001 (0.004)	0.097 (0.003)	0.162 (0.002)	0.205 (0.002)	0.155 (0.004)

NOTE: Accuracy, specificity, Matthew correlation coefficient (MCC), F1 score, sensitivity, and residual mean squared error (RMSE) of precision matrix estimates, for each state $j = 1, \ldots, \hat{M}$. Standard deviations over the 30 simulations are displayed in parentheses. Results are reported for `sggmDAR`, `mvHMM`, and `glassoSlide`. Results for `sggmDAR` are conditioned upon the modal number of states and autoregressive order. A hyphen is used for those metrics that cannot be computed due to the structure of the underlying truth (e.g., TP+FN = 0).

symmetric Dirichlet distributed, with concentration parameter equal to one. Since this HMM approach does not estimate precision entries as exact zeros, we once again used 95% posterior credible intervals to perform edge selection. In the second approach, named `glassoSlide`, we followed Allen et al. (2014) and employed a sliding window to compute time-varying sparse inverse precision matrices via graphical lasso (Friedman, Hastie, and Tibshirani 2008) using the R package `glasso`. In order to obtain an estimate of the latent state sequence, the windowed estimates of the precision matrices were then used as feature vectors in the k-means clustering algorithm. Finally, sparse state-specific precision matrices were estimated by applying graphical lasso to the MLE estimates of the covariances relative to the set of observations corresponding to each distinct state. The number of states was set to five (i.e., the truth), while the size of the sliding window and the magnitude of the penalization parameter were selected in such a way to maximize model selection performances averaged over the different states.

To assess model selection performances we computed accuracy, sensitivity, specificity, $F1$-score and Matthew correlation coefficient (MCC), for each regime $j = 1, \ldots, \hat{M}$. In addition, to evaluate estimation accuracy, we calculated residual mean squared error (RMSE) of state-specific off-diagonal entries of the precision matrices as $\text{RMSE}_j = \sqrt{\frac{1}{D}\sum_{i<l}\left(\omega_{jil} - \hat{\omega}_{jil}\right)^2}$. Results from these measures are summarized in Table 1. Note that MCC, F1, and sensitivity for the Identity state are not presented since these metrics cannot be computed due to the structure of the underlying truth (e.g., TP+FN = 0). Overall, `sggmDAR` produced the best results both in estimation accuracy and model selection performances. Though accuracy and specificity of `glassoSlide` are somewhat high, this frequentist method is by far the worse, as illustrated by very low MCC scores. We remark that, while both `mvHMM` and `glassoSlide` need to specify the number of states in advance, our proposed approach produces an estimate of this parameter. In the supplementary material, we further investigate the performance of our proposed methodology for data-generating emissions with zero-mean, that is $\boldsymbol{\mu}_j = \mathbf{0}$, for $j = 1, \ldots, M$. The results confirm the superiority of our approach over both `mvHMM` and `glassoSlide` in terms of estimation and model selection accuracy.Additionally, the supplementary material contains a sensitivity analysis study that focuses on examining the impact of the hyperparameters a_v, b_v associated with the zero-inducing cumulative shrinkage prior (4). Results show that, for small and moderate T, different combinations of the hyperparameters may yield varying dynamics of the process. However, as T increases, such differences are not noticeable.

3.4. Simulations for Varying T and D

Next, we investigated performance for different values of the sample size T. For this, we generated 30 distinct time series for different sample sizes, $T = 100, 500, 1000, 5000, 10{,}000$, assuming $M = 3$ states and DAR order $P = 2$, with autoregressive probabilities specified as $\boldsymbol{\phi} = (0.2, 0.5, 0.3)$ and innovations set to $\boldsymbol{\pi} = (0.5, 0.3, 0.2)$. The emission means were generated as in the main simulation above, whereas the precision matrices were constructed using patterns (i), (iii), and (v) from Section 3.1. Here, to perform Bayesian inference we fixed the maximum number of states to $M_{\max} = 3$ and maximum DAR order to $P_{\max} = 2$. The hyperparameters were specified as in Section 3.1. Figure 4 displays boxplots over the 30 simulations of the posterior distributions of $\boldsymbol{\phi}$ and $\boldsymbol{\pi}$, for the different values of T, conditioned upon the modal number of states and autoregressive order. As it was to be expected, estimates for both ϕ_j and π_j showed larger variability for small sample sizes. Conversely, as T increases, the generating DAR dynamics became more apparent, and our inference procedure is indeed able to retrieve the correct parameters more accurately, for most cases.

