Bayesian hierarchical modeling on covariance valued data

Satwik Acharyya, Zhengwu Zhang, Anirban Bhattacharya, Debdeep Pati

Analysis of structural and functional connectivity (FC) of human brains is of pivotal importance for diagnosis of cognitive ability. The Human Connectome Project (HCP) provides an excellent source of neural data across different regions of interest (ROIs) of the living human brain. Individual specific data were available in the form of time varying covariance matrices representing the brain activity as the subjects perform a specific task. As a preliminary objective of studying the heterogeneity of brain connectomics across the population, we develop a probabilistic model for a sample of covariance matrices using a scaled Wishart distribution. We stress here that our data units are available in the form of covariance matrices, and we use the Wishart distribution to create our likelihood function rather than its more common usage as a prior on covariance matrices. Based on empirical explorations suggesting the data matrices to have a low effective rank, we further model the center of the Wishart distribution using an orthogonal factor model type decomposition. We encourage shrinkage toward a low rank structure through a novel shrinkage prior and discuss strategies to sample from the posterior distribution using a combination of Gibbs and slice sampling. The efficacy of the approach is explored in various simulation settings and exemplified on several case studies including our motivating HCP data. We extend our modeling framework to a dynamic setting to detect change points.

KEYWORDS
change point, covariance matrix, functional connectivity, low rank, Stiefel manifold, Wishart distribution

1 | INTRODUCTION

Functional connectomes play a critical role in determining how the brain responds to everyday tasks and life's challenges (Glasser et al., 2016; Jbabdi et al., 2015; Park & Friston, 2013). In recent years, there has been an abundance of literature focusing on understanding the variation of functional connectomes in healthy and diseased people and their relationships to various covariates and phenotypes (Finn et al., 2015; Smith et al., 2015; Zhang et al., 2018). Such interests are inspired and propelled by large scale neuroimaging studies, such as the Human Connectome Project (HCP) (Glasser et al., 2016; Van Essen et al., 2013), the Alzheimer's Disease Neuroimaging Initiative (ADNI) (Weiner et al., 2010) and the UK Biobank (Miller et al., 2016). In this article, we focus our attention to functional connectome (FC) inferred from functional magnetic resonance imaging (fMRI) data that measures the blood oxygen level dependent (BOLD) contrast signals of each brain voxel. As opposed to the anatomical axon connections (also referred as structural connectome), FC quantifies functional dependences between brain regions through correlations or covariances of BOLD signals. Conventional FC is often represented as a covariance or correlation matrix of fMRI data over a long recording time (Friston, 2011; Hutchison et al., 2013), where the matrix size equals the number of ROIs being considered.

While FC is assumed to be fixed or static over time in earlier studies, there is an abundance of evidence (Hindriks et al., 2016; Hutchison et al., 2013; Monti et al., 2014) in recent studies showing that FC is a dynamic process. The dynamic FC (dFC) is represented as a time series of short-term FCs, which are calculated using functional MRI data over short time intervals. Due to limitations of the fMRI blood oxygen level...
dependent (BOLD) contrast signals, fMRI signals are not directly analyzed (Glover, 2011; Turner, 2016). The most popular way to transfer BOLD signals into something that is reasonable to analyze is to calculate coherence between different brain regions, for example, correlation or covariance. We choose to use the covariance matrix, which in general carries more information than the correlation matrix. The goal of this paper is to understand and infer on the structure of dFC and detect change points in the dFC as the subjects perform a specific action. We first model the short-term FC using a scaled Wishart distribution and then generalize the static model to a hierarchical model of a time series of covariance matrices. Our final goal is to detect and compare individual specific change points along the dFC based on this hierarchical model.

As argued before, a first step toward change point detection is to model a population of covariance matrices. This is entirely different from covariance matrix estimation from multivariate data, which is a well-studied problem; see Daniels and Kass (1999), Leonard and Hsu (1992), and Pati et al. (2014) as some representative examples of Bayesian inference for covariance matrices and Pourahmadi (2011) for a more comprehensive review. In the covariance estimation context, the observational data vectors are directly available and the goal is to characterize the dependence among different variables in the data from multiple independent and identically distributed samples. On the other hand, our observational units are covariance matrices corresponding to different individuals observed over time, which we shall henceforth refer to as covariance-valued data. Hierarchical models for capturing heterogeneity in multiple related groups based on covariance matrices are relatively fewer in numbers (Boik, 2002; Flury, 1984, 1987; Franks & Hoff, 2019; Hoff, 2009a). Schott (1999, 2001) developed hypothesis testing methods based on covariance structures. In a Bayesian context, Barnard et al. (2000), Daniels (2006), and Pourahmadi et al. (2007) considered parsimonious modeling of covariance matrices, which were extended to a longitudinal setting by Das and Daniels (2014), Gaskins and Daniels (2013), Gaskins et al. (2014), and Gaskins and Daniels (2016). A parallel sequence of works proposes modeling of fMRI data-matrices via Gaussian graphical modeling techniques (Stingo et al., 2013; Warnick et al., 2018). In contrast to the existing studies, we built a hierarchical model on observed covariance valued datasets to detect individual specific change points. The literature on probabilistic modeling for covariance-valued data in the time series context (Golosnoy et al., 2012; Gouriéroux et al., 2009; Yu et al., 2017) is focused on maximum likelihood estimation using a non-central Wishart distribution as the likelihood. For example, the dataset considered in Yu et al. (2017) comprises low-dimensional ($5 \times 5$) daily realized covariance (RCOV) matrices for five stocks observed across 2274 time points. This single time series sequence is modeled using a generalized conditional auto-regressive Wishart (GCAW) model. In presence of smaller number of parameters and a huge collection of time points, maximum likelihood estimation is a natural choice for model fitting. On the other hand, since we are dealing with $10 \times 10$ covariance matrices observed over 26 time points for 500 individuals, it is important to borrow information across individuals and seek for a parsimonious modeling framework.

