

Approximate Hidden Semi-Markov Models for Dynamic Connectivity Analysis in Resting-State fMRI

Mark B. Fiecas¹, Christian Coffman¹, Meng Xu¹, Timothy J. Hendrickson², Bryon A. Mueller³, Bonnie Klimes-Dougan⁴, and Kathryn R. Cullen³

¹*Division of Biostatistics, University of Minnesota*

²*University of Minnesota Informatics Institute*

³*Department of Psychiatry and Behavioral Sciences, University of Minnesota*

⁴*Department of Psychology, University of Minnesota*

Abstract

Motivated by a study on adolescent mental health, we conduct a dynamic connectivity analysis using resting-state functional magnetic resonance imaging (fMRI) data. A dynamic connectivity analysis investigates how the interactions between different regions of the brain, represented by the different dimensions of a multivariate time series, change over time. Hidden Markov models (HMMs) and hidden semi-Markov models (HSMMs) are common analytic approaches for conducting dynamic connectivity analyses. However, existing approaches for HSMMs are limited in their ability to incorporate covariate information. In this work, we approximate an HSMM using an HMM for modeling multivariate time series data. The approximate HSMM model allows one to explicitly model dwell-time distributions that are available to HSMMs, while maintaining the theoretical and methodological advances that are available to HMMs. We conducted a simulation study to show the performance of the approximate HSMM relative to other approaches. Finally, we used the approximate HSMM to conduct a dynamic connectivity analysis, where we showed how dwell-time distributions vary across the severity of non-suicidal self-injury (NSSI) in female adolescents, and how those with moderate or severe NSSI have greater state-switching frequency.

Keywords: Dynamic functional connectivity, fMRI, Hidden Markov models, Hidden semi-Markov models, Multivariate time series

1 Introduction

The motivating data for this article is the resting-state functional magnetic resonance imaging (fMRI) data obtained from female adolescents in the Brain Imaging Development of Girls' Emotion and Self (BRIDGES) Study (<http://radlab.umn.edu/current-research/bridges-brain-imaging-development-girls-emotion-and-self>). For this article, the adolescents were categorized into two clinical groups: adolescents with moderate or severe non-suicidal self-injury (NSSI), and those with no or mild NSSI. The resting-state fMRI data is represented as a multivariate time series, and the time series are independent across the individuals. Our overall objective in this article is to develop a statistical model for *dynamic connectivity analysis*, and determine how dynamic connectivity differs between the two clinical groups.

Dynamic connectivity analysis is the study of how the regions of the brain, which are represented by the different dimensions of the time series, are correlated and how this correlation is potentially changing over time. Dynamic connectivity analysis is still a relatively young approach for fMRI data, but there have been substantial evidence of its utility (Allen et al., 2014; Calhoun et al., 2014; Liégeois et al., 2019; Rashid et al., 2016; Shappell et al., 2021). Existing analytic approaches range from the use of sliding-window correlations (Allen et al., 2014), hidden Markov models (HMM) and hidden semi-Markov models (HSMM) (Vidaurre et al., 2017; Shappell et al., 2019), change-point analysis (Cribben et al., 2012), and time series models with time-varying parameters (Lindquist et al., 2014). Calhoun et al. (2014) gave a thorough overview of analytic approaches for dynamic connectivity analysis.

We focus our attention on HMMs and HSMMs. These models assume that there is a sequence of unobserved states over time, and the properties of the data at each time point within an individual is determined by the current state at that time point for that individual. In our case, we will characterize the states depending on the patterns of *connectivity*, i.e., the structure of the correlation matrix of the multivariate time series. Thus, as the state sequence moves from one state to another state, the correlation matrix will also change. After fitting an HMM or HSMM, one could extract summary statistics such as the number of state switches or the amount of time spent in a state, and one could conduct post-hoc analyses about these summary statistics. HMMs and HSMMs have been useful analytic approaches for modeling multivariate time series data from neuroimaging studies. Indeed, a number of studies have previously used HMMs, HSMMs, or their variants for dynamic connectivity analyses using fMRI data (Baker et al., 2014; Shappell et al., 2019, 2021; Taghia et al., 2017; Vidaurre et al., 2017; Warnick et al., 2018). In addition, neuroimaging studies using other modalities such as electroencephalograms and magnetoencephalograms have also used this modeling framework (Obermaier et al., 2001; Ombao et al., 2018; Quinn et al., 2018).

Thus, HMMs and HSMMs have broad appeal for modeling time series data in neuroimaging studies.

HSMMs contain HMMs as a special case. In an HMM, the state sequence is governed by transition probabilities such that the state sequence will either stay in its current state or transition to a different state with a certain probability. Implicit here is that the duration within a state follows a geometric distribution. Consequently, shorter durations occur with higher probabilities. In an HSMM, state transitions are still governed by transition probabilities, but the duration of the state sequence within a state prior to transitioning to a different state is governed by a probability distribution. Thus, an HSMM explicitly models the distribution of the duration within a state, e.g., by assuming that the distribution follows a Poisson distribution or some other discrete distribution. This additional flexibility can affect summary statistics such as the number of times points in a state or how often a state transition occurs. Depending on the context of the problem, practitioners may need the explicit modeling of each state's duration distribution.

There have been a number of methodological advances for HMMs for multivariate time series. In contrast, the methodological advances for HSMMs are limited in number, especially in the context where we observe multivariate time series data from independent sources, e.g., the female adolescents in the BRIDGES Study. Bulla and Bulla (2006) developed HSMMs for univariate time series data, and software implementation was later provided by Bulla et al. (2010). O'Connell et al. (2011) developed the `mhsmm` package in R, which is capable of fitting HSMMs for multivariate time series data from multiple independent sources. Shappell et al. (2019) and Shappell et al. (2021) used this package for the dynamic connectivity analyses in their fMRI studies. An important limitation with their approach is that observed covariates, e.g., NSSI severity, cannot be embedded directly into the model and so the impact of covariates must be investigated in post-hoc analyses. Langrock and Zucchini (2011) showed how one could structure an HMM such that it *approximates* an HSMM and its dwell-time distributions with arbitrary accuracy, and Langrock et al. (2012) illustrated the utility of this approach in modeling animal telemetry data. We build on the modeling framework by Langrock and Zucchini (2011) for multivariate time series data and use it to conduct a dynamic connectivity analysis on the resting-state fMRI time series data from the BRIDGES Study. This modeling framework allows us to have explicit models for the dwell-time distributions that are potentially modulated by covariate information, and we will have available to us the computational tractability and the theoretical and methodological advances already available for HMMs.

The rest of this article is organized as follows. In Section 2, we give a more detailed overview of HMMs and HSMMs, and we describe how one could use an HMM to approximate an HSMM. In Section 3, we use a simulation study to illustrate the performance of the approximate HSMM relative to other models. In Section 4, we show the empirical utility of the approximate HSMM by using it in a dynamic connectivity analysis on the resting-state fMRI data from the BRIDGES Study. Finally, in Section 5 we end with a

discussion of the analysis, potential extensions, and limitations.

2 Modeling Dynamic Connectivity

2.1 Hidden Markov Models and Hidden Semi-Markov Models

We now give an overview of the context of the problem. Let $\{\mathbf{y}_{t,n}\}_{n=1}^N$ be a collection of P -variate time series observed from N independent subjects, and we assume that the time series for each subject we have $t = 1, \dots, T$. Denote $\mathbf{y}_n = (\mathbf{y}_{1,n}, \dots, \mathbf{y}_{T,n})$. Furthermore, let \mathbf{Z} be the matrix of covariates whose n -th row contains the covariate values for subject n . For subject n , let $S_{t,n}$ be an unobserved finite-state process that takes on discrete values in $\{1, \dots, M\}$ such that, conditional on the states, the observed data \mathbf{y}_n are independent and follow the P -variate normal distribution, also called the *emission distribution*. The mean vectors and covariance matrices of the emission distributions, $\{\boldsymbol{\mu}_m\}_{m=1}^M$ and $\{\boldsymbol{\Sigma}_m\}_{m=1}^M$, respectively, therefore characterize the states. As the state process switches from one state to the next, the correlation between the dimensions of $\mathbf{y}_{t,n}$ also changes. Thus, to model the temporal dynamics of the correlation structure of the time series, our goal is to estimate the covariance matrix in the emission distribution for each of the M unobserved states using the collection of both the observed time series $\{\mathbf{y}_{t,n}\}_{n=1}^N$ and covariate matrix \mathbf{Z} .

We now give an overview of the Markov and semi-Markov assumptions for $\{S_{t,n}\}_{n=1}^N$ and the hidden Markov and semi-Markov models that arise. If we assume that the n -th subject's finite-state process $S_{t,n}$ is a Markov process, then for each n , we have transition probabilities $a_{ij} = \Pr(S_{t,n} = j | S_{t-1,n} = i)$ with $\sum_j a_{ij} = 1$, and these form the transition probability matrix $\mathbf{A} = (a_{ij})$. Note that the Markov assumption for the state sequence leads to the assumption that the *dwelt-time* (or *sojourn time*) within a state, i.e., the number of consecutive time points that the Markov chain spends in a state, follows a geometric distribution. A hidden *semi*-Markov model (HSMM) relaxes this assumption by allowing the dwell-time for state i follow a discrete distribution with probability mass function $p_i(\cdot)$ that is potentially parameterized by a vector $\boldsymbol{\beta}_i$. For example, the dwell-time distribution for state i could be a Poisson distribution with rate parameter λ_i , and we could have a model for this rate parameter, e.g., $\log(\lambda_i) = \mathbf{Z}\boldsymbol{\beta}_i$. The dwell-time distribution characterizes how long the state sequence stays in a state before switching to a different state (i.e., not itself) with some probability. Like the HMM, the state switches in the HSMM are governed by a transition probability matrix $\mathbf{A} = (a_{ij})$, though the diagonal elements of this matrix are 0. The HSMM contains the HMM as a special case if we let each $p_i(\cdot)$ denote the probability mass function for the geometric distribution. We give an illustration of both the HMM and the HSMM in Figure 1.

[Figure 1 about here.]