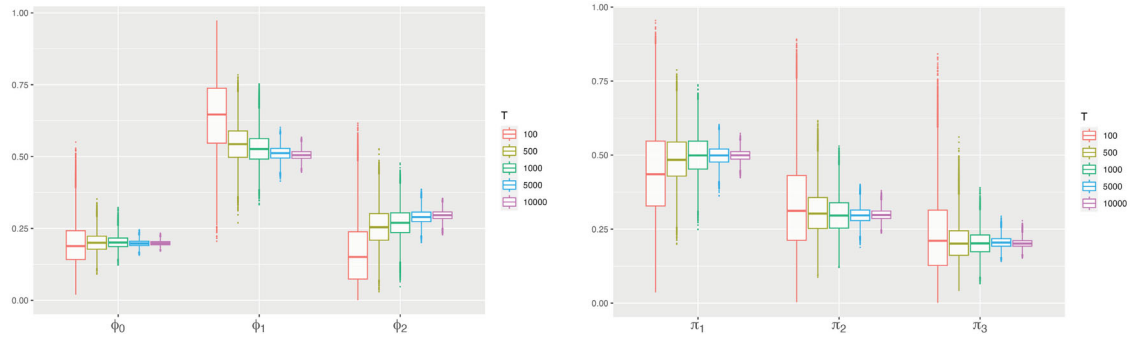

Figure 4. Simulation study. Boxplots over 30 simulations of posterior distributions for (left) autoregressive probabilities $\phi_l, l = 0, \ldots \hat{P}$, and (right) innovation probabilities $\pi_j, j = 1, \ldots, \hat{M}$, for different sample size $T = 100, 500, 1000, 5000, 10{,}000$. These results are conditioned on the modal number of states and autoregressive order.

Finally, we explored the performance of our approach in a scenario where the dimension D of the data is large. Here, we focused on assessing the ability of our proposed method in recovering the number of states, number of DAR parameters, and true sparse precision matrices. We simulated 30 time series, each consisting of $D = 100$-dimensional time series of length $T = 2000$, with $M = 3$ and $P = 2$, and with the emissions generated as in Section 3.1 and the precision matrices for the three states specified as Identity, Hub (with four blocks) and Random, respectively. The innovations were set to $\boldsymbol{\pi} = (0.6, 0.2, 0.2)$, while the rest of the data-generating parameters were set as in Section 3.1. Here we report results obtained by specifying the hyper-parameters as described in Section 3.1 and by running MCMC chains for 4000 iterations, with 1200 iterations discarded as burn-in.

As in the previous simulations, the correct number of states and DAR order were identified as those with the highest posterior probability for all replicated datasets. In the supplementary material, we report model selection and estimation accuracy performances for the off-diagonal component of the precision matrices, for `sggmDAR`, `mvHMM`, and `glassoSlide`. The MCC scores highlight the advantage of choosing our proposed method in large settings. Indeed, the number of parameters for each individual state is substantial, as there are 4950 distinct off-diagonal coefficients to be inferred for each precision matrix.

3.5. Additional Simulations

We conducted further simulations to assess the robustness of our model under two complementary settings: (i) a model misspecification scenario, where data are generated from a stationary vector autoregressive (VAR) process; and (ii) a realistic synthetic fMRI dataset designed to mimic experimental BOLD signals under structured noise conditions. These simulations provide insight into how the model behaves when its assumptions are violated and how well it performs in capturing task-driven neural dynamics. Full details are provided in the supplementary material (Sections F and G).

3.5.1. Misspecified Model

VAR is a popular model for both task and resting state fMRI, commonly used to capture temporal dependencies in neural time series. To explore the model's robustness under model misspecification, we generated multivariate time series from a stationary first-order VAR process with either sparse or dense autoregressive coefficient matrices. We summarize here the major findings. A complete description of the simulation setup and additional diagnostic plots are provided in the supplementary material (Section F).

In the sparse setting, where dependencies among dimensions are limited, the model consistently favored a single latent regime, with an estimated number of states $\hat{M} = 1$. This outcome reflects the relative homogeneity of the time series structure under sparsity, which does not induce detectable changes in statistical properties over time. In contrast, under a dense VAR structure, where most components influence one another, the model identified multiple regimes, with $\hat{M} = 3$. Despite the absence of true regime-switching in the data-generating process, the inferred latent states captured subtle shifts in the local mean and variance, as well as changes in the estimated connectivity patterns. Residual diagnostics, including standardized residuals and Q-Q plots, confirmed the adequacy of the model fit, with deviations confined to minor departures in the tails. These results highlight the model's capacity to adapt to a variety of temporal structures, effectively extracting interpretable latent states even under substantial model misspecification.

3.5.2. Simulated fMRI Data

To further assess the performance of our model in a realistic neuroimaging context, we generated synthetic fMRI data using the `neuRosim` package in R (Welvaert et al. 2011), which provides a principled framework for simulating blood-oxygen-level-dependent (BOLD) signals under controlled experimental and noise conditions. This simulation mimics a block-design experiment in which an external stimulus is applied every 80 sec and persists for 40 sec. The BOLD response is constructed using a canonical double-gamma hemodynamic response function (HRF), and several structured noise components-namely, temporal drift, physiological oscillations, and scanner-induced artifacts-are independently added to each brain region. The resulting signal is standardized to reflect typical preprocessing steps in fMRI analysis.