To that end, we develop a suite of hierarchical modeling techniques for covariance-valued data to provide insight into the structural connectivity of human brains. We use a scaled version of the Wishart distribution to model the covariance-valued observations. While the Wishart distribution is commonly used as a prior distribution on inverse-covariance or precision matrices in Bayesian inference, its usage as a likelihood is novel in the Bayesian context to the best of our knowledge. The presence of a modest number of observations further necessitates structured modeling of the center of the Wishart distribution, which itself is a covariance matrix. Based on empirical evidence of low effective ranks of the data matrices, we modeled the center of the Wishart model using an orthogonal factor model type decomposition and encouraged shrinkage toward a low rank structure through the development of a novel shrinkage prior. We use a combination of Gibbs and slice sampling to sample from the posterior distribution whose steps are mostly standard.

Our primary objective is to explore the dynamic nature of FC between different brain regions during performances of certain tasks. A dynamical FC model provides an overall architecture of how the brain functions as an individual performs certain tasks. An important scientific goal is to identify change points (Barry & Hartigan, 1993) in the time series of covariances that split the data into contiguous segments. Difference in the change points across individuals is indicative of behavioral and cognitive differences (Dai et al., 2017). To address this, we extend our hierarchical model to accommodate a single or multiple change points in a fully Bayesian framework. A novel combination of existing MCMC algorithms renders sampling from the joint posterior distribution tractable. The change point model is then implemented on both the HCP and the ADNI datasets to extract scientifically meaningful conclusions. For the HCP dataset, we studied the change point pattern during the motor task and discovered the primary FC change point occurs when people switch the movement from hand and foot to the tongue. For the ADNI dataset, we compared FCs in two groups of older people (supernormal subjects and normal controls) and found that supernormal subjects have higher strength of connectivity within posterior regions or between posterior and anterior regions of their brain.

In this paper, we begin with an illustration of our motivating data set in Section 2 followed by a model for covariance matrices (Section 3.1), a hierarchical covariance model (Section 3.2) and a hierarchical change point model (Section 4). Results of detailed simulation studies are provided in the corresponding sections. In Section 5, we provide the results obtained from our motivating HCP dataset under the hierarchical change point model.

2 | DATA DESCRIPTION

We utilize functional MRI data from two large datasets, ADNI (Weiner et al., 2010) and HCP (Van Essen et al., 2013), to illustrate the proposed method. ADNI was initiated by National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, the Food and Drug
Administration, and some private pharmaceutical companies and non-profit organizations. ADNI assesses clinical, imaging, genetic, and bio specimen biomarkers through the process of normal aging to early mild cognitive impairment, to late mild cognitive impairment, to dementia or Alzheimer’s disease (AD). Participants were recruited across North America to participate in three phases of the study: ADNI1, ADNI GO, and ADNI2. A variety of imaging and clinical assessments were conducted for each participant. Results were then shared by ADNI through the Laboratory of Neuro Imaging’s Image Data Archive (https://ida.loni.usc.edu/). In our study, we focus on a subset of healthy subjects that were previously identified in (Lin et al., 2017). These subjects were AD free but were clustered in two groups. The first group is called supernormals who exhibited excellent episodic memory and executive function. The other group is age-matched healthy control subjects. All their resting-state fMRI data were collected using a 3.0 Tesla Phillips MRI with an echo-planar imaging sequence (spatial resolution = 3 × 3 × 3 mm³). Structural images were obtained using an MPRAGE sequence (spatial resolution 1 × 1 × 1 mm³), which were then used for registration during preprocessing. Across individuals, the first 10 volumes were discarded to avoid potential noise related to the equilibrium of the scanner and participant’s adaptation process. The remaining 130 volumes were preprocessed using slice time correction and head motion correction. The images were then registered to each individual’s own structural image, normalized to the Montreal Neurological Institute standard space and spatially smoothed using a Gaussian kernel (FWHM = 4 mm). We utilized the automated anatomical labeling (AAL) (Tzourio-Mazoyer et al., 2002) to percolate the whole brain into 116 regions of interest (ROIs).