2.2 Approximate Hidden Semi-Markov Models

The greater flexibility of the HSMM relative to the HMM results in the likelihood function of the HSMM being more complicated relative to the likelihood function of the HMM, and modeling the dwell-time distribution in the HSMM leads to a substantial computational cost for fitting the model. Our goal is to work in a middle ground in having flexibility in modeling dwell-time distributions with computational complexity. To this end, we follow Langrock and Zucchini (2011) in structuring an HMM in a very deliberate way such that the resulting HMM approximates an HSMM with any form for the dwell-time distributions. In the following, for each n , suppose that $S_{t,n}$ is a M -state semi-Markov process with $M \times M$ transition probability matrix $\mathbf{A} = (a_{ij})$. We now create a Markov process $\tilde{S}_{t,n}$ such that an aggregate of its states will approximate $S_{t,n}$ and its dwell-time. Let m_1, \dots, m_M be integers with each $m_i \geq 2$, and let the state sequence $\tilde{S}_{t,n}$ take on values in $\{1, 2, \dots, \sum_{i=1}^M m_i\}$. Let \mathbf{B} be the $\sum_{i=1}^M m_i \times \sum_{i=1}^M m_i$ transition probability matrix for $\tilde{S}_{t,n}$, such that it is composed of submatrices, namely,

$$\mathbf{B} = \begin{pmatrix} \mathbf{B}_{11} & \cdots & \mathbf{B}_{1M} \\ \vdots & \ddots & \vdots \\ \mathbf{B}_{M1} & \cdots & \mathbf{B}_{MM} \end{pmatrix},$$

where \mathbf{B}_{ii} are $m_i \times m_i$ matrices and \mathbf{B}_{ij} are $m_i \times m_j$ matrices. These submatrices have the form

$$\mathbf{B}_{ii} = \left(\begin{array}{c|cccc} 0 & 1 - c_i(1) & 0 & \cdots & 0 \\ \vdots & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ \hline 0 & 0 & \cdots & 0 & 1 - c_i(m_i - 1) \\ 0 & 0 & \cdots & 0 & 1 - c_i(m_i) \end{array} \right), \text{ and for } i \neq j, \mathbf{B}_{ij} = \begin{pmatrix} a_{ij}c_i(1) & 0 & \cdots & 0 \\ a_{ij}c_i(2) & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & 0 \\ a_{ij}c_i(m_i) & 0 & \cdots & 0 \end{pmatrix},$$

where $c_i(r) = p_i(r)/(1 - F_i(r - 1))$ for $F_i(r - 1) < 1$ and $c_i(r) = 1$ for $F_i(r - 1) = 1$, and $p_i(\cdot)$ is the probability mass function with cumulative distribution function $F_i(\cdot)$ for the dwell-time distribution of state i in the HSMM, with both functions potentially indexed by a parameter vector β_i that may be associated with the covariates in \mathbf{Z} . For $i = 1, \dots, M$, let $\mathcal{I}_i = \{k : \sum_{j=0}^{i-1} m_j < k \leq \sum_{j=0}^i m_j\}$, where $m_0 = 0$, be the i -th *state aggregate*. \mathcal{I}_i is a collection of states in $\tilde{S}_{t,n}$ for the approximate HSMM, and is constructed such that \mathcal{I}_i corresponds to state i for the state sequences $\{S_{t,n}\}$ in the HSMM. Note the deliberate construction of the transition probability matrix \mathbf{B} . The properties of the M states for $S_{t,n}$ in the HSMM are approximated by the submatrices of \mathbf{B} . The diagonal block \mathbf{B}_{ii} represents transitions within

state aggregate i in the approximate HSMM, and this corresponds to the dwell-time in the HSMM. The off-diagonal block $\mathbf{B}_{ij}, i \neq j$, characterizes the transitions between state aggregates in the approximate HSMM, corresponding to the transitions between states in the HSMM. Note that a transition from state aggregate \mathcal{I}_i to state aggregate \mathcal{I}_j in the approximate HSMM must go to the state with the smallest index in \mathcal{I}_j .

We now give a few remarks about the approximate HSMM being in an HMM framework, and its relation to the HSMM, which Langrock and Zucchini (2011) discussed in greater detail:

- All entries of \mathbf{B} lie in the interval $[0, 1]$, and its row sums equal to 1, and so \mathbf{B} is a valid transition probability matrix.
- One can show that the transition from \mathcal{I}_i to a different state aggregate \mathcal{I}_j in the approximate HSMM has the same probability of state transition governed by the transition probability matrix \mathbf{A} in the HSMM.
- Different dwell-time distributions will lead to different formulations for $c_i(r)$, and thus different entries for each submatrix in \mathbf{B} . However, the deliberate structure for \mathbf{B} will remain the same.
- Since the states within \mathcal{I}_i in the approximate HSMM correspond to state i in the HSMM, the parameters of the emission distribution for all states within \mathcal{I}_i must be constrained to be the same.

The cardinality of \mathcal{I}_i is m_i , and this is a parameter that is set *a priori* for each state aggregate. Langrock and Zucchini (2011) showed that the errors in the approximation is in the right tail of the dwell-time distribution, and that a larger value of m_i leads to a better approximation of the dwell-time distribution. In Figure 2, we show how the choice of m_i affects the approximation of the dwell-time distribution. In the example in the figure, suppose the dwell-time distribution for one of the states in the HSMM is the shifted Poisson distribution with shift parameter 1 and rate parameter λ , and suppose the approximate HSMM uses the correctly specified dwell-time distribution. The rows in Figure 2 correspond to dwell-time distributions with different rate parameters, where in the top row we have $\lambda = 3$ and the bottom row we have $\lambda = 10$. The three columns in Figure 2 correspond to the different sizes of the state aggregates, namely, $m_i = 5, 10$, or 15 . The vertical lines correspond to the actual values of the dwell-time distribution, and the dots correspond to the values of the dwell-time distribution induced by the approximation that arises for each m_i . In the top row, we see that setting $m_i = 10$ or 15 approximates the dwell-time distribution really well, but setting $m_i = 5$ leads to errors in the right tail of the distribution. When $\lambda = 3$ we see that the error in the approximation when using $m_i = 5$ is in the right-tail of the distribution, but using $m_i = 10$ or 15 leads to good approximations. When $\lambda = 10$ we see that the errors are very pronounced for $m_i = 5$, but the approximation is much better for larger values of m_i . In fact, we see that the larger the mean for the

dwell-time distribution the larger we need to set m_i . Altogether, m_i is a parameter where larger values lead to better approximations, with the trade-off that larger values of m_i will increase the dimensions of \mathbf{B} and hence will increase the computational cost.

[Figure 2 about here.]

The above describes how we can approximate the HSMM with an HMM using state aggregates and a transition probability matrix with a specific structure. The use of state aggregates allows us to approximate dwell-time distributions within a state aggregate such that dwell-times are not necessarily geometric, and since the approximate HSMM is itself an HMM constructed in a special way we therefore inherit the theoretical and methodological properties and benefits of an HMM. Langrock and Zucchini (2011) developed the above strategy for univariate time series data, and we show using synthetic and empirical data that this strategy also works well for multivariate time series.

2.3 Estimation

Our use of an HMM to approximate an HSMM allows us to use the approaches developed for HMMs to estimate the parameters, being mindful of constraints, namely, the deliberate form of the transition probability matrix and the emission distribution parameters being the same within a state aggregate. In this work, we use the Expectation-Maximization (EM) algorithm (Dempster et al., 1977; Zucchini et al., 2017). We provide an overview of the algorithm here, adopting the notation by Zucchini et al. (2017).

If we know the true state sequences $S_{t,n}$ for each subject n , then we could optimize the complete likelihood given the observed data $\mathbf{y}_1, \dots, \mathbf{y}_n$ and the known state sequences $S_{t,1}, \dots, S_{t,N}$. For $t = 1, \dots, T$ and $n = 1, \dots, N$, let $u_{i,n}(t) = 1$ if $S_{t,n} = i$ and 0 otherwise, and for $t = 2, \dots, T$ let $v_{i,j,n}(t) = 1$ if both $c_{t-1,n} = i$ and $c_{t,n} = j$ and 0 otherwise. Let $\tilde{M} = \sum_{i=1}^M m_i$, and let $b_{ij} = b_{ij}(\boldsymbol{\beta})$ denote the (i, j) -th entry of the matrix $\mathbf{B} = \mathbf{B}(\boldsymbol{\beta})$, where the parameter $\boldsymbol{\beta}$ contains the parameters for the dwell-time distributions. Given the observed data and known state sequences, the complete data log-likelihood is

$$\begin{aligned} \ell(\Omega | \mathbf{y}_1, \dots, \mathbf{y}_n, S_{t,1}, \dots, S_{t,N}) &= \sum_{n=1}^N \sum_{i=1}^{\tilde{M}} u_{i,n}(1) \log \delta_{i,n} + \sum_{n=1}^N \sum_{i=1}^{\tilde{M}} \sum_{j=1}^{\tilde{M}} \left(\sum_{t=2}^T v_{i,j,n}(t) \right) \log b_{ij}(\boldsymbol{\beta}) \\ &+ \sum_{n=1}^N \sum_{i=1}^{\tilde{M}} \sum_{t=1}^T u_i(t) \log f_i(\mathbf{y}_n; \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i), \end{aligned} \quad (1)$$

where Ω is the set of all parameters and f_i is the probability density function for the P -variate normal distribution with mean vector $\boldsymbol{\mu}_i$ and covariance matrix $\boldsymbol{\Sigma}_i$. The EM algorithm iterates between the E step and the M step for optimizing Equation (1). The goal for the E step is to calculate the conditional

expectation of $\{u_{i,n}(t)\}_{i,n}$ and $\{v_{i,j,n}(t)\}_{i,j,n}$ given an estimate for Ω . On the other hand, the goal for the M step is to optimize Equation (1) replacing $\{u_{i,n}(t)\}_{i,n}$ and $\{v_{i,j,n}(t)\}_{i,j,n}$ with their conditional expectations obtained from the E step. The EM algorithm alternates between these two steps until a convergence criterion is reached, which, in our case, is when there is a small relative change in the log-likelihood.

The E step remains the same as in standard HMMs, and so in the following we focus only on the M step. To update the transition probabilities in the M step we optimize the second term in Equation (1). There are two sets of parameters in this term: the parameter vector $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_M)$ related to the dwell-time distribution and the transition probabilities $\mathbf{A} = (a_{ij})$. In this work for the dwell-time distribution we use a shifted Poisson distribution with shift parameter set to 1 and rate parameter λ_i , which we relate to the covariates using the model $\log(\lambda_i) = \mathbf{Z}\boldsymbol{\beta}_i$, though other discrete distributions for the dwell-time distributions are possible. The shift parameter can be estimated, but we fixed that parameter here for simplicity. Thus, dwell-times potentially vary across states and could relate to covariates in different ways. To update $\boldsymbol{\beta}$, we numerically optimize the second term of Equation (1) using the `optim` function in R. Next, we use a closed-form solution to update \mathbf{A} using $\{v_{i,j,n}(t)\}$ that was updated in the previous E step, namely, $\hat{a}_{ij} = \nu_{ij} / \sum_{k=1}^M \nu_{ik}$, where $\nu_{ij} = \sum_{n=1}^N \sum_{t=2}^T v_{i,j,n}(t)$. We point out that $\mathbf{B}(\boldsymbol{\beta})$ is a sparse matrix and so a number of its elements should be constrained to 0 and thus do not need to be optimized.