Figure 5 displays this process: panel (a) shows the final simulated multivariate fMRI time series; panel (b) presents the clean BOLD signal prior to noise addition; and panel (c) shows the posterior state probability plot obtained from our model. In this setting, the proposed approach correctly infers $\hat{M} = 2$ latent states, with state transitions that align closely with the stimulus schedule-despite the unsupervised nature of the analysis. This

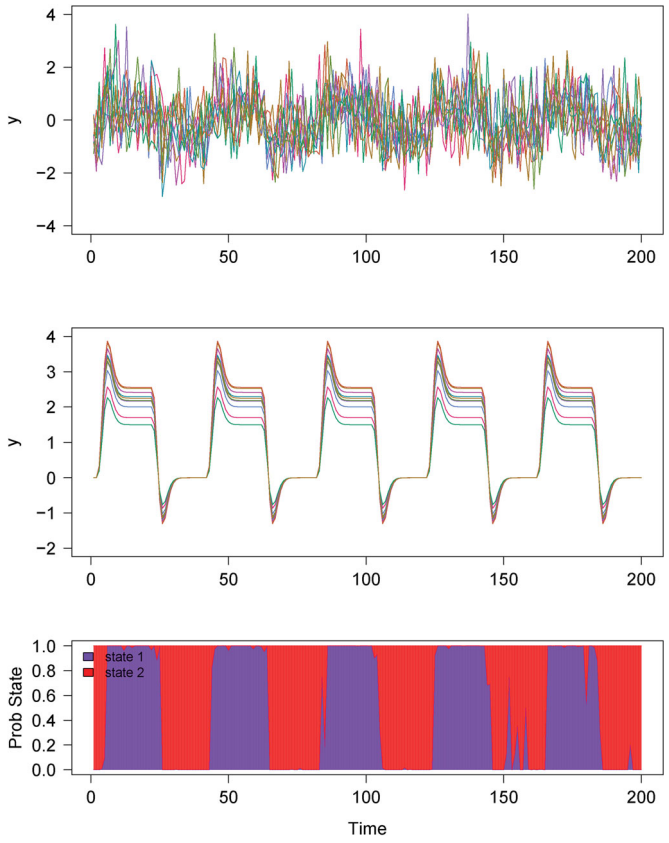

Figure 5. Simulated fMRI Data. (a) Simulated fMRI time series obtained by summing the clean BOLD signal with structured noise components and standardizing each region; (b) True underlying BOLD signal convolved with a canonical double-gamma HRF; (c) Estimated time-varying state probability plot.

confirms the model's ability to recover task-related structure in high-dimensional noisy time series data.

The model's estimated state-specific mean and variance profiles (reported in the supplementary material, Section G) highlight interpretable differences between the two inferred states. One state corresponds to higher overall activation across regions, while the other reflects baseline or resting activity. These findings demonstrate the method's ability to isolate functionally relevant brain states. Additional diagnostics, including standardized residuals and Q-Q plots, further confirm the model's goodness-of-fit, suggesting that the proposed framework effectively characterizes stimulus-driven neural dynamics even under complex noise conditions.

4. Application to fMRI Data

Identifying the dynamic nature of brain connectivity is critical for understanding and advancing our current knowledge about human brain functioning. Functional magnetic resonance imaging (fMRI), which measures brain activity by detecting changes associated with blood flow, has become a successful and effective instrument for studying how the brain functions. Here we consider the problem of estimating brain connectivity, that is, statistical dependence between fMRI time series in distinct regions of the brain. Recent evidence has shown that the interactions among brain regions vary over the course of an fMRI experiment, suggesting that brain connectivity is a dynamic process (Cribben, Wager, and Lindquist 2013; Lindquist et al. 2014; Xu and Lindquist 2015; Vidaurre, Smith, and Woolrich 2017; Vidaurre et al. 2018; Warnick et al. 2018). In this context, our proposed modeling framework is particularly well suited to address two key challenges in the analysis of fMRI data. First, the number of distinct cognitive or connectivity states that manifest during the course of the task is unknown and likely to vary across individuals or experimental conditions. Rather than assuming a fixed number of states, our model infers the number of latent regimes in a data-driven fashion using a sparsity-inducing prior. Second, the temporal structure governing transitions between these regimes may involve dependencies that extend beyond first-order Markovian assumptions. Employing a DAR process of unknown order allows for more flexible and biologically plausible switching dynamics. Together, these features enable the identification of meaningful patterns in brain connectivity that evolve over time, offering insight into latent cognitive processes.