The HCP project aims at characterizing human brain connectivity in >1000 healthy adults and to enable detailed comparisons between brain circuits, behavior and genetics at the level of individual subjects. The HCP raw and preprocessed data can be easily accessed through ConnectomeDB (https://www.humanconnectome.org). The high-quality imaging data and easy accessibility make it an ideal dataset for this paper. Majority of the HCP fMRI data were acquired at 3T with a 2 × 2 × 2 mm³ resolution. Preprocessing steps using the HCP pipeline (Glasser et al., 2013, 2016) were performed before any data analysis, for example, removing spatial distortions, realigning volume to compensate for subject motion, registering the fMRI to the structural MRI, reducing the bias field, normalizing the 4D image to a global mean, masking the data with the final brain mask, and aligning the brain to a standard space. Figure 1 provides an overview of the preprocessing steps. The Destrieux atlas (Destrieux et al., 2010) was used to percolate cortical regions into 74 nodes per hemisphere. Similar to Dai et al. (2017), for the fMRI BOLD signal in each ROI, we first calculate a mean time series, and then we utilize a sliding window method to calculate a covariance trajectory \( S_t^{i} \) for subject \( i \) at the \( t \)th window based on selected ROIs. Therefore, \( S_t \) is a \( p \times p \) covariance matrix representing the short time functional connectome, where \( p \) denotes the number of ROIs.

3 | HIERARCHICAL MODELING FOR COVARIANCE DATASET

Since the covariance matrices are observed for multiple individuals and time points, a natural course of action is to build a parsimonious model that borrows strength across all observational units. We first discuss an independence model with scaled Wishart distribution for covariance-valued data that serves as the basic building block for the forthcoming extensions. Both the central and non-central Wishart distributions have full support on the space of covariance matrices. However, the non-central Wishart distribution has more parameters making prior elicitation more complicated. Moreover, the presence of key parameters inside a hypergeometric function complicates posterior computation; in particular, the nice conjugate structure that we exploit throughout the article is lost. Due to these reasons, we consider a central Wishart likelihood through the rest of the paper.

---

**FIGURE 1** An overview of preprocessing steps for extracting dynamic FC from fMRI data.
Motivated by a pattern we observe in the functional connectivity data, the mean structure of the independence model is encouraged to shrink toward low rank matrices via a parsimonious shrinkage prior. We develop an MCMC algorithm to fit the independence model to data and show its efficacy in a simulation study. Next, the independence model is extended to a Bayesian hierarchical model to incorporate multiple individuals, allowing for subject specific deviations from a common mean structure. Fitting the hierarchical model requires sampling from a class of distributions on the Stiefel manifold, which can be done efficiently using the algorithm in § 3.3 of Hoff (2009b); refer to Section 3.2 herein for more details. The hierarchical model leads to our eventual goal of detecting subject specific change points in the functional connectivity data.

3.1 Independence model

We begin by describing the details of the independence model. Let \( \{S_j\}_{j=1}^N \) be a collection of independent and identically distributed \( p \times p \) covariance matrices. We probabilistically model the \( S_j \)s using a Wishart distribution, which is arguably the most recognized distributional family for covariance matrices. We shall use the standard \( W_p(\nu, V) \) notation to denote the Wishart distribution on the space of \( p \times p \) positive definite matrices, with degrees of freedom \( \nu \times p - 1 \) and a \( p \times p \) positive definite scale matrix \( V \). The density \( W_p(\nu, V) \) distribution has a density (in \( X \)) proportional to

\[
|V|^{-\nu/2} |X|^{-(\nu - p - 1)/2} e^{-\text{tr}(V^{-1}X)/2}.
\]

Specifically, we use a scaled Wishart distribution \( W_p(\phi, \phi^{-1}\Omega) \) to model the \( S_j \)s,

\[
S_j \overset{\text{iid}}{\sim} W_p(\phi, \phi^{-1}\Omega), \quad j = 1...N.
\]

The introduction of the parameter \( \phi \) in the scale matrix is to decouple its presence in both the mean and covariance. For \( S_1 \sim W_p(\phi, \Omega) \), one has

\[
E[S_1] = \phi\Omega, \quad \text{Var}(S_{1i}) = \phi(\omega_i^2 + \omega_j^2), \quad \text{Cov}(S_{1i}, S_{1j}) = \phi(\omega_i\omega_j + \omega_j\omega_i),
\]

whereas for \( S_2 \sim W_p(\phi, \phi^{-1}\Omega) \),

\[
E[S_1] = \Omega, \quad \text{Var}(S_{1i}) = \phi^{-1}(\omega_i^2 + \omega_j^2), \quad \text{Cov}(S_{1i}, S_{1j}) = \phi^{-1}(\omega_i\omega_j + \omega_j\omega_i).
\]

Thus, in the parameterization we work with, \( \Omega \) is the population mean. We henceforth fix \( \phi \) at \( (p+1) \) and validate our assumption in the Section S2.1 of Supporting Information. Now, we focus our attention on modeling the mean \( \Omega \).