To update the parameters of the emission distribution in the M step we optimize the third term in Equation (1). A closed-form solution for these parameters exists. Since the parameters of the emission distribution must be the same across states within a state aggregate, then for state aggregate $i = 1, \dots, M$, we have

$$\hat{\boldsymbol{\mu}}_i = \frac{1}{N} \sum_{n=1}^N \frac{\sum_{t=1}^T \left(\sum_{j \in \mathcal{I}_i} u_{j,n}(t) \right) \mathbf{y}_{t,n}}{\sum_{j \in \mathcal{I}_i} u_{j,n}(t)}, \quad (2)$$

and

$$\hat{\boldsymbol{\Sigma}}_i = \frac{1}{N} \sum_{n=1}^N \frac{\sum_{t=1}^T \left(\sum_{j \in \mathcal{I}_i} u_{j,n}(t) \right) (\mathbf{y}_{t,n} - \hat{\boldsymbol{\mu}}_i)(\mathbf{y}_{t,n} - \hat{\boldsymbol{\mu}}_i)'}{\sum_{t=1}^T \sum_{j \in \mathcal{I}_i} u_{j,n}(t)}. \quad (3)$$

Recall that larger state aggregates improve the approximation to the dwell-time distribution but at a computational cost. Since the approximate HSMM is in the framework of an HMM, then the algorithmic complexity for the EM algorithm for the approximate HSMM is $\mathcal{O}(\tilde{M}^2 T)$ (Rabiner, 1989). In contrast, for a standard HSMM, the worst-case computational complexity for the EM algorithm is $\mathcal{O}(M^2 T + MT^2)$ (Guédon, 2003). The computational complexity for the approximate HSMM remains linear with respect to the length of the time series even though we increase the dimensionality of the problem through the use of state aggregates. Larger state aggregates may be necessary depending on the context of the problem. In that case, one may want to take advantage of the sparse structure of the transition probability matrix $\mathbf{B}(\boldsymbol{\beta})$

to improve computational speed and for more efficient use of memory (Hadj-Amar et al., 2020).

3 Simulation Study

3.1 Simulation Settings

We conducted a number of simulations to show the performance of the approximate HSMM relative to other approaches under various conditions. In all cases, we simulated zero-mean P -variate time series data, $P \in \{10, 30\}$, from a 3-state HSMM for $N = 100$ independent subjects, each time series having length $T = 500$. The emission distribution was a P -variate zero-mean normal distribution. The covariance matrices were as follows:

- **State 1.** Σ_1 had a first-order autoregressive structure, such that its (i, j) -th element was set to $0.7^{|i-j|}$;
- **State 2.** Σ_2 had a fourth-order autoregressive structure. The (i, j) -th element of the *precision matrix* (i.e., Σ_2^{-1}) was set to $\mathbf{1}(|i-j|=0) + 0.4\mathbf{1}(|i-j|=1) + 0.2\mathbf{1}(|i-j|=2) + 0.2\mathbf{1}(|i-j|=3) + 0.1\mathbf{1}(|i-j|=4)$, where $\mathbf{1}(\cdot)$ is the indicator function that evaluates to 1 if its argument is true and 0 otherwise;
- **State 3.** $\Sigma_3 = (\Phi + \text{Id})^{-1}$, where Id is the $P \times P$ identity matrix and Φ is a random symmetric matrix such that an off-diagonal element was set to $1/P$ with probability 0.5 or 0 with probability 0.5. Thus, the degree of sparsity remained constant as P varied, but the values of these elements were closer to 0 for larger P .

The above covariance matrices were used in the simulation study by Rothman et al. (2008) and Städler and Mukherjee (2013), though we made slight modifications to Σ_3 . We simulated time series from an HSMM with transition probability matrix

$$\mathbf{A} = \begin{pmatrix} 0 & 0.5 & 0.5 \\ 0.3 & 0 & 0.7 \\ 0.7 & 0.3 & 0 \end{pmatrix},$$

using a dwell-time distribution that was either i) a shifted Poisson distribution whose shift and rate parameters for the i -th state $(1, \beta_i)$, respectively, where $\log(\beta) = (2.5, 0.5, 1.5)$, or ii) a shifted negative binomial distribution whose shift, size, and mean parameters were $(1, 10, \beta_i)$, respectively, where $\beta = (5, 10, 15)$. The purpose of the shift parameter is so that the dwell-time within a state is at least 1.

We assessed the performance of three competing models:

- **MHSMM.** We used the implementation of the HSMM in the `mhsmm` package in R (O’Connell et al., 2011). We set this model to have a Poisson distribution with rate parameter initialized to 1 for

each state’s dwell-time distribution, transition probability matrix initialized to have 0.45 along the off-diagonal elements, and emission distribution parameters initialized to the maximum likelihood estimates obtained from segmenting each subject’s time series into three equal-sized segments.

- **HMM.** We fit an HMM using custom code. For the emission distribution, we initialized the mean vectors to the zero vector, and we initialized each state’s covariance matrix to have a compound symmetry structure by drawing the correlation from the Uniform($-1/P, 1/P$) distribution and drawing the common standard deviation from the Uniform($1, \sqrt{3}$) distribution. To ensure positive-definiteness of each state’s covariance matrix, we added 0.1 to the diagonal elements. We initialized the transition probability matrix to have 0.8 along the diagonals and 0.1 in the off-diagonals. We ran the EM algorithm and declared convergence when the relative change in the log-likelihood was $< 10^{-5}$. We repeated this procedure 20 times, yielding 20 HMM fits, and we kept the fit that yielded the largest log-likelihood.
- **Approximate HSMM.** We fit the approximate HSMM described in Section 2. We set the size of the state aggregate corresponding to state i to $m_i = m \in \{5, 10, 15\}$. We initialized the parameters of the emission distribution and the transition probability matrix to the estimates obtained by the HMM. We initialized the parameters of the dwell-time distribution (a shifted Poisson distribution with shift parameter 1 and rate parameter $\log(\lambda_i) = \beta_i$) to a random draw from the normal distribution with mean 1.0 and standard deviation 0.5. Just as in the HMM, we used a tolerance criterion of 10^{-5} for the relative change in the log-likelihood for the EM algorithm, and after 20 random initializations, we kept the fit that yielded the largest log-likelihood.

After fitting each model, we addressed the label-switching problem by relabeling the states such that we minimize the Frobenius norm, denoted as $\|\cdot\|_F$, between the estimated covariance matrices and the true covariance matrices across the three states. To assess model performance, we calculated the Frobenius norm between the estimated covariance matrix and the true covariance matrix for each state, and we report the estimate for $E(\hat{\beta})$. Note that whenever the true dwell-time distribution was the negative binomial distribution, both the MHSMM and the approximate HSMM were therefore deliberately misspecified since their dwell-time distributions were the Poisson distribution. Due to this misspecification, we instead compare $E(\hat{\beta}_i)$ to the log of the center of dwell-time distribution for the i -th state. Finally, after relabeling the states, we reconstructed each subject’s state sequence using the Viterbi algorithm (Rabiner, 1989; Zucchini et al., 2017), and then calculated the misclassification rate averaged across time and subjects. We repeat each simulation study 100 times and averaged our assessments across the 100 simulations.

3.2 Simulation Results

In Table 1 we see how the approximate HSMM compares to the HMM and we also see the effects of different sizes of the state aggregates. The approximate HSMM for any state aggregate size has similar or better performances at estimating each state’s covariance matrix relative to the HMM. It may seem, however, that the approximate HSMM with state aggregate size $m = 15$ performed the worst at estimating the covariance matrix for State 1 in the scenario with $P = 10$ and shifted Poisson dwell-time distribution. Further investigation, however, is that there were some outliers that skewed the result. Indeed, the median error for the HMM was 0.134 and the median errors for the approximate HSMM was 0.061, 0.061, and 0.060 for $m = 5, 10$, and 15, respectively. It is not clear what led to the outliers in performance for $m = 15$. We used the shifted Poisson distribution as the dwell-time distribution, and thus it was correctly specified when the true dwell-time distribution was the shifted Poisson distribution and misspecified when it was the shifted negative binomial distribution. Thus, $E(\hat{\beta}_i)$ is an estimate of the mean of the dwell-time distribution for state i . When the true dwell-time distribution was the shifted Poisson distribution, the approximate HSMM generally estimated the true β well except for State 1, where we see a slight downward bias when $m = 10$ and a larger downward bias when $m = 5$. For State 1, the dwell-time distribution was truly centered at $\exp(2.5)$, and so the state aggregates were too small, similar to the example we showed in Figure 5. We draw similar conclusions when the true dwell-time distribution was the shifted negative binomial distribution. Note that due to the misspecification of the dwell-time distribution, we compare $\hat{\beta}$ to the log of the true means, $(\log(5), \log(10), \log(15))$. As before, we see the approximate HSMM underestimating the true mean of the dwell-time distribution for States 2 and 3 for $m = 5$ and State 3 for $m = 10$. While larger values of m led to a reduction in bias in estimating the mean of the dwell-time distribution, we also see that larger values of m also led to greater variability in the estimates. Finally, we see that the misclassification rate for the approximate HSMM were comparable over the different values of m . In all cases, the approximate HSMM for any state aggregate size had a similar or better performance than the HMM with respect to misclassification. The low misclassification rate for all methods under all scenarios is likely because of how distinct the covariance matrices were across the three states.

[Table 1 about here.]

In Table 2 we compare the performances of the MHSMM, the HMM, and the approximate HSMM. First, we point out that in some instances for MHSMM, the estimates of β diverged to $-\infty$. We removed these instances when tabulating results, and so the results for MHSMM reported here are more optimistic than they truly are. With respect to estimating each state’s covariance matrix, as we saw before generally the

approximate HSMM performed as well as or improved on the HMM. We also see that the MHSMM and the approximate HSMM have similar performances in estimating each state’s covariance matrix. However, when it comes to estimating the mean of the dwell-time distributions, we see that the MHSMM has mixed results. Indeed, the MHSMM had good performances on average, but the estimates could have higher variance or could diverge, as stated earlier. Finally, the misclassification performance of the HMM and the approximate HSMM were either as good as or better than the misclassification performance of the MHSMM.

[Table 2 about here.]

In summary, our simulation study shows that the approximate HSMM performs well at estimating the covariance matrices, and its performance at estimating the mean of the dwell-time distribution depends on the state aggregates. We initialized the emission distribution parameters of the approximate HSMM to the estimates of the emission distribution parameters of the HMM, which explains why the approximate HSMM improved on the HMM. We believe that this is good practice given how fast one can fit the HMM. Finally, while the approximate HSMM seems to have greater bias than the MHSMM in estimating the mean of the dwell-time distribution, the estimates of the MHSMM can be relatively unstable.