We apply the proposed model to fMRI data from a subject performing a task-based experiment where the interest is to identify the neural representations that are formed during latent learning of predictive sequences (Bornstein and Daw 2012). In this experiment, participants carried out a task in which they observed a sequence of black-and-white natural scene images. They were instructed only to press a keyboard key ("d", "f", "j", "k") that they had previously been trained to associate with each image. Throughout the trials, the series of pictures were generated according to a first-order Markov process, though the participants were not aware of this structure. This form of task has been used to examine the cognitive and neural architecture of latent learning and the use of learned representations in support of predictive lookahead, a core computational process supporting decision-making in humans and animals (Strange et al. 2005; Harrison, Duggins, and Friston 2006; Bornstein and Daw 2013; Hunter, Bornstein, and Hartley 2018; Morris, Bornstein, and Shenhav 2018; Rmus et al. 2022; Yoo, Bornstein, and Chrastil 2024). Here, response times indicated the degree to which the participant implicitly expected the currently presented image, on the basis of the previously presented one. A consistent finding in these tasks is that participants implicitly learn the sequential structure, and that neural regions signal the degree to which they anticipate the upcoming image in the sequence. Several studies have identified more than one representation of sequential structure, each of which has an influence on behavior as estimated across the entire experiment. However, it is unclear how these multiple representations are arbitrated among to influence behavior—for example, as a weighted mixture at the single-trial level (Wang, Feng, and Bornstein 2022; Khoudary, Peters, and Bornstein 2022; Nicholas, Daw, and Shohamy 2022), in alternation according to regimes of task statistics (Poldrack et al. 2001; Daw, Niv, and Dayan 2005; Lengyel and Dayan 2007; Yoo, Bornstein, and Chrastil 2024), or as a fixed proportion that varies according to individual differences such as in memory encoding precision Noh et al. (2023). Full details of the experiment are provided in Bornstein and Daw (2012).

The scanning session proceeded with four blocks consisting of 275 fMRI acquisitions. For the analyses of this article we concatenated the four blocks and subtracted the mean of each block. $D = 18$ lateralized regions of interest (ROIs) were selected on the basis of prior findings using this task

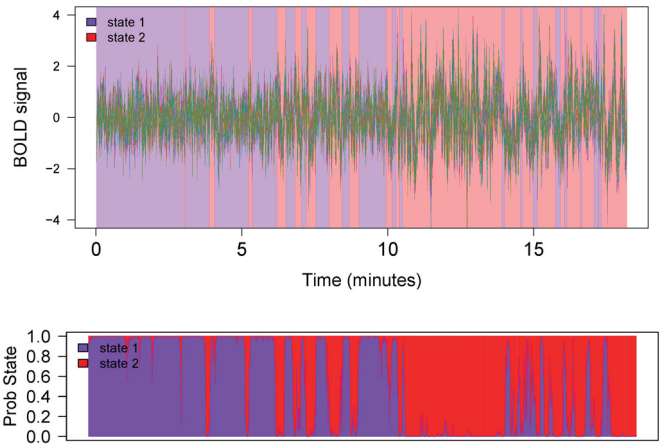

Figure 6. Application to FMRI data. (top) BOLD time series data from one participant, with each brain region represented by a different line; vertical bands represent the estimated state sequence. (bottom) estimated time-varying state probability plot.

(Bornstein and Daw 2012, 2013) as those most sensitive to one or more of the identified representations of sequential structure (dorsal and ventral striatum, hippocampus), to the degree of conflict between the representations (anterior cingulate cortex), or to the stimulus content (scene images; ventral visual stream regions). The observed time series correspond to the preprocessed and standardized blood-oxygen-level-dependent (BOLD) signals recorded from each of the selected brain regions, serving as proxies for local neural activity across time. We scaled each dimension $d = 1, \ldots, D$, to have variance one. In our analysis, we seek to identify distinct regimes of functional connectivity that can be mapped onto cognitive interpretation—specifically, to identify the manner in which multiple representations combine to control behavior. Since we do not have prior information on the cognitive states that are manifested during the experiment, we assumed that the number of the states is unknown, as well as the latent learning structure driving the switching of regimes. We set the hyperparameters of the model as described in Section 3.1 and ran MCMC chains with 4000 iterations, 1200 of which were discarded as burn-in.

The BOLD time series data for the $D = 18$ selected ROIs are shown in the top panel of Figure 6, along with the estimated latent state sequence. Our model identified a mode at $\hat{K} = 2$ distinct states and a DAR order $\hat{P} = 2$, with estimated values of innovations and autoregressive parameters $\hat{\boldsymbol{\pi}} = [0.50, 0.50]$ and $\hat{\boldsymbol{\phi}} = [0.08, 0.87, 0.05]$, respectively. The bottom panel of Figure 6 displays the time-varying probability plot, namely the local decoding of the hidden state at time t, $p(\gamma_t = j \,|\, \boldsymbol{y},)$, for $t = 1, \ldots, T$, as described in Section 2.4. These probabilities are represented with a different color for each of the two inferred states, and they cumulatively add to one for each t. The state probability plot displays a clear transition from state 1 to state 2, approximately halfway through the task.

The top panel of Figure 7 shows the estimated state-specific partial correlation matrices, for the two estimated states, and the bottom panel the corresponding estimated connectivity graphs, with edges identified through the procedure described in Section 2.2. State 1 has relatively stronger connectivity between hippocampus (HC) and anterior cingulate cortex (ACC), with mean difference between states across ROI pairs equal to 0.028; whereas State 2 shows stronger connectivity between Caudate and ACC, with mean difference between states across ROI pairs equal to 0.048. Across all ROI pairs, the average difference in partial correlation values between states is 0.002. In the supplementary material, Section C.2, we provide additional analysis of the fMRI data, including state-specific mean and variance estimates, standardized residual diagnostics, and ACF/PACF plots to further assess model fit and temporal structure. These results show that while mean activation remains largely stable across states, variability differs meaningfully, suggesting distinct dynamic regimes. Residual diagnostics, including standardized residuals and Q-Q plots, further support model adequacy, indicating no major misspecification and approximate normality of the residuals. Together, these findings suggest that our model effectively captures the dominant temporal structure of the data, achieving a balance between flexibility and interpretability in the context of dynamic brain connectivity.