An unstructured \( p \times p \) covariance matrix has \( p(p+1)/2 \) free elements (e.g., in the HCP dataset, \( p = 10 \) leading to a total number of 55 parameters and in the ADNI dataset, \( p = 7 \) in the present case leading to 28 parameters). Given that we only have a modest number of time points, it is important to make meaningful structural assumptions on \( \Omega \) to reduce the effective number of parameters to be estimated. We conducted an exploratory analysis to find patterns in the data matrices that could direct us toward a parsimonious model. The data matrices and their inverses did not contain any obvious sparsity pattern. Next, we investigated the effective ranks of the data matrices. For a \( p \times p \) positive definite matrix \( A \) with eigenvalues \( s_1(A) \geq s_2(A) \geq ... \geq s_p(A) \geq 0 \), its effective or intrinsic rank (Vershynin, 2012),

\[
r_e(A) := \frac{\sum_{k=1}^{p} s_k(A)}{s_1(A)}
\]

is the ratio of its trace and the largest eigenvalue. The effective rank satisfies \( 1 \leq r_e(A) \leq \text{rank}(A) \), so that it always provides a lower bound to the actual rank. Further, the effective rank is a smooth function of its argument. For example, consider the class of matrices

\[
M_{\lambda} := uu^T + \lambda I_p
\]

for \( \lambda > 0 \) and \( u \) a \( p \)-dimensional vector of unit length. The matrices \( M_{\lambda} \) increasingly get close to being rank deficient as \( \lambda \downarrow 0 \); however, this is not captured by the rank as \( \text{rank}(M_{\lambda}) = p \) for any \( \lambda > 0 \). On the other hand, \( r_e(M_{\lambda}) = 1 + (p - 1)\lambda/(1 + \lambda) \), which smoothly decays to 1 as \( \lambda \downarrow 0 \). These features render the effective rank a suitable measure to capture the intrinsic dimensionality of a matrix and indicate potential near rank-deficiencies.
Figure 2 shows boxplots of the effective ranks of the data matrices across the 26 time points for 50 randomly selected individuals from the motor task of the HCP dataset. It is evident that the 10 × 10 data matrices have a low effective rank, with the bulk of the empirical effective rank distribution between 1.5 and 3. This observation motivated us to consider an orthogonal factor model type decomposition (Hoff, 2009a) for Ω to exploit the near low-rank structure as

\[
Ω = VDV^T + \sigma^2 I_p,
\]

where \(V \in \mathbb{R}^{p \times r^*}\) for some \(r^* \leq p\) is a semi-orthogonal matrix satisfying \(V^TV = I_{r^*}\), and \(D = \text{diag}(d_1, \ldots, d_{r^*})\) is a diagonal matrix with non-negative diagonal entries. Such a decomposition readily satisfies the positive definiteness constraint on Ω. Different variance components are typically employed in factor analysis to account for variables with possibly different scales. However, since all the entries represent functional connectivities, we make the simple assumption of using the same \(\sigma^2\) for all the components.

We operate in a Bayesian framework to perform inference based on the posterior distribution of the model parameters. Before proceeding to describe our prior specifications, it is important to discuss the role of \(r^*\) in what follows. In a fully Bayesian framework, one may treat \(r^*\) as a parameter that designates the effective rank of Ω and assign it a prior distribution; the discrete uniform distribution on \([1, \ldots, p]\) being a default choice. Under this prior, the posterior distribution of \(r^*\) is proportional to the marginal likelihood of the data given \(r^*\), which is intractable in the present context. While it is possible to sample \(r^*\) inside a larger trans-dimensional MCMC algorithm such as the reversible jump MCMC (RJMC), its implementation remains computationally challenging, especially when considering extensions to the hierarchical modeling setup later on. Moreover, the effective rank does not have a clear biological interpretation in our real application and is purely a modeling device to induce parsimony. Based on these considerations, we undertake a shrinkage approach rather than explicit selection of the rank. Specifically, we set \(r^*\) to a conservative upper bound, with \(p\) being a default choice, and encourage a subset of the diagonal entries of \(D\) to shrink toward zero. If \(A \subset \{1, \ldots, p\}\) denotes the active subset, that is, the subset of diagonal entries of \(D\) that are left unshrunk, then \(VDV^T \approx V_A D_A V_A^T\), where \(V_A\) denotes the \(p \times |A|\) sub-matrix of \(V\) corresponding to the columns in \(A\), and \(D_A\) denotes the corresponding \(|A| \times |A|\) diagonal sub-matrix of \(D\). This leads to an approximately low rank decomposition under the posterior, which is sufficient for our purpose. In the factor modeling context, Bhattacharya and Dunson (2012) considered a shrinkage prior on the factor loadings matrix rather than placing a prior on the number of factors, for example, as in Lopes and West (2004). We have a very different shrinkage mechanism as our shrinkage operates on the diagonal matrix \(D\).