4 Applications to Dynamic Connectivity Analysis

4.1 Data Description

The data came from the BRain Imaging Development of Girls’ Emotion and Self (BRIDGES) Study at the University of Minnesota. A central goal of the BRIDGES Study is to understand how the development of brain systems may be aberrant in adolescents with non-suicidal self-injury (NSSI). This longitudinal study recruited adolescent females 12-16 years of age who exhibited a continuum of NSSI severity which was then classified into 4 categories of NSSI severity: no NSSI, mild NSSI, moderate NSSI, or severe NSSI. The adolescents were enrolled to participate in three annual evaluations that involved clinical, physiologic and neuroimaging assessments. The current study focused on the baseline neuroimaging assessment, and considered two clinical groups for comparison: adolescents with moderate or severe NSSI versus those with no or mild NSSI.

Brain scanning sessions were conducted at the Center for Magnetic Resonance Research at the University of Minnesota. The resting-state functional magnetic resonance imaging (fMRI) data consisted of a 12-minute scan during which each participant was instructed to stay awake, keep their eyes open focused on a fixation cross, and to “not think about anything in particular”. These fMRI scans consisted of whole brain T2*-weighted functional volumes with 2 mm isotropic voxel resolution, with the following fMRI parameters:

TR = 800ms, TE = 37 ms, flip angle = 52°, FOV = 212 mm, 2 mm isotropic voxel, Multiband factor=8. All functional data were acquired using the Human Connectome Project multiband echo planar imaging sequence.

Group level spatial ICA was utilized to estimate intrinsic connectivity networks (ICNs) using the GIFT toolbox (<https://trendscenter.org/software/gift/>) (Calhoun et al., 2001). Subject level time courses were reduced to 110 principal components and concatenated along the time dimension. The concatenated time courses were reduced to 100 principal components through a group level PCA. Group level ICN's were estimated using the infomax algorithm to optimize temporal independence (Calhoun et al., 2001). Calculations on group level ICN's were repeated 20 times using the ICASSO technique for reliability (Himberg and Hyvarinen, 2003). The spatial maps from the infomax ICA output were used as spatial templates for Group Information Guided ICA (GIG-ICA). The GIG-ICA calculations estimated 100 group level ICN's also optimizing the independence at the subject level (Du and Fan, 2013; Du et al., 2020). 35 ICN spatial maps were selected to be binarized, setting the bottom 45% nonzero absolute voxel intensities to zero, and evaluated for mutual overlap with ROIs from the Yeo 17 network and the Harvard-Oxford cortical and subcortical structural atlases (<https://identifiers.org/neurovault.collection:262>) (Yeo et al., 2011). Each ICN was labeled using the atlas region with which it had the most overlap.

The final data used in our analysis was a time series with length $T = 912$, dimension $P = 35$, observed from $N = 100$ subjects, and with each time series having zero mean and unit variance. Of these 100 subjects, 61 were classified to have moderate or severe nonsuicidal self-injury (NSSI), and 39 were classified to have no or mild NSSI.

4.2 Overview of the Statistical Procedure

Vidaurre et al. (2017), Vidaurre et al. (2018), and Shappell et al. (2019) each fit a 12-state model to their resting-state fMRI data, and so we selected the same model to reproduce their results. We set the covariate matrix \mathbf{Z} to have two columns, where the first column is an indicator for moderate or severe NSSI and the second column is an indicator for no or mild NSSI. For the dwell-time distribution for the i -th state we used a shifted Poisson distribution with shift parameter set to 1 and rate parameter $\log(\lambda_i) = \mathbf{Z}\boldsymbol{\beta}_i$, where $\boldsymbol{\beta}_i = (\beta_{i1}, \beta_{i2})'$ captures the relationship between NSSI severity and dwell-time. For the i -th state, we set $m_i = 20$. Thus, the approximate HSMM had a total of 240 states.

We now describe how we fit the approximate HSMM. As in our simulation study, we first fit a standard HMM with 12 states using the EM algorithm. We used the same strategy as in our simulation study for the random initializations, so we refer to Section 3 for details. We ran the EM algorithm and declared

convergence when the relative change in the log-likelihood was $< 10^{-6}$. After obtaining fits from 50 random initializations, we kept the fit that yielded the largest log-likelihood.

We used the estimates of the parameters of the emission distribution for the HMM and its transition probability matrix (setting its diagonal entries to 0 and then standardizing the off-diagonals to have a unit row sum) to initialize the parameters for the EM algorithm to fit the approximate HSMM. To initialize the dwell-time distribution β , we randomly drew each coefficient from the normal distribution with mean 2.5 and standard deviation 0.5, so that the resulting prior dwell-time for each state would be roughly between 1 and 32.5 TR units (or between 0.8 and 26 seconds). We ran the EM algorithm and declared convergence when the relative change in the log-likelihood was $< 10^{-5}$. We repeated this procedure 50 times, yielding 50 approximate HSMM fits, and we kept the fit that yielded the largest log-likelihood. We then used the `numDeriv` package in R to numerically estimate the Hessian matrix for β to obtain standard errors and to construct Wald test statistics for post-hoc analyses about these coefficients. We used the Viterbi algorithm for state reconstruction for each subject (Zucchini et al., 2017), and then mapped each state sequence over the 240 states back to the 12 state aggregates. Given each subject’s state reconstruction, we computed their state switching frequency, defined as a transition from one state aggregate to another, and their fractional occupancy, defined as the proportion of time spent in a state. Finally, to assess the similarity between the 12 state aggregates, we used hierarchical clustering, using the Kullback-Liebler divergence between covariance matrices as a measure of distance between states.

4.3 Results

Henceforth, in the context of the approximate HSMM we will use the terms “state” and “state aggregate” interchangeably. Figure 3 shows the correlation matrix for each of the 12 states in the approximate HSMM. We see across all 12 states that that the different time series within a region generally maintained strong correlations. However, across the 12 states we see varying degrees of the strength of the correlation between regions. For instance, States 8, 9, and 12 are characterized by strong inter-region correlations, whereas States 3 and 7 are characterized by weak inter-region correlations.

[Figure 3 about here.]

Figure 4(a) shows the similarity between the 12 states. We see that the 12 states can be separated into two *metastates*, in agreement with previous findings (Vidaurre et al., 2017; Shappell et al., 2019), with States 1-7 form Metastate 1, and States 8-12 form Metastate 2. Vidaurre et al. (2017) and Shappell et al. (2019) showed that transitions from one state to another are greater if the two states are within the same metastate. Our results are in partial agreement. Figure 4(b) shows the transition probability matrix. States 2 and 3

tend to transition to each other. State transitions between the states in metastate 1 tend to be to other states within metastate 1, specifically to States 2 or 3, but we see that states within metastate 2 tend to transition to States 1-3 which are in metastate 1.

[Figure 4 about here.]

Table 3 shows the estimates of β across all states, which we then mapped back to the rate parameter of the shifted Poisson distribution as shown in Figure 5. The dwell-time distributions were generally centered around $TR = 10$, corresponding to 8 seconds. Looking at Figure 5, we see that in many of the states the Moderate/Severe group tended to have shorter dwell-times. To test if the dwell-time distributions differed between NSSI severity groups, we tested $H_0 : \beta_{i2} - \beta_{i1} = 0$ for each i , then used the Benjamini-Hochberg procedure to correct for the false discovery rate (FDR) over the 12 comparisons (Benjamini and Hochberg, 1995). The difference was statistically significant (FDR-adjusted p -value < 0.05) for all states except for State 4.

[Table 3 about here.]

[Figure 5 about here.]

As discussed above, the shorter dwell-times for those with Moderate/Severe NSSI suggest a greater number of state switches over the course of the resting-state scan. In Figures 6(a) and 6(b), we see this to be the case. The mean (SD) number of state switches is 102.69 (9.01) in the Moderate/Severe NSSI severity group and 94.08 (9.49) in the No/Mild NSSI severity group (p -value < 0.001 using a two-sample t -test). The mean (SD) number of metastate switches is 56.07 (8.00) in the Moderate/Severe NSSI severity group and 49.21 (8.67) in the No/Mild NSSI severity group. Table 4 shows the fractional occupancy for each of the 12 states and each of the 2 metastates. We see that States 2 and 3 were the most visited states for both groups, with the remaining 10 states being visited somewhat equally. However, the dwell-time parameters shown in Table 3 do not suggest that States 2 and 3 have longer dwell-times relative to the other states. Thus, the larger fractional occupancy is due to the number of transitions into States 2 and 3 as opposed to the longer dwell-times. There is no evidence of a difference in fractional occupancy between the NSSI severity groups for any state or metastate (p -value > 0.05 for all comparisons).

[Figure 6 about here.]

[Table 4 about here.]

5 Discussion

We developed a statistical model for approximating HSMMs using an HMM, and showed its utility in analyzing resting-state fMRI time series. Our analysis showed that dwell-time distributions vary over states, and differ between NSSI severity. Our results are consistent with previous work on evidence of two metastates, but we did not see strong evidence that states tended to transition to other states within the same metastate (Vidaurre et al., 2017; Shappell et al., 2019). It may be the case that the transition probabilities between states is associated with NSSI severity which we did not account for in our model. Indeed, in our analysis NSSI severity was only included in the dwell-time distributions. Previous work has shown that the strength of connectivity and the variability in the temporal dynamics of connectivity between specific regions of the brain can be useful for characterizing depression and suicide ideation (Demirtaş et al., 2016; Kaiser et al., 2016; Jiao et al., 2019; Wise et al., 2017). It is possible for our approximate HSMM model to focus on the dynamic connectivity between specific region pairs by fitting a collection of bivariate models. Our findings show that female adolescents with moderate or severe NSSI severity have more state switches, suggesting greater temporal variability in the correlations. Although no prior studies have examined dynamic connectivity in relation to NSSI, a small number of prior studies utilized different techniques to study resting-state connectivity in patients with depression. Different directions of association between temporal variability and connectivity have been previously reported, but these studies used different analytic approaches (e.g., sliding-window correlations, cluster analysis, graph theory), and time series data from different regions of the brain (Demirtaş et al., 2016; Kaiser et al., 2016; Jiao et al., 2019; Wise et al., 2017; Zheng et al., 2018; Zhi et al., 2018). Notably, these prior studies have been primarily conducted on adults, and no prior studies have applied HMMs or HSMMs to examine state duration and transition frequency in depressed patients. Furthermore, while depression and NSSI are related constructs in the sense that NSSI usually occurs in the context of negative affect (Klonsky, 2011) and the two problems commonly co-occur, growing evidence suggests that depression and NSSI might have both overlapping and distinct underlying biological patterns (Klimes-Dougan et al., 2019). Finally, we showed that fractional occupancy was the greatest for State 3. We point out that State 3 is characterized by weak correlations averaged across all the off-diagonal entries of its correlation matrix (mean correlation = 0.173), with only State 7 having weaker correlations (mean correlation = 0.145), although Marusak et al. (2017) showed that healthy children spend more time in a state characterized by weak regional connections.