These observations are consistent with findings in the literature that at least two distinct networks mediate expectations in this task: one centered on hippocampus and thought to encode stimulus-stimulus predictive relationships (e.g., "cognitive maps"), and the other centered on striatum and thought to encode response-response sequences (Bornstein and Daw 2012, 2013). Each has different dynamics with regard to the predictiveness of the learned sequences: activity in the hippocampal network scales with increasing uncertainty about the next item in the sequence, consistent with its proposed role in "pre-fetching" upcoming states in support of decision-making (Johnson and Redish 2007); separately, activity in the striatal network *decreases* with uncertainty about the next item in the sequence, consistent with observations that this structure is more strongly activated by highly predictive associations (Smith and Graybiel 2016). The observation that the network regime corresponds to shifts in its connectivity with anterior cingulate cortex is consistent with theoretical accounts of this region as signaling the "expected value of control," mediating the influence of internal representations on behavior (Shenhav, Botvinick, and Cohen 2013). The transition between hippocampal and striatally-mediated regimes is consistent with extensive empirical findings that these regions "trade-off" in control of behavior across highly repeated tasks, with hippocampus driving responses early on and striatum taking over when sequences are more well-practiced (Poldrack et al. 2001; Lengyel and Dayan 2007).

5. Concluding Remarks

We have presented a flexible Bayesian approach for estimating sparse Gaussian graphical models based on time series data. In order to represent switching dynamics of the time series data, we have assumed an unobserved hidden process underlying the data, with observations generated from state-specific multivariate Gaussian emission distributions. We have modeled the temporal structure of the hidden state sequence based on a DAR process, as a flexible approach to incorporate temporal dynamics that extend beyond simple Markovian structures. We have modeled the time-varying mixing probabilities capturing the state-switching behavior of the DAR process via a cumulative shrinkage nonparametric prior that accommodates zero-

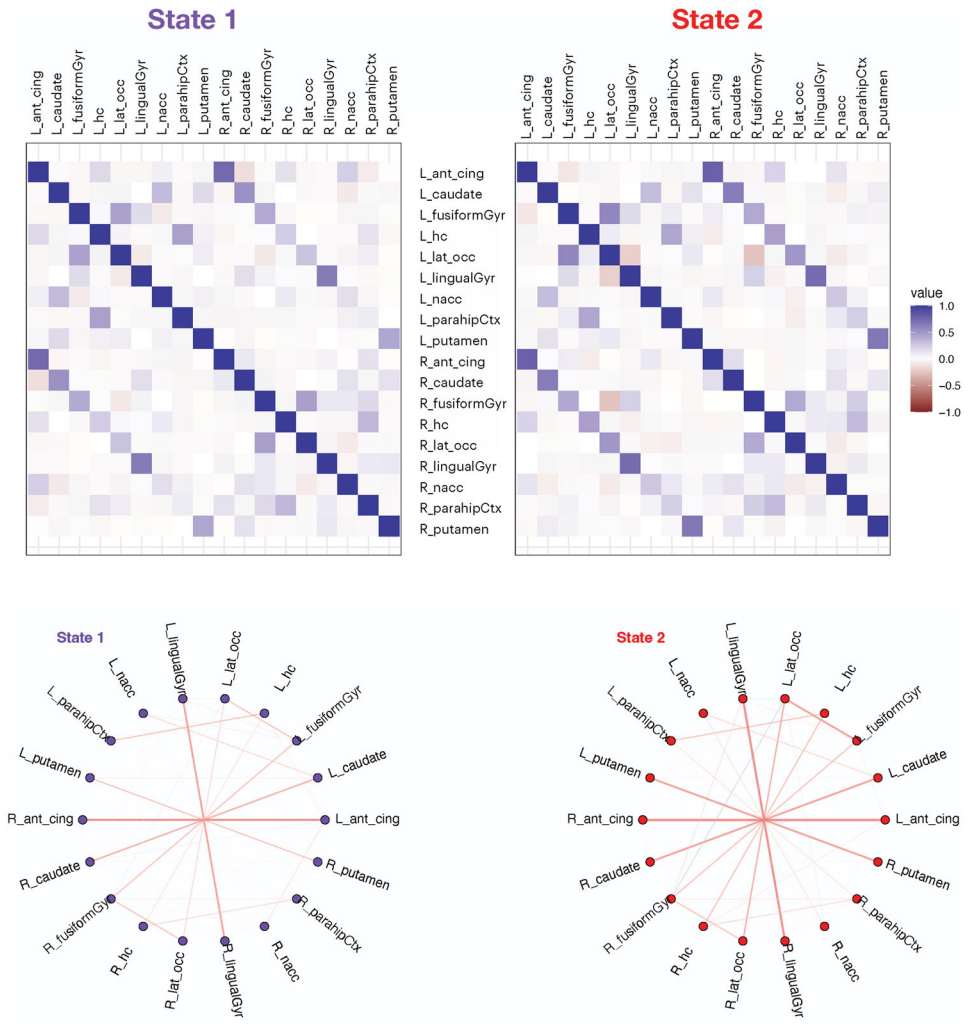

Figure 7. Application to FMRI data. (top) estimated partial correlation matrices, for each of the two inferred states. (bottom) estimated state-specific connectivity graphs.