Fixing \(r^*\), the unknown parameters in our model are \((V, D, \sigma^2)\) with parameter space \(V_{p \times r^*} \otimes D_{r^*} \otimes \mathbb{R}_{++}\), where \(V_{p \times r^*}\) denotes the Stiefel manifold of \(p \times r^*\) semi-orthogonal matrices, and \(D_{r^*}\) is the collection of \(r^*\) dimensional diagonal matrices with non-negative entries. The likelihood function for the parameters is given by

\[
L(V, D, \sigma^2) = |Ω|^\frac{N}{2} \prod_{j=1}^{N} \exp \left\{ -\frac{\phi}{2} \text{tr}(Ω^{-1}S_j) \right\}.
\]

We now discuss prior choices on the parameters. For computational convenience, we reparameterize to \((V, D, \sigma^2)\) where \(\bar{D} = D/\sigma^2\), so that \(Ω = \sigma^2(\bar{V}D\bar{V}^T + I_p)\). We place a uniform prior on \(V\) supported on the Stiefel manifold \(V_{p \times r^*}\), and an inverse-gamma IG(\(\alpha_\sigma, \beta_\sigma\)) prior on \(\sigma^2\). To set up our sparsity favoring shrinkage prior on the diagonal entries \(\tilde{d}_j\)s of \(\bar{D}\), first decompose
\[ \tilde{d}_h = \tau d_h, \ h = 1, ..., r. \]  

(4)

In (4), \( r \) plays the role of a global shrinkage parameter while the \( d_h \)'s allow for coordinate specific deviations, much in the spirit of the global-local shrinkage priors popularly used in regression (Carvalho et al., 2010). We place independent half-Cauchy priors on the \( d_h \)'s, \( d_h \sim \text{Cauchy}(0,1) \), with density proportional to \( 1/(1+t^2) \) for \( t \). The half-Cauchy prior is a popular choice as a prior distribution of shrinkage parameters due to its positive density at zero and heavy tails (Carvalho et al., 2010; Polson & Scott, 2012). We complete the prior specification by placing a half-Cauchy prior truncated to \((0,1)\) on \( r \). Truncating the prior on the global parameter leads to better identifiability and is recommended by van der Pas et al. (2014) in the context of the horseshoe prior. The multiplicative prior on the \( d_h \)’s can also be interpreted as an additive one-way ANOVA type decomposition in the logarithmic scale,

\[ \log \tilde{d}_h = \mu + \beta_h, \ \mu = \log (\tau ), \ \beta_h = \log (d_h), \ h = 1, ..., r. \]

with grand mean \( \mu \) and main effects \( \beta_h \)’s. The posterior computation is also conveniently carried out in the logarithmic scale, which we describe next.

We develop a fully automated and easy to implement Markov chain Monte Carlo algorithm to sample from the joint posterior distribution of \((V, D, \sigma^2)\) given the data. Specifically, we use a combination of Gibbs sampling with slice sampling and Metropolis-within-Gibbs to iteratively sample from the full-conditional distribution of each parameter block given the rest. The sampler iterates through the following steps; the derivations are deferred to the Supporting Information (Section S6). We use the notation \( [\theta | - ] \) to denote the full conditional distribution of a parameter.

- Sample \( V \) from its matrix Bingham \((S^V, \phi E^{-1}/2\sigma^2)\) full-conditional distribution. The matrix Bingham \((A, B)\) distribution has a density with respect to the uniform distribution on the Stiefel manifold given by

\[ p_B(X|A, B) \propto \text{etr}(BX^TAX), \]

where \( A \) and \( B \) are symmetric and diagonal matrices, respectively. In our case, \( S^V = \sum_{j=1}^N S_j \) is a symmetric matrix by definition and \( E = (D^{-1} + I_r) \). The matrix Bingham distribution is conveniently sampled using the \texttt{R} package \texttt{ratiefel} (Hoff, 2013).