Our approximate HSMM has natural extensions that can provide the practitioner with further flexibility. In our application, we accounted for the NSSI Severity group in the dwell-time distribution, but our use of

the HMM framework means we could model the parameters of the emission distribution or in the transition probability between state aggregates (Zucchini et al., 2017). However, we may encounter challenges with respect to updating the parameters in the M-step of the EM algorithm. Furthermore, we restricted our analysis to $P = 35$ dimensions. One could consider a model with higher dimensions, though one may need to include some form of regularization in the model. While there have been theoretical and methodological developments for HMMs for high-dimensional time series (Fiecas et al., 2017b; Städler and Mukherjee, 2013), to our knowledge there have been no developments for HSMMs. Since our approximate HSMM is in the HMM framework, one could therefore utilize the developments for HMMs for modeling high-dimensional time series data. Another potential extension is to account for the autocorrelation of the data, which could improve estimates of and statistical inference on functional connectivity (Fiecas et al., 2017a; Hart et al., 2018). One possibility is to use a switching vector autoregressive (VAR) model, which uses a stationary VAR conditional on the state sequence (Ting et al., 2017; Ombao et al., 2018). Finally, our analysis only used the baseline data. One could conduct a longitudinal dynamic connectivity analyses by incorporating random effects into the model. There have been developments for HMMs with random effects (Altman, 2007; DeRuiter et al., 2017), and one could potentially adapt these developments to the approximate HSMM.

We now describe some limitations with our modeling approach. In our analyses, we assumed that the number of states was known *a priori*. This will not always be the case in practice, and so one would need to use, e.g., the Akaike or Bayesian information criterion (AIC or BIC, respectively) or other model comparison metrics to select the number of states (Zucchini et al., 2017). In our approximate HSMM, we also have the size of the state aggregates as a tuning parameter, and we showed that this can affect performance if it is too small with respect to the true underlying dwell-time distribution. In our data analysis, we set the size $m_i = 20$ for all state aggregates and so our approximation of dwell-time distributions will be accurate for up to 16 seconds. We saw in our analysis that dwell times were generally centered at $TR = 10$, corresponding to 8 seconds, and thus the right-tail of the dwell-time distribution was likely not affected by our choice for m_i . Indeed, the flexibility of the HSMM means that there one would need to make more decisions about the model and tune more parameters. There are a number of open problems and questions surrounding dynamic connectivity (Lurie et al., 2020). For instance, it is important to demonstrate if the summary measures extracted from our approximate HSMM (e.g., state-specific correlations, state reconstructions) are robust and reproducible. We showed the performance of the approximate HSMM using synthetic data, but empirical evidence will be more useful. The literature is mixed on the test-retest reliability of summary measures extracted from dynamic connectivity analyses (Choe et al., 2017; Liao et al., 2017). On the other hand, Vidaurre et al. (2018) showed that HMMs yielded reproducible results in modeling brain dynamics. Finally, it is important to be clear on what we mean when we say that connectivity is “dynamic”, since this

can influence the appropriate “null model” to use to test for the presence of such dynamics and can also affect the interpretation of results (Liégeois et al., 2017; Lurie et al., 2020). In our case, the approximate HSMM, like the HMM, assumes that the correlations vary over time conditional on the state sequence. However, we point out that unconditional on the state sequence the first and second-order moments do not vary over time.

Altogether, we showed using synthetic data and resting-state fMRI data that the approximate HSMM provides an excellent modeling framework for conducting dynamic connectivity analyses in resting-state fMRI. The theoretical and methodological foundations of the approximate HSMM are based on those already established for HMMs, and thus this framework has the flexibility to be extended or adapted based on the needs of the practitioner. Finally, the model is fairly general, and even though we used this model to analyze resting-state fMRI data, it can be adapted to analyze multivariate time series data from other scientific fields.

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(3):663–676.
- Altman, R. M. (2007). Mixed hidden Markov models: an extension of the hidden Markov model to the longitudinal data setting. *Journal of the American Statistical Association*, 102(477):201–210.
- Baker, A. P., Brookes, M. J., Rezek, I. A., Smith, S. M., Behrens, T., Smith, P. J. P., and Woolrich, M. (2014). Fast transient networks in spontaneous human brain activity. *Elife*, 3:e01867.
- Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1):289–300.
- Bulla, J. and Bulla, I. (2006). Stylized facts of financial time series and hidden semi-Markov models. *Computational Statistics & Data Analysis*, 51(4):2192–2209.
- Bulla, J., Bulla, I., and Nenadić, O. (2010). hsmm—an R package for analyzing hidden semi-Markov models. *Computational Statistics & Data Analysis*, 54(3):611–619.
- Calhoun, V. D., Adali, T., Pearlson, G. D., and Pekar, J. J. (2001). A method for making group inferences from functional MRI data using independent component analysis. *Human Brain Mapping*, 14(3):140–151.

- Calhoun, V. D., Miller, R., Pearlson, G., and Adalı, T. (2014). The chronnectome: time-varying connectivity networks as the next frontier in fMRI data discovery. *Neuron*, 84(2):262–274.
- Choe, A. S., Nebel, M. B., Barber, A. D., Cohen, J. R., Xu, Y., Pekar, J. J., Caffo, B., and Lindquist, M. A. (2017). Comparing test-retest reliability of dynamic functional connectivity methods. *NeuroImage*, 158:155–175.
- Cribben, I., Haraldsdottir, R., Atlas, L. Y., Wager, T. D., and Lindquist, M. A. (2012). Dynamic connectivity regression: determining state-related changes in brain connectivity. *NeuroImage*, 61(4):907–920.
- Demirtaş, M., Tornador, C., Falcón, C., López-Solà, M., Hernández-Ribas, R., Pujol, J., Menchon, J. M., Ritter, P., Cardoner, N., Soriano-Mas, C., et al. (2016). Dynamic functional connectivity reveals altered variability in functional connectivity among patients with major depressive disorder. *Human Brain Mapping*, 37(8):2918–2930.
- Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22.
- DeRuiter, S. L., Langrock, R., Skirbutas, T., Goldbogen, J. A., Calambokidis, J., Friedlaender, A. S., Southall, B. L., et al. (2017). A multivariate mixed hidden Markov model for blue whale behaviour and responses to sound exposure. *The Annals of Applied Statistics*, 11(1):362–392.
- Du, Y. and Fan, Y. (2013). Group information guided ICA for fMRI data analysis. *NeuroImage*, 69:157–197.
- Du, Y., Fu, Z., Sui, J., Gao, S., Xing, Y., Lin, D., Salman, M., Abrol, A., Rahaman, M. A., Chen, J., Hong, L. E., Kochunov, P., Osuch, E. A., and Calhoun, V. D. (2020). Neuromark: An automated and adaptive ica based pipeline to identify reproducible fMRI markers of brain disorders. *NeuroImage: Clinical*, 28:102375.
- Fiecas, M., Cribben, I., Bahktiari, R., and Cummine, J. (2017a). A variance components model for statistical inference on functional connectivity networks. *NeuroImage*, 149:256–266.
- Fiecas, M., Franke, J., von Sachs, R., and Tadjuidje Kamgaing, J. (2017b). Shrinkage estimation for multivariate hidden Markov models. *Journal of the American Statistical Association*, 112(517):424–435.
- Guédon, Y. (2003). Estimating hidden semi-Markov chains from discrete sequences. *Journal of Computational and Graphical Statistics*, 12(3):604–639.
- Hadj-Amar, B., Jewson, J., and Fiecas, M. (2020). Bayesian Approximations to Hidden Semi-Markov Models. *arXiv preprint arXiv:2006.09061*.

- Hart, B., Cribben, I., and Fiecas, M. (2018). A longitudinal model for functional connectivity networks using resting-state fMRI. *NeuroImage*, 178:687–701.
- Himberg, J. and Hyvarinen, A. (2003). Icasso: software for investigating the reliability of ica estimates by clustering and visualization. In *2003 IEEE XIII Workshop on Neural Networks for Signal Processing (IEEE Cat. No. 03TH8718)*, pages 259–268. IEEE.
- Jiao, L., Duan, X., Cui, Q., Chen, H., and Liao, W. (2019). More than just statics: temporal dynamics of intrinsic brain activity predicts the suicidal ideation in depressed patients. *Psychological Medicine*, 49(5):852–860.
- Kaiser, R. H., Whitfield-Gabrieli, S., Dillon, D. G., Goer, F., Beltzer, M., Minkel, J., Smoski, M., Dichter, G., and Pizzagalli, D. A. (2016). Dynamic resting-state functional connectivity in major depression. *Neuropsychopharmacology*, 41(7):1822–1830.
- Klimes-Dougan, B., Begnel, E., Almy, B., Thai, M., Schreiner, M. W., and Cullen, K. R. (2019). Hypothalamic-pituitary-adrenal axis dysregulation in depressed adolescents with non-suicidal self-injury. *Psychoneuroendocrinology*, 102:216–224.
- Klonsky, E. D. (2011). Non-suicidal self-injury in united states adults: prevalence, sociodemographics, topography and functions. *Psychological medicine*, 41(9):1981.
- Langrock, R., King, R., Matthiopoulos, J., Thomas, L., Fortin, D., and Morales, J. M. (2012). Flexible and practical modeling of animal telemetry data: hidden Markov models and extensions. *Ecology*, 93(11):2336–2342.
- Langrock, R. and Zucchini, W. (2011). Hidden Markov models with arbitrary state dwell-time distributions. *Computational Statistics & Data Analysis*, 55(1):715–724.
- Liao, X., Cao, M., Xia, M., and He, Y. (2017). Individual differences and time-varying features of modular brain architecture. *NeuroImage*, 152:94–107.
- Liégeois, R., Laumann, T. O., Snyder, A. Z., Zhou, J., and Yeo, B. T. (2017). Interpreting temporal fluctuations in resting-state functional connectivity MRI. *NeuroImage*, 163:437–455.
- Liégeois, R., Li, J., Kong, R., Orban, C., Van De Ville, D., Ge, T., Sabuncu, M. R., and Yeo, B. T. (2019). Resting brain dynamics at different timescales capture distinct aspects of human behavior. *Nature Communications*, 10(1):1–9.