inflated parameters for non-active components. The proposed formulation ensures that if a parameter in the DAR model is zero, then all subsequent lag parameters are also zero, yielding a flexible and computationally efficient modeling framework for estimating the time-varying mixing probabilities as well as the effective order of the process. This considerably speeds up the posterior sampler, especially in regard to the forward-backward scheme for updating the latent state sequence. We have additionally integrated a sparsity-inducing Dirichlet prior to estimate the effective number of states in a data-driven manner. At the network level, we have assumed a graphical horseshoe prior to induce sparsity in the state-specific precision matrices. We have thoroughly investigated the performance of our methods through simulation studies and performed comparisons with competing approaches. We have further illustrated our proposed approach for the estimation of dynamic brain connectivity based on fMRI data collected on a subject performing a task-based experiment on latent learning.

Supplementary Materials

Supplement: We include further details about backward and forward messages for our sampling algorithm. We also report results from additional simulations, sensitivity analyses and convergence diagnostics of the MCMC.

Software: `sggmDAR` - a Julia software implementing the methodology outlined in the article, accompanied by a comprehensive tutorial designed to guide users through replicating the findings detailed in the article. The software `sggmDAR` is also available on GitHub at https://github.com/Beniamino92/sggmDAR.

Acknowledgments

B.H.-A. completed part of this work as a postdoctoral fellow in the Department of Statistics at Rice University, Houston, TX.

Disclosure Statement

No potential competing interest was reported by the authors.

ORCID

Beniamino Hadj-Amar http://orcid.org/0000-0002-2265-0712
Michele Guindani http://orcid.org/0000-0002-6363-9907
Marina Vannucci http://orcid.org/0000-0002-7360-5321

References

Allen, E. A., Damaraju, E., Plis, S. M., Erhardt, E. B., Eichele, T., and Calhoun, V. D. (2014), "Tracking Whole-Brain Connectivity Dynamics in the Resting State," *Cerebral Cortex*, 24, 663–676. [9]

Biswas, A., and Song, P. X.-K. (2009), "Discrete-Valued ARMA Processes," *Statistics & Probability Letters*, 79, 1884–1889. [1,3]

Bornstein, A. M., and Daw, N. D. (2012), "Dissociating Hippocampal and Striatal Contributions to Sequential Prediction Learning," *European Journal of Neuroscience*, 35, 1011–1023. [11,12]

Bornstein, A. M., and Daw, N. D. (2013), "Cortical and Hippocampal Correlates of Deliberation During Model-based Decisions for Rewards in Humans," *PLoS Computational Biology*, 9, e1003387. [11,12]

Cappé, O., Moulines, E., and Rydén, T. (2005), *Inference in Hidden Markov Models*, New York: Springer-Verlag. [2]

Cribben, I., Wager, T. D., and Lindquist, M. A. (2013), "Detecting Functional Connectivity Change Points for Single-Subject fMRI Data," *Frontiers in Computational Neuroscience*, 7, 143. [1,11]

Danaher, P., Wang, P., and Witten, D. M. (2014), "The Joint Graphical Lasso for Inverse Covariance Estimation Across Multiple Classes," *Journal of the Royal Statistical Society*, Series B, 76, 373–397. [1]

Daw, N. D., Niv, Y., and Dayan, P. (2005), "Uncertainty-based Competition between Prefrontal and Dorsolateral Striatal Systems for Behavioral Control," *Nature Neuroscience*, 8, 1704–1711. [11]

DeRuiter, S. L., Langrock, R., Skirbutas, T., Goldbogen, J. A., Calambokidis, J., Friedlaender, A. S., and Southall, B. L. (2017), "A Multivariate Mixed Hidden Markov Model for Blue Whale Behaviour and Responses to Sound Exposure," *The Annals of Applied Statistics*, 11, 362–392. [2]

Fox, E. B., Sudderth, E. B., Jordan, M. I., and Willsky, A. S. (2011), "A Sticky HDP-HMM with Application to Speaker Diarization," *The Annals of Applied Statistics*, 5, 1020–1056. [5,6]

Friedman, J., Hastie, T., and Tibshirani, R. (2008), "Sparse Inverse Covariance Estimation with the Graphical LASSO," *Biostatistics*, 9, 432–441. [9]

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (1995), *Bayesian Data Analysis*, New York: Chapman and Hall/CRC. [5]

Green, P. J. (1995), "Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination," *Biometrika*, 82, 711–732. [5]

Hadj-Amar, B., Finkenstädt, B., Fiecas, M., and Huckstepp, R. (2021), "Identifying the Recurrence of Sleep Apnea Using a Harmonic Hidden Markov Model," *The Annals of Applied Statistics*, 15, 1171. [5,6]