- Update the \((\beta_h)\)’s from their independent full conditional distributions using slice sampling. Set \( M = V^T S^V V \) and consider the transformation \( w_h = \frac{1 + \beta_h}{\mu } \) for \( h = 1, ..., r \). Then sample

\[ w_h | w_h, - \sim \text{Uniform} \left[ 0, \left( \mu ^2 + (1 - w_h)/w_h^2 \right)^{-1} \right], \]

\[ w_h | w_h, - \sim \text{Gamma} \left( \text{shape} = N\phi /2 - 1, \ \text{rate} = \phi M_{hh}/(2\sigma^2) \right) \] truncated to the region \( \left\{ 1 + \sqrt{1/w_h - \mu ^2} \right\}^{-1}, \infty \), and set \( \beta_h = (1 - w_h)/(w_h\mu ) \).

- To sample \( \mu \), propose \( \mu ^* \sim N(\mu, \sigma^2) \) and compute the Metropolis ratio

\[ a(\mu, \mu ^*) = \frac{\Pi (\mu ^* | - )}{\Pi (\mu | - )} \]

where \( \Pi (\mu | - ) \) denotes the full-conditional of \( \mu \). Accept \( \mu ^* \) with probability \( \min \{a(\mu, \mu ^*), 1\} \).

- Sample \( \sigma^2 \) from its inverse-gamma full conditional distribution as

\[ [\sigma^2 | - ] \sim \text{InvGamma} \left( \alpha_e - 1 + \frac{N\phi}{2}, \beta_e + \frac{\phi \text{etr}(Q S^V)}{2} \right) \]

where \( Q = (VDV^T + I_r)^{-1} \).

We observed good mixing and convergence of the above MCMC sampler based on standard MCMC diagnostics. Although not our primary motivation, one can estimate the effective rank based on a simple post-processing step of the MCMC samples for the \( \{d_h\} \)’s. As in Bhattacharya et al. (2015) and Li and Pati (2017) at each MCMC iteration, we cluster the \( \{d_h\} \)’s into two groups using 2-means clustering and save the size of the group having the larger mean. The mode of these numbers across the MCMC iteration is then used as an estimate of the effective or intrinsic rank. We find that this approach performs well in our simulation and real examples. A more nuanced approach for post-processing was proposed by Li and Pati (2017), which can also be used in the present context. A detailed simulation study of the independence model is deferred to Section S1 of the Supporting information.
3.2 Hierarchical covariance model

In this subsection, we extend the independence model to a hierarchical modeling framework encompassing all the individuals. Our hierarchical modeling framework lets different individuals share common parameters while allowing for subject specific deviations, striking a balance between pooling of information across different individuals while retaining flexibility. Letting $S_t$ denote the observed covariance matrix for individual $i$ at time $t$, we let

$$
S_t \sim W_p(\phi, \phi^{-1} \Omega_i) \quad t = 1, \ldots, T,
\Omega_i = V D_i V^T + \sigma_i^2 I_p \quad i = 1, \ldots, n.
$$

(5)

The first line of (5) posits the same scaled Wishart model as in the previous subsection with individual specific mean $\Omega_i$. As discussed earlier, we only have data on $T = 26$ time points for each individual. On the other hand, there are a relatively larger number of individuals in the study. For this reason, rather than separately fitting the independence model for each individual, we consider a structured decomposition of $\Omega_i$ that lets $D_i$ and $\sigma_i^2$ vary across individuals, while keeping $V$ fixed. This is akin to an expansion of the $\Omega_i$s in terms of a fixed dictionary $V$, with subject specific loadings. This fixed dictionary expansion vastly reduces the number of model parameters and allows one to borrow information across individuals to estimate the common dictionary $V$. We continue using $\sigma_i^2 I$ for the residual part in the covariance decomposition for model parsimony. We later conduct model validation (Section S2 of the Supporting Information) to show that model (5) provides an adequate fit to the data compared to separately fitting the independence model.

We continue to use the uniform prior on the Stiefel manifold for $V$. After reparameterizing to $\tilde{D}_i$, we place independent copies of the shrinkage prior introduced earlier on the $\tilde{D}_i$s, and independent inverse-gamma priors on the $\sigma_i^2$s.

We extend the MCMC algorithm for the independence model to the hierarchical setting. The updates for $\tilde{D}_i$ and $\sigma_i^2$ proceed independently across $i$ exactly along same lines as before. However, since $V$ is common to all individuals, its full conditional no longer remains a matrix Bingham distribution. We show in the Supporting Information (Section S7) that the full-conditional distribution of $V$ is given by

$$
|V|^{-1} \propto \exp \left[ \frac{d}{2} \sum_{i=1}^n \text{tr} \left( \frac{V E^{-1} V^T}{\sigma_i^2} \sum_{t=1}^T S_t \right) \right] \times \prod_{j=1}^K \exp(v_j^T H_j v_j),
$$

(6)

where $H_j = \sum_{i=1}^n \left( \frac{\phi S_i^{-1}}{2 \epsilon_i \sigma_i^2} \right) S_i^* = \sum_{t=1}^T S_t$, and $E_j = (D_j^{-1} + b_j)$.