- 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.
- Lurie, D. J., Kessler, D., Bassett, D. S., Betzel, R. F., Breakspear, M., Kheilholz, S., Kucyi, A., Liégeois, R., Lindquist, M. A., McIntosh, A. R., et al. (2020). Questions and controversies in the study of time-varying functional connectivity in resting fMRI. *Network Neuroscience*, 4(1):30–69.
- Marusak, H. A., Calhoun, V. D., Brown, S., Crespo, L. M., Sala-Hamrick, K., Gotlib, I. H., and Thomason, M. E. (2017). Dynamic functional connectivity of neurocognitive networks in children. *Human Brain Mapping*, 38(1):97–108.
- Obermaier, B., Guger, C., Neuper, C., and Pfurtscheller, G. (2001). Hidden Markov models for online classification of single trial EEG data. *Pattern Recognition Letters*, 22(12):1299–1309.
- Ombao, H., Fiecas, M., Ting, C.-M., and Low, Y. F. (2018). Statistical models for brain signals with properties that evolve across trials. *NeuroImage*, 180:609–618.
- O’Connell, J., Højsgaard, S., et al. (2011). Hidden semi-Markov models for multiple observation sequences: The mhsmm package for R. *Journal of Statistical Software*, 39(4):1–22.
- Quinn, A. J., Vidaurre, D., Abeyesuriya, 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.
- Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286.
- Rashid, B., Arbabshirani, M. R., Damaraju, E., Cetin, M. S., Miller, R., Pearlson, G. D., and Calhoun, V. D. (2016). Classification of schizophrenia and bipolar patients using static and dynamic resting-state fMRI brain connectivity. *NeuroImage*, 134:645–657.
- Rothman, A. J., Bickel, P. J., Levina, E., Zhu, J., et al. (2008). Sparse permutation invariant covariance estimation. *Electronic Journal of Statistics*, 2:494–515.
- Shappell, H., Caffo, B. S., Pekar, J. J., and Lindquist, M. A. (2019). Improved state change estimation in dynamic functional connectivity using hidden semi-Markov models. *NeuroImage*, 191:243–257.
- Shappell, H. M., Duffy, K. A., Rosch, K. S., Pekar, J. J., Mostofsky, S. H., Lindquist, M. A., and Cohen, J. R. (2021). Children with attention-deficit/hyperactivity disorder spend more time in hyperconnected network states and less time in segregated network states as revealed by dynamic connectivity analysis. *NeuroImage*, 229:117753.

- Städler, N. and Mukherjee, S. (2013). Penalized estimation in high-dimensional hidden Markov models with state-specific graphical models. *The Annals of Applied Statistics*, pages 2157–2179.
- Taghia, J., Ryali, S., Chen, T., Supekar, K., Cai, W., and Menon, V. (2017). Bayesian switching factor analysis for estimating time-varying functional connectivity in fMRI. *NeuroImage*, 155:271–290.
- Ting, C.-M., Ombao, H., Samdin, S. B., and Salleh, S.-H. (2017). Estimating dynamic connectivity states in fMRI using regime-switching factor models. *IEEE Transactions on Medical Imaging*, 37(4):1011–1023.
- Vidaurre, D., Abeyesuriya, 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.
- 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(48):12827–12832.
- 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(521):134–151.
- Wise, T., Marwood, L., Perkins, A., Herane-Vives, A., Joules, R., Lythgoe, D., Luh, W., Williams, S., Young, A., Cleare, A., et al. (2017). Instability of default mode network connectivity in major depression: a two-sample confirmation study. *Translational Psychiatry*, 7(4):e1105–e1105.
- Yeo, B. T. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., Roffman, J. L., Smoller, J. W., Zöllei, L., Polimeni, J. R., Fisch, B., Liu, H., and Buckner, R. L. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology*, 106(3):1125–1165.
- Zheng, H., Li, F., Bo, Q., Li, X., Yao, L., Yao, Z., Wang, C., and Wu, X. (2018). The dynamic characteristics of the anterior cingulate cortex in resting-state fMRI of patients with depression. *Journal of Affective Disorders*, 227:391–397.
- Zhi, D., Calhoun, V. D., Lv, L., Ma, X., Ke, Q., Fu, Z., Du, Y., Yang, Y., Yang, X., Pan, M., et al. (2018). Aberrant dynamic functional network connectivity and graph properties in major depressive disorder. *Frontiers in Psychiatry*, 9:339.
- Zucchini, W., MacDonald, I. L., and Langrock, R. (2017). *Hidden Markov models for time series: an introduction using R*. CRC press.

List of Figures

1	A graphical model representation for a Hidden Markov Model (HMM; left) and a Hidden semiMarkov Model (HSMM; right) for the n -th subject. White circles correspond to unobserved variables, and shaded circles correspond to the observed data. For both models, $S_{t,n}$ is the hidden state sequence that give rise to the observed data $\mathbf{y}_{1,n}, \dots, \mathbf{y}_{T,n}$. For the HSMM, $d_{1,n}(\boldsymbol{\beta}), \dots, d_{\tau,n}(\boldsymbol{\beta})$ are random dwell-times associated with each state that come from a dwell-time distribution with parameter $\boldsymbol{\beta}$	24
2	In each plot, we have the shifted Poisson distribution with shift parameter 1 and rate parameter $\lambda = 3$ (top row) and $\lambda = 10$ (bottom row). The columns correspond to the different sizes m_i of the state aggregates. The vertical lines in each plot shows the probability mass function of the shifted Poisson distribution, and the dots correspond to the induced dwell-time distribution of the approximated HSMM for a given m_i	25
3	The correlation matrix for each state, with upper and lower triangles corresponding to different states. The black squares within the plot correspond to the ICNs mapped to the same Yeo regions.	26
4	(a) Dendrogram for the covariance matrices across the 12 states. (b) Transition probability matrix for the approximate HSMM. States 1-7 and 8-12 correspond to metastates 1 and 2, respectively. These states are outlined in the black squares.	27
5	Dwell-time distributions for each state, stratified by NSSI severity group.	28
6	Boxplots of (a) state-switching frequency and (b) metastate-switching frequency per NSSI severity group.	29

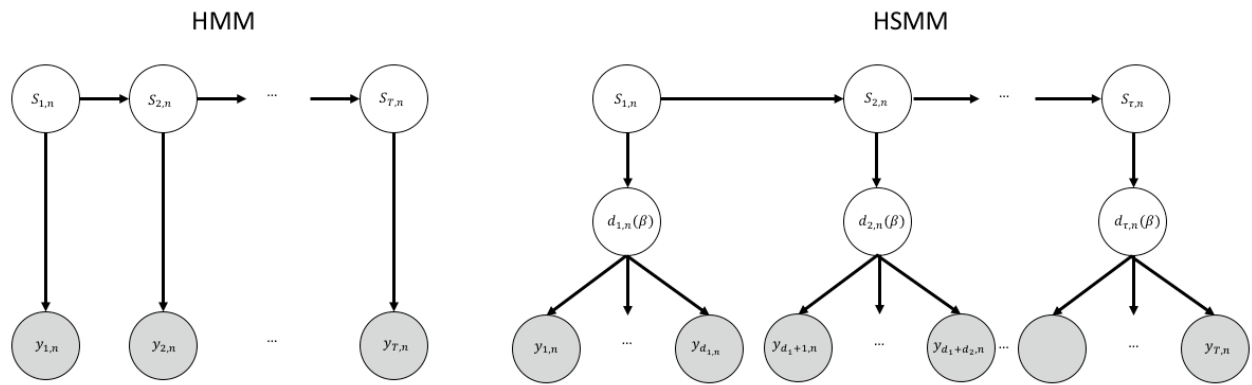


Figure 1: A graphical model representation for a Hidden Markov Model (HMM; left) and a Hidden semi-Markov Model (HSMM; right) for the n -th subject. White circles correspond to unobserved variables, and shaded circles correspond to the observed data. For both models, $S_{t,n}$ is the hidden state sequence that give rise to the observed data $\mathbf{y}_{1,n}, \dots, \mathbf{y}_{T,n}$. For the HSMM, $d_{1,n}(\beta), \dots, d_{T,n}(\beta)$ are random dwell-times associated with each state that come from a dwell-time distribution with parameter β .

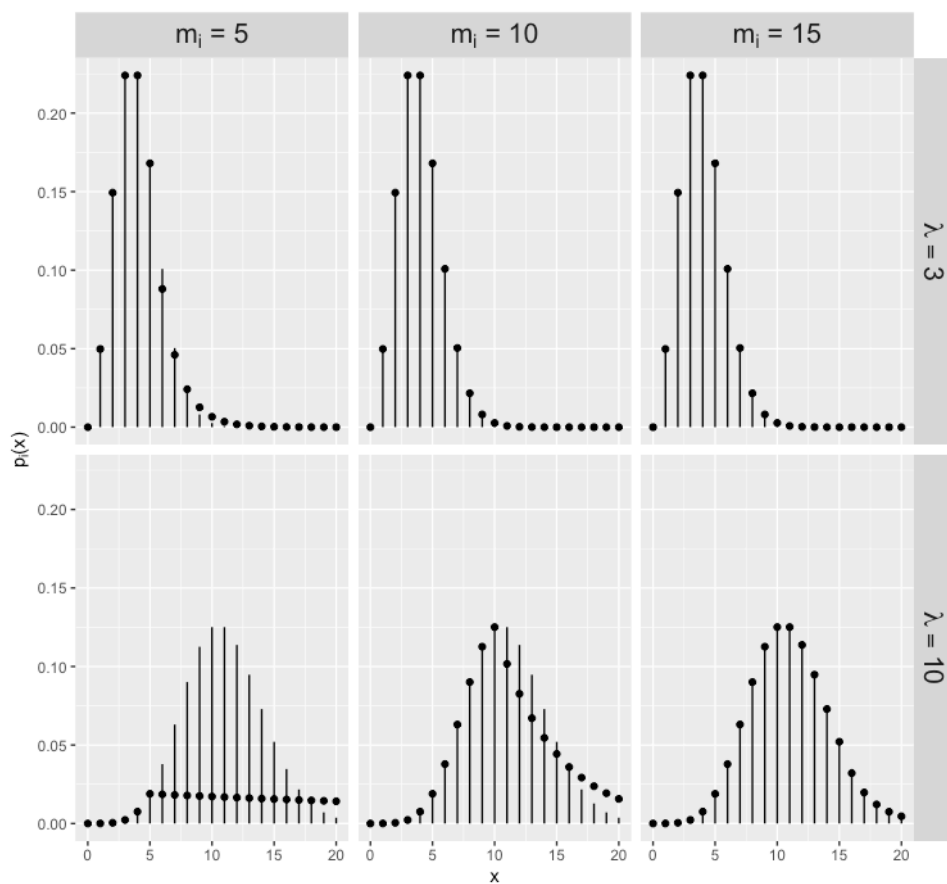


Figure 2: In each plot, we have the shifted Poisson distribution with shift parameter 1 and rate parameter $\lambda = 3$ (top row) and $\lambda = 10$ (bottom row). The columns correspond to the different sizes m_i of the state aggregates. The vertical lines in each plot shows the probability mass function of the shifted Poisson distribution, and the dots correspond to the induced dwell-time distribution of the approximated HSMM for a given m_i .