Harrison, L. M., Duggins, A., and Friston, K. J. (2006), "Encoding Uncertainty in the Hippocampus," *Neural Networks*, 19, 535–546. [11]

Heidelberger, P., and Welch, P. D. (1981), "A Spectral Method for Confidence Interval Generation and Run Length Control in Simulations," *Communications of the Association for Computing Machinery*, 24, 233–245. [8]

Heiner, M., Kottas, A., and Munch, S. (2019), "Structured Priors for Sparse Probability Vectors with Application to Model Selection in Markov Chains," *Statistics and Computing*, 29, 1077–1093. [2]

Holsclaw, T., Greene, A. M., Robertson, A. W., and Smyth, P. (2017), "Bayesian Nonhomogeneous Markov Models via Pólya-Gamma Data Augmentation with Applications to Rainfall Modeling," *The Annals of Applied Statistics*, 11, 393–426. [2]

Hunter, L. E., Bornstein, A. M., and Hartley, C. A. (2018), "A Common Deliberative Process Underlies Model-based Planning and Patient Intertemporal Choice," bioRxiv, p. 499707. [11]

Ishwaran, H., and James, L. F. (2001), "Gibbs Sampling Methods for Stick-Breaking Priors," *Journal of the American Statistical Association*, 96, 161–173. [2]

Jasra, A., Holmes, C. C., and Stephens, D. A. (2005), "Markov Chain Monte Carlo Methods and the Label Switching Problem in Bayesian Mixture Modeling," *Statistical Science*, 20, 50–67. [6]

Johnson, A., and Redish, A. D. (2007), "Neural Ensembles in CA3 Transiently Encode Paths Forward of the Animal at a Decision Point," *Journal of Neuroscience*, 27, 12176–12189. [12]

Kastner, G., and Huber, F. (2020), "Sparse Bayesian Vector Autoregressions in Huge Dimensions," *Journal of Forecasting*, 39, 1142–1165. [1]

Khoudary, A., Peters, M. A., and Bornstein, A. M. (2022), "Precision-Weighted Evidence Integration Predicts Time-Varying Influence of Memory on Perceptual Decisions," *Cognitive Computational Neuroscience*. DOI:10.32470/CCN.2022.1292-0. [11]

Kolar, M., Song, L., Ahmed, A., and Xing, E. P. (2010), "Estimating Time-Varying Networks," *The Annals of Applied Statistics*, 4, 94–123. [1]

Lam, C., and Yao, Q. (2012), "Factor Modeling for High-Dimensional Time Series: Inference for the Number of Factors," *The Annals of Statistics*, 40, 694–726. [1]

Legramanti, S., Durante, D., and Dunson, D. B. (2020), "Bayesian Cumulative Shrinkage for Infinite Factorizations," *Biometrika*, 107, 745–752. [2,3]

Lengyel, M., and Dayan, P. (2007), "Hippocampal Contributions to Control: The Third Way," in *Advances in Neural Information Processing Systems* (Vol. 20). [11,12]

Li, Y., Craig, B. A., and Bhadra, A. (2019), "The Graphical Horseshoe Estimator for Inverse Covariance Matrices," *Journal of Computational and Graphical Statistics*, 28, 747–757. [4,5,6]

Lindquist, M. A., Xu, Y., Nebel, M. B., and Caffo, B. S. (2014), "Evaluating Dynamic Bivariate Correlations in Resting-State fMRI: A Comparison Study and a New Approach," *NeuroImage*, 101, 531–546. [1,11]

Malsiner-Walli, G., Frühwirth-Schnatter, S., and Grün, B. (2016), "Model-based Clustering based on Sparse Finite Gaussian Mixtures," *Statistics and Computing*, 26, 303–324. [2,4]

Morris, R. W., Bornstein, A., and Shenhav, A. (2018), *Goal-Directed Decision Making: Computations and Neural Circuits*, Amsterdam: Elsevier. [11]

Neal, R. M. (2003), "Slice Sampling," *The Annals of Statistics*, 31, 705–767. [5]

Nicholas, J., Daw, N. D., and Shohamy, D. (2022), "Uncertainty Alters the Balance between Incremental Learning and Episodic Memory," *Elife*, 11, e81679. [11]

Noh, S. M., Singla, U. K., Bennett, I. J., and Bornstein, A. M. (2023), "Memory Precision and Age Differentially Predict the Use of Decision-Making Strategies Across the Lifespan," *Scientific Reports*, 13, 17014. [11]

Papastamoulis, P., and Iliopoulos, G. (2013), "On the Convergence Rate of Random Permutation Sampler and ECR Algorithm in Missing Data Models," *Methodology and Computing in Applied Probability*, 15, 293–304. [6]

Pegram, G. (1980), "An Autoregressive Model for Multilag Markov Chains," *Journal of Applied Probability*, 17, 350–362. [3]