We sample from the above density by adopting the Gibbs sampling scheme of Hoff (2009b) to sample from a class of matrix Bingham–von Mises–Fisher (BMF) distributions (Khatri & Mardia, 1977). The BMF distribution has a density on the Stiefel manifold given by $p_{BMF}(V|A, B, C) \propto \text{etr}(C^T V + B V^T A V)$. Hoff considers the case when $B$ is diagonal, noting that the general symmetric case can be handled by a transformation, when the density assumes the form

$$
p_{BMF}(V|A, B, C) \propto \prod_{j=1}^K \exp(c_j^T v_j + b_j v_j^T A v_j).
$$

Hoff used Gibbs sampling to sample from $p_{BMF}$ by alternately sampling from the full-conditional distributions of each column $v_j$ given the rest; see § 3.3 of Hoff (2009b) for details.

The distribution of $|V|^{-1}$ in Equation (6) is almost identical to the above BMF distribution; we have $C = 0$ in our case and matrices $H_j$ in place of $b_j A$. This minor difference is immaterial from a Gibbs sampling standpoint; the steps can be found in the Supporting Information: Section S7.

3.3 Simulation study for the hierarchical covariance model

We conduct a replicated simulation study to illustrate the operating characteristics of the hierarchical model. We set $n = 100, T = 26$ and $p = 50$ for our simulations. The true $V_0$ is generated uniformly on the Stiefel manifold. Also, for each $i$, the diagonal entries of the true $D_{ij}$ are generated uniformly between 0 and 5, while the $\sigma_i^2$s are generated uniformly between 0.25 and 0.50. We generate 100 independent simulation replications as above.

We fit the hierarchical model using the MCMC outlined in the previous subsection. We set $r^* = 10$ and use a modification of the empirical Bayes procedure to elicit the hyperparameters $\alpha_\nu$ and $\beta_\nu$. As metrics of parameter recovery, we considered
where \( \hat{\Omega}_i \) and \( \hat{\sigma}^2_i \) are the posterior means of \( \Omega_i \) and \( \sigma^2_i \) for \( i = 1, \ldots, n \). \( d^{(i)}_\Omega \) and \( d^{(i)}_\sigma \) are individual specific measures of the distance of the posterior mean from the truth, while \( d_\Omega \) and \( d_\sigma \) are average measures over all the individuals.

As a point of comparison, we also fit the independence model in the previous subsection separately for each individual. Figure 3 shows boxplots of \( \{d^{(i)}_\Omega\}_{i=1}^n \) averaged over the simulation replicates for the hierarchical and independence models. The tighter spread of the boxplot for the hierarchical model indicates the gains from borrowing information across subjects. The hierarchical model also successfully recovered the true ranks as shown in Figure 4.

We conducted a second set of simulations by varying \( n \in \{100, 200, 300, 400, 500\} \) and \( p \in \{50, 100, 150, 200, 250\} \). A summary is presented in Figure 5. In the top left panel, we provide the boxplot of \( d_\Omega \) across the 100 simulation replicates for the different values of \( n \) keeping \( p \) fixed, while the bottom left panel provides the same for varying \( p \) and fixed \( n \). As expected, the estimation performance improves for larger \( n \) and smaller \( p \).

We observe similar patterns in the density plots of \( d_\sigma \) w.r.t. increasing \( n \), fixed \( p \) (Figure 5b) and fixed \( n \), increasing \( p \) (Figure 5d).

### 3.4 Real data analysis for ADNI dataset

In this case study, we utilize 18 subjects’ resting-state fMRI data from ADNI. Half of them are from supernormal (SN) subjects who possess excellent (Lin et al., 2017), and the other half are healthy control (HC) subjects. Each group contains nine individuals with its resting state fMRI data at baseline preprocessed. From the previous literature, we identified seven interesting ROIs: left occipital cortex, left occipital cortex, left precuneus, left superior temporal cortex, right middle frontal gyrus, right parahippocampus, and right thalamus (indexed as ROI 1, 2, …, 7), which are potentially linked to cognition, emotional regulation, and memory. After preprocessing, we obtained a mean BOLD signal within each ROI and then applied a...
sliding window method to obtain a $7 \times 7 \times 24$ covariance matrix time series. We applied our proposed hierarchical model on this dataset for different groups of individuals and obtained Bayes estimates of individual specific covariance matrix $\Omega_i$ for $i = 1, \ldots, 9$. (Note that in Section S2.2 we have validated that there is no change point in these covariance trajectories.)