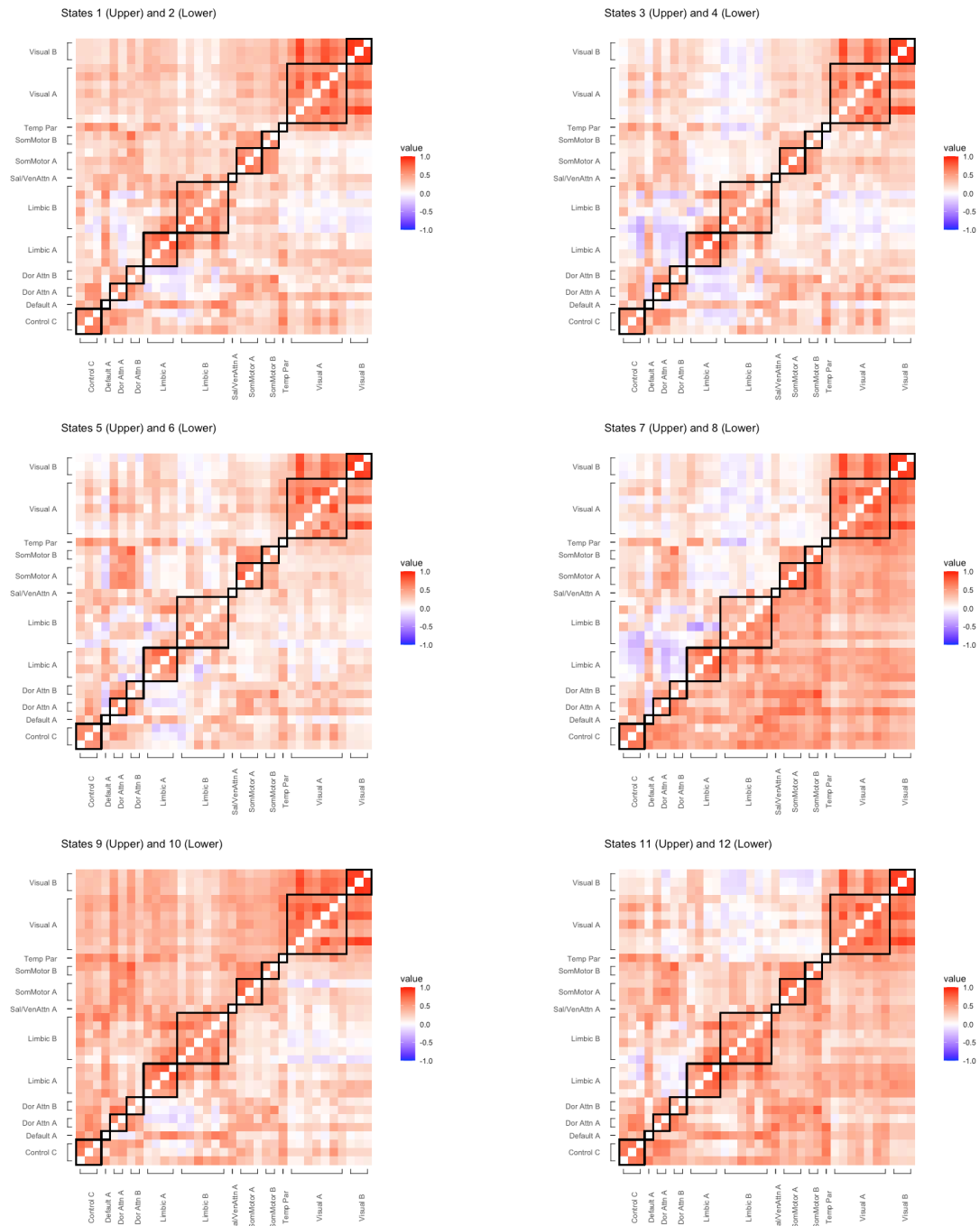


Figure 3: The correlation matrix for each state, with upper and lower triangles corresponding to different states. The black squares within the plot correspond to the ICNs mapped to the same Yeo regions.

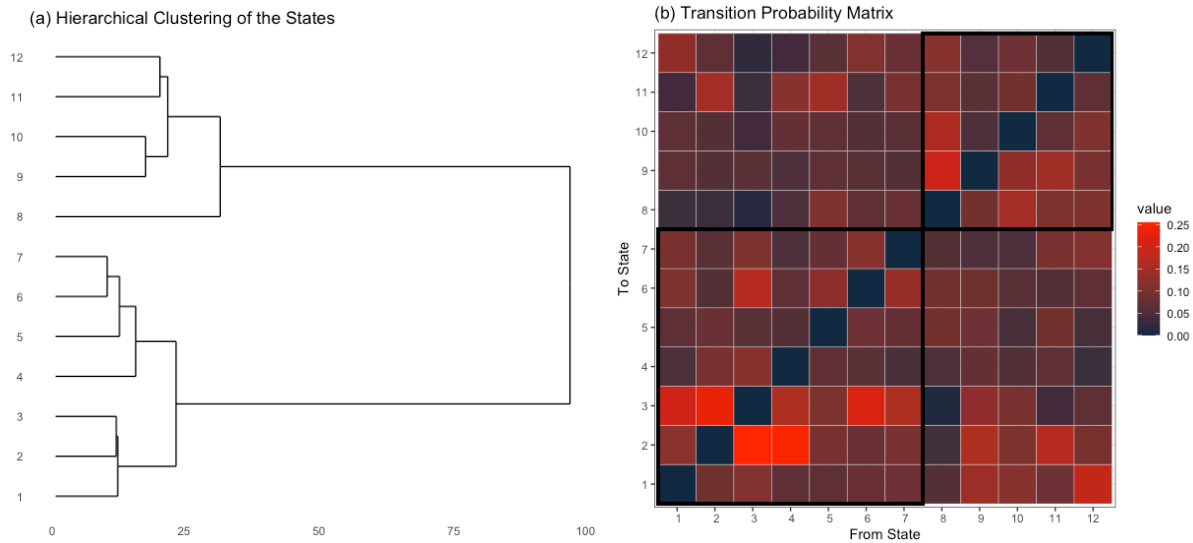


Figure 4: (a) Dendrogram for the covariance matrices across the 12 states. (b) Transition probability matrix for the approximate HSMM. States 1-7 and 8-12 correspond to metastates 1 and 2, respectively. These states are outlined in the black squares.

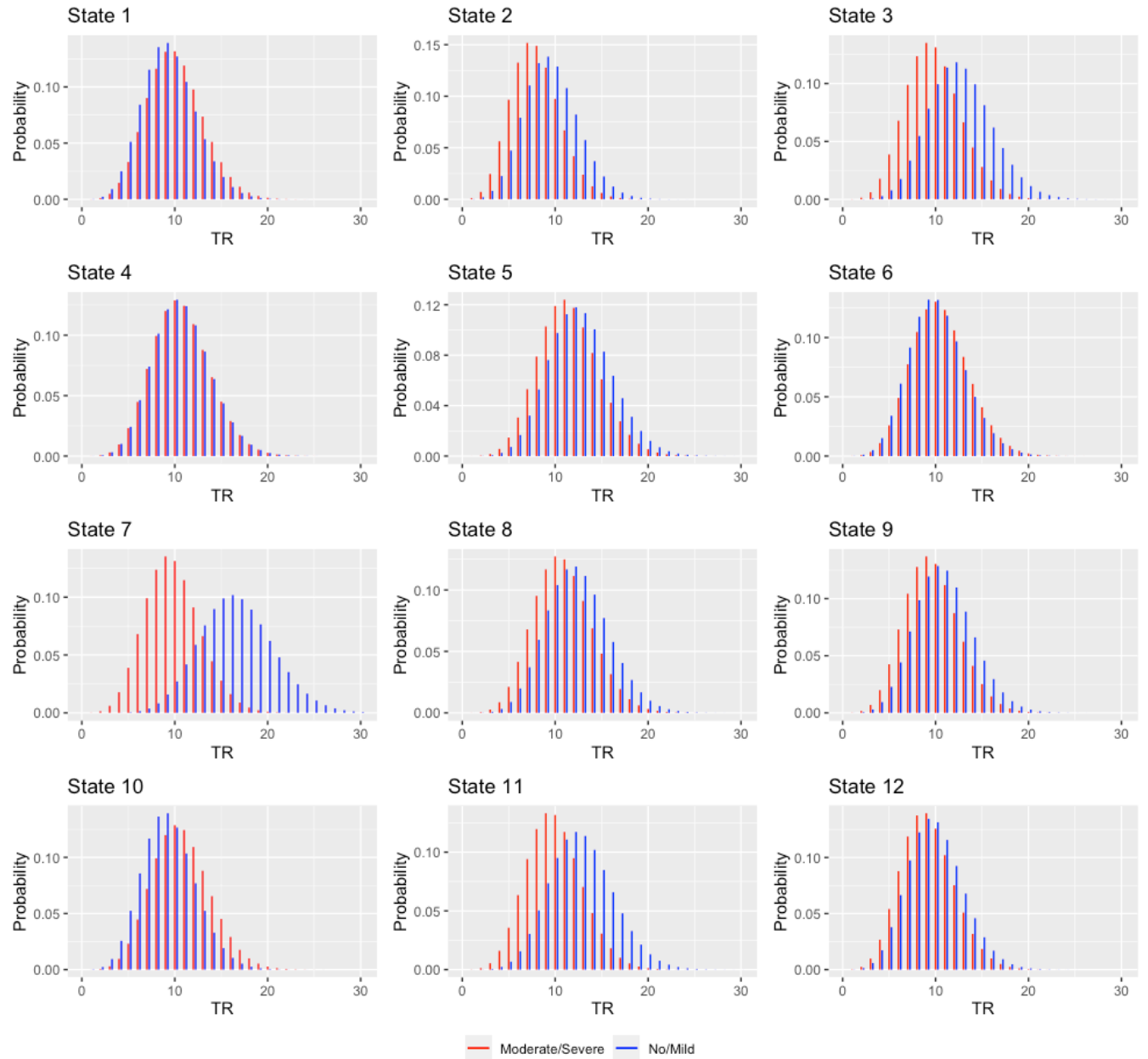


Figure 5: Dwell-time distributions for each state, stratified by NSSI severity group.