Poldrack, R. A., Clark, J., Pare-Blagoev, E., Shohamy, D., Creso Moyano, J., Myers, C., and Gluck, M. A. (2001), "Interactive Memory Systems in the Human Brain," *Nature*, 414, 546–550. [11,12]

Qiu, H., Han, F., Liu, H., and Caffo, B. (2016), "Joint Estimation of Multiple Graphical Models from High Dimensional Time Series," *Journal of the Royal Statistical Society*, Series B, 78, 487–504. [1]

Quinn, A. J., Vidaurre, D., Abeysuriya, R., Becker, R., Nobre, A. C., and Woolrich, M. W. (2018), "Task-Evoked Dynamic Network Analysis through Hidden Markov Modeling," *Frontiers in Neuroscience*, 12, 603. [1,2]

Rmus, M., Ritz, H., Hunter, L. E., Bornstein, A. M., and Shenhav, A. (2022), "Humans Can Navigate Complex Graph Structures Acquired During Latent Learning," *Cognition*, 225, 105103. [11]

Rodríguez, C. E., and Walker, S. G. (2014), "Label Switching in Bayesian Mixture Models: Deterministic Relabeling Strategies," *Journal of Computational and Graphical Statistics*, 23, 25–45. [6]

Rousseau, J., and Mengersen, K. (2011), "Asymptotic Behaviour of the Posterior Distribution in Overfitted Mixture Models," *Journal of the Royal Statistical Society*, Series B, 73, 689–710. [2,4]

Safikhani, A., and Shojaie, A. (2022), "Joint Structural Break Detection and Parameter Estimation in High-Dimensional Nonstationary VAR Models," *Journal of the American Statistical Association*, 117, 251–264. [1]

Sarkar, A., and Dunson, D. B. (2019), "Bayesian Higher Order Hidden Markov Models," arXiv preprint arXiv:1805.12201. [2]

Savitsky, T., Vannucci, M., and Sha, N. (2011), "Variable Selection for Nonparametric Gaussian Process Priors: Models and Computational Strategies," *Statistical Science: A Review Journal of the Institute of Mathematical Statistics*, 26, 130–149. [4]

Shenhav, A., Botvinick, M. M., and Cohen, J. D. (2013), "The Expected Value of Control: An Integrative Theory of Anterior Cingulate Cortex Function," *Neuron*, 79, 217–240. [12]

Smith, K. S., and Graybiel, A. M. (2016), "Habit Formation Coincides with Shifts in Reinforcement Representations in the Sensorimotor Striatum," *Journal of Neurophysiology*, 115, 1487–1498. [12]

Song, L., Kolar, M., and Xing, E. (2009), "Time-Varying Dynamic Bayesian Networks," in *Advances in Neural Information Processing Systems* (Vol. 22). [1]

Strange, B. A., Duggins, A., Penny, W., Dolan, R. J., and Friston, K. J. (2005), "Information Theory, Novelty and Hippocampal Responses: Unpredicted or Unpredictable?" *Neural Networks*, 18, 225–230. [11]

Tang, Z.-Z., and Chen, G. (2019), "Zero-Inflated Generalized Dirichlet Multinomial Regression Model for Microbiome Compositional Data Analysis," *Biostatistics*, 20, 698–713. [2]

Vidaurre, D., Abeysuriya, R., Becker, R., Quinn, A. J., Alfaro-Almagro, F., Smith, S. M., and Woolrich, M. W. (2018), "Discovering Dynamic Brain Networks from Big Data in Rest and Task," *NeuroImage*, 180, 646–656. [11]

Vidaurre, D., Smith, S. M., and Woolrich, M. W. (2017), "Brain Network Dynamics are Hierarchically Organized in Time," *Proceedings of the National Academy of Sciences*, 114, 12827–12832. [11]

Wang, S., Feng, S. F., and Bornstein, A. M. (2022), "Mixing Memory and Desire: How Memory Reactivation Supports Deliberative Decision-Making," *Wiley Interdisciplinary Reviews: Cognitive Science*, 13, e1581. [11]

Warnick, R., Guindani, M., Erhardt, E., Allen, E., Calhoun, V., and Vannucci, M. (2018), "A Bayesian Approach for Estimating Dynamic Functional Network Connectivity in fMRI Data," *Journal of the American Statistical Association*, 113, 134–151. [1,2,11]

Welvaert, M., Durnez, J., Moerkerke, B., Berdoolaege, G., and Rosseel, Y. (2011), "neurosim: An r Package for Generating fMRI Data," *Journal of Statistical Software*, 44, 1–18. [10]

Xu, Y., and Lindquist, M. A. (2015), "Dynamic Connectivity Detection: An Algorithm for Determining Functional Connectivity Change Points in fMRI Data," *Frontiers in Neuroscience*, 9, 285. [1,11]

Yoo, J., Bornstein, A., and Chrastil, E. R. (2024), "Cognitive Graphs: Representational Substrates for Planning," *Decision*, 11, 537–556. [11]

Zhang, W., Cribben, I., Petrone, S., and Guindani, M. (2021), "Bayesian Time-Varying Tensor Vector Autoregressive Models for Dynamic Effective Connectivity," ArXiv:2106.14083. [1,3]