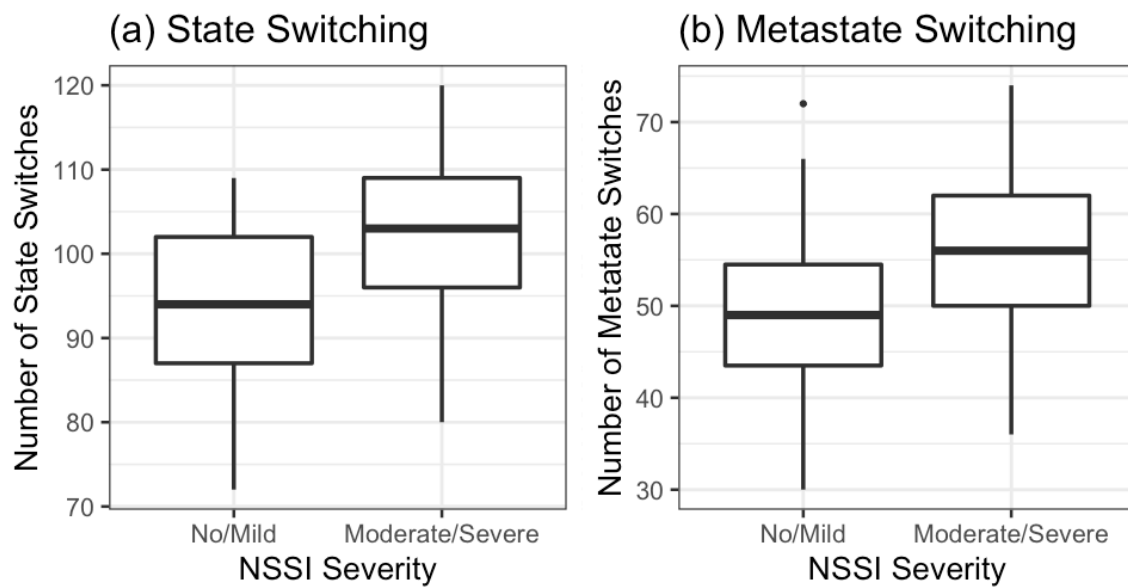


Figure 6: Boxplots of (a) state-switching frequency and (b) metastate-switching frequency per NSSI severity group.

List of Tables

1	Simulation results showing the performance under various conditions of the HMM and the approximate HSMM with different sizes, m , for the state aggregates, reported as the average (SD) across the 100 simulations. The HMM does not have a parameterized dwell-time distribution, hence its entries corresponding to β are marked with a hyphen (-).	31
2	Simulation results showing the performance under various conditions of the MSHMM, the HMM, and the approximate HSMM $m = 10$ reported as the average (SD) across the 100 simulations. The HMM does not have a parameterized dwell-time distribution, hence its entries corresponding to β are marked with a hyphen (-).	32
3	Parameter estimate (standard error) for the dwell-time distribution for each state and each NSSI severity group. We controlled the p -values for the 12 comparisons using the false discovery rate procedure.	33
4	Mean (SD) fractional occupancy per state and metastate for each NSSI severity group.	34

Dwell-time Distribution	P	Model	$E(\ \hat{\Sigma}_1 - \Sigma_1\ _F)$	$E(\ \hat{\Sigma}_2 - \Sigma_2\ _F)$	$E(\ \hat{\Sigma}_3 - \Sigma_3\ _F)$	$E(\hat{\beta}_1)$	$E(\hat{\beta}_2)$	$E(\hat{\beta}_3)$	Misclassification Rate
Shifted Poisson	10	HMM	0.067 (0.028)	0.611 (1.082)	0.241 (0.591)	-	-	-	0.114 (0.092)
		Approx. HSM (m=5)	0.065 (0.020)	0.512 (1.098)	0.214 (0.598)	1.930 (0.088)	0.627 (0.263)	1.463 (0.062)	0.102 (0.099)
		Approx. HSM (m=10)	0.065 (0.024)	0.504 (1.100)	0.214 (0.603)	2.326 (0.236)	0.604 (0.286)	1.533 (0.111)	0.099 (0.104)
		Approx. HSM (m=15)	0.093 (0.156)	0.499 (1.063)	0.209 (0.546)	2.334 (0.780)	0.594 (0.423)	1.489 (0.348)	0.105 (0.133)
Shifted Negative Binomial	30	HMM	0.179 (0.016)	0.664 (0.033)	0.278 (0.009)	-	-	-	0.028 (0.001)
		Approx. HSM (m=5)	0.179 (0.016)	0.641 (0.033)	0.274 (0.009)	1.954 (0.002)	0.514 (0.023)	1.426 (0.011)	0.023 (0.001)
		Approx. HSM (m=10)	0.179 (0.016)	0.639 (0.033)	0.274 (0.009)	2.371 (0.003)	0.505 (0.022)	1.492 (0.013)	0.022 (0.001)
		Approx. HSM (m=15)	0.179 (0.016)	0.639 (0.028)	0.274 (0.009)	2.491 (0.005)	0.505 (0.023)	1.493 (0.014)	0.022 (0.001)
Shifted Negative Binomial	10	HMM	0.125 (0.034)	0.147 (0.020)	0.076 (0.009)	-	-	-	0.085 (0.006)
		Approx. HSM (m=5)	0.119 (0.029)	0.136 (0.018)	0.073 (0.009)	1.508 (0.011)	1.843 (0.008)	1.998 (0.005)	0.082 (0.006)
		Approx. HSM (m=10)	0.119 (0.030)	0.135 (0.018)	0.071 (0.008)	1.631 (0.015)	2.251 (0.011)	2.471 (0.006)	0.077 (0.005)
		Approx. HSM (m=15)	0.120 (0.030)	0.137 (0.019)	0.071 (0.008)	1.628 (0.016)	2.321 (0.017)	2.672 (0.010)	0.077 (0.005)
Shifted Negative Binomial	30	HMM	0.329 (0.029)	0.375 (0.015)	0.193 (0.007)	-	-	-	0.014 (0.001)
		Approx. HSM (m=5)	0.330 (0.030)	0.373 (0.015)	0.192 (0.007)	1.503 (0.010)	1.875 (0.005)	2.018 (0.004)	0.014 (0.001)
		Approx. HSM (m=10)	0.330 (0.030)	0.372 (0.015)	0.192 (0.007)	1.613 (0.013)	2.239 (0.008)	2.466 (0.005)	0.014 (0.001)
		Approx. HSM (m=15)	0.330 (0.030)	0.372 (0.015)	0.192 (0.007)	1.613 (0.013)	2.308 (0.010)	2.654 (0.008)	0.014 (0.001)

Table 1: Simulation results showing the performance under various conditions of the HMM and the approximate HSM with different sizes, m , for the state aggregates, reported as the average (SD) across the 100 simulations. The HMM does not have a parameterized dwell-time distribution, hence its entries corresponding to β are marked with a hyphen (-).

Dwell-time Distribution	P	Model	$E(\ \hat{\Sigma}_1 - \Sigma_1\ _F)$	$E(\ \hat{\Sigma}_2 - \Sigma_2\ _F)$	$E(\ \hat{\Sigma}_3 - \Sigma_3\ _F)$	$E(\hat{\beta}_1)$	$E(\hat{\beta}_2)$	$E(\hat{\beta}_3)$	Misclassification Rate
Shifted Poisson	10	MHSM	0.066 (0.020)	0.371 (0.773)	0.119 (0.065)	2.498 (0.106)	0.489 (4.199)	1.388 (1.366)	0.122 (0.055)
		HMM	0.066 (0.020)	0.471 (0.770)	0.149 (0.069)	-	-	-	0.099 (0.045)
		Approx. HSMM ($m = 10$)	0.066 (0.020)	0.371 (0.773)	0.119 (0.065)	2.361 (0.099)	0.554 (0.163)	1.525 (0.066)	0.082 (0.047)
	30	MHSM	0.178 (0.018)	0.642 (0.026)	0.278 (0.009)	2.459 (0.181)	-1.062 (6.696)	1.346 (0.829)	0.085 (0.102)
		HMM	0.178 (0.018)	0.669 (0.032)	0.281 (0.009)	-	-	-	0.028 (0.001)
		Approx. HSMM ($m = 10$)	0.178 (0.018)	0.642 (0.026)	0.278 (0.009)	2.371 (0.004)	0.506 (0.026)	1.492 (0.012)	0.022 (0.001)
Shifted Negative Binomial	10	MHSM	0.121 (0.033)	0.134 (0.014)	0.071 (0.007)	1.612 (0.259)	2.328 (0.066)	2.698 (0.048)	0.082 (0.077)
		HMM	0.127 (0.033)	0.146 (0.019)	0.075 (0.008)	-	-	-	0.085 (0.007)
		Approx. HSMM ($m = 10$)	0.121 (0.033)	0.134 (0.014)	0.071 (0.007)	1.632 (0.013)	2.252 (0.013)	2.472 (0.006)	0.077 (0.005)
	30	MHSM	0.329 (0.028)	0.373 (0.016)	0.192 (0.006)	1.455 (0.480)	2.314 (0.097)	2.731 (0.086)	0.015 (0.001)
		HMM	0.329 (0.028)	0.375 (0.016)	0.194 (0.006)	-	-	-	0.014 (0.001)
		Approx. HSMM ($m = 10$)	0.329 (0.028)	0.373 (0.016)	0.192 (0.006)	1.614 (0.012)	2.239 (0.008)	2.467 (0.005)	0.014 (0.001)

Table 2: Simulation results showing the performance under various conditions of the MSHMM, the HMM, and the approximate HSMM $m = 10$ reported as the average (SD) across the 100 simulations. The HMM does not have a parameterized dwell-time distribution, hence its entries corresponding to β are marked with a hyphen (-).

State	NSSI Severity		Adjusted p -value
	No/Mild	Moderate/Severe	
State 1	2.107 (0.019)	2.201 (0.014)	< 0.001
State 2	2.126 (0.017)	1.926 (0.015)	< 0.001
State 3	2.437 (0.018)	2.169 (0.015)	< 0.001
State 4	2.261 (0.021)	2.268 (0.016)	0.788
State 5	2.445 (0.020)	2.344 (0.016)	< 0.001
State 6	2.196 (0.018)	2.248 (0.016)	0.035
State 7	2.737 (0.018)	2.168 (0.014)	< 0.001
State 8	2.418 (0.016)	2.284 (0.015)	< 0.001
State 9	2.271 (0.016)	2.149 (0.013)	< 0.001
State 10	2.101 (0.015)	2.268 (0.012)	< 0.001
State 11	2.455 (0.016)	2.186 (0.015)	< 0.001
State 12	2.174 (0.021)	2.094 (0.017)	0.003

Table 3: Parameter estimate (standard error) for the dwell-time distribution for each state and each NSSI severity group. We controlled the p -values for the 12 comparisons using the false discovery rate procedure.

State	Fractional Occupancy	
	No/Mild	Moderate/Severe
1	0.075 (0.034)	0.088 (0.042)
2	0.105 (0.050)	0.109 (0.056)
3	0.124 (0.095)	0.137 (0.076)
4	0.069 (0.034)	0.071 (0.032)
5	0.083 (0.037)	0.071 (0.036)
6	0.093 (0.049)	0.100 (0.049)
7	0.079 (0.050)	0.083 (0.034)
8	0.080 (0.080)	0.065 (0.066)
9	0.081 (0.033)	0.076 (0.034)
10	0.063 (0.046)	0.060 (0.039)
11	0.086 (0.047)	0.076 (0.044)
12	0.061 (0.028)	0.063 (0.033)
Metastate 1	0.544 (0.097)	0.531 (0.090)
Metastate 2	0.456 (0.097)	0.469 (0.090)

Table 4: Mean (SD) fractional occupancy per state and metastate for each NSSI severity group.