To compare the functional connectivity between ROIs across SN and HC, we look at all the off-diagonal elements of $\Omega_i(i = 1, \ldots, 9)$ for both SN and HC using an overlaid histogram in Figure 6. The overlaid histograms clearly show that on average the FC for the SN individuals is higher. In addition, it is also important to know which one out of the 21 pairs of regions accounts for the maximum separation in $\Omega_i$ between SN

**FIGURE 5** Left panel: Boxplot of $d_{\Omega}$ across simulation replicates for varying $n$ (top panel) and varying $p$ (bottom panel). Right panel: Density plots of $d_{\sigma}$ across simulated replicates for varying $n$ (top panel) and varying $p$ (bottom panel).

**FIGURE 6** Histogram of off-diagonal elements that denote FC between ROIs among supernormals (SN) and health controls (HC).
and HC. A simple multiple comparison test reveals that the FC difference between SN and HC for the ROI-pairs (2, 3) and (3, 6) are statistically significant (p-values 0.038 and 0.013), where (2, 3) represents a posterior regions’ connection and (3, 6) represents an anterior-posterior connection. This finding is in line with the literature (Lin et al., 2017): the SN group has higher strength of connectivity within posterior regions or between posterior and anterior regions. Box plots for ROI-pairs in Figure 7 clearly show that the FC for SN is higher than HC for both the ROI-pairs.

4 | HIERARCHICAL CHANGE POINT MODEL

Although a majority of previous works on modeling functional connectivity assumes stationarity (Friston, 2011; Hutchison et al., 2013), recent developments in Dynamic Connectivity Regression (Cribben et al., 2013) suggest the necessity of incorporating non-stationary modeling of the time series of covariance matrices. It is reasonable to assume that different parts of brains will react distinctly under the effect of external stimuli, so assuming a common mean for the Wishart distribution in (5) is not warranted unless the subjects are in a resting state. Moreover, in the presence of multiple subjects, it becomes necessary to borrow information across multiple subjects while retaining some commonality features. Preliminary time series models based on the sliding window technique (Lindquist et al., 2014) and asymptotic tests are based on a single subject and do not naturally extend to the case when multiple subjects are concerned.

In the following, we extend our hierarchical model in (5) to include the most simple departure from stationarity, which is accommodating a single change point in the mean of the Wishart distribution. Focusing on one action, the hierarchical change point model across individuals is

\[
S^{(1)}_i \sim W(\phi_1, \phi_1^1 \Omega_1), t = 1, \ldots, c_i
\]

\[
S^{(1)}_i \sim W(\phi_2, \phi_2^2 \Omega_2), t = c_i + 1, \ldots, T
\]

where \(c_i\) represents the change point specific to subject \(i\). We used scaled Wishart distributions with different individual specific means \(\Omega_1\) and \(\Omega_2\) before and after the change points, respectively. A similar orthogonal factor model type decomposition (discussed in (2)) is proposed on \(\Omega_1\) and \(\Omega_2\).

\[
\Omega_1 = V_1 D_1 V_1^T + \sigma_1^2 I_p, \quad \Omega_2 = V_2 D_2 V_2^T + \sigma_2^2 I_p, \quad i = 1, \ldots, n.
\]

Observe that the orthogonal matrices \(V_1\) and \(V_2\) are fixed across individuals and thus viewed as a common dictionary on which individual specific loadings \(D_1\) and \(D_2\) act on to create subject specific deviations. We place independent uniform prior distributions on the Stiefel manifold for \(V_1\) and \(V_2\) along with an independent global-local prior on \(D_1\) and \(D_2\) exactly as in Section 3.1. Independent inverse-gamma priors are chosen on the \(\sigma_1^2\) and \(\sigma_2^2\). We assumed that a priori any time point is equally probable to be a change-point, that is,

\[
c_i \sim \text{Discrete-Uniform}\{1, \ldots, T\}.
\]

FIGURE 7  Left panel represents the boxplot of \(\Omega_i^{[2,3]}, i = 1, \ldots, 9\), which are basically FC between ROI 2 and ROI 3 among healthy controls (HC) and supernormals (SN). Right panel represents the same between ROI 3 and ROI 6.
An efficient Gibbs sampler is developed mimicking Section 3.1 with an additional step to update the change-points \( c_i, i = 1, \ldots, n \). A detailed calculation of the steps is provided in the Supporting Information (Section S7). We discussed a dynamic extension of the hierarchical change model (8) in Section S3 of the Supporting Information.

### 4.1 Simulation study for hierarchical change point model

To demonstrate the the hierarchical change point model (8) on simulated datasets, we consider \( n = 100, p = 50 \) and \( T = 26 \) with \( n_2 = 40 \) individuals having change-points \( c_{10} \in \{2, \ldots, T - 1\} \) and the remaining individuals with size \( n_1 = 60 \) having no change points. For simplicity and to develop a simulation scenario analogous to the HCP dataset, we assume all the individuals are observed at the same time points and the boundary points cannot be considered as a candidate for a change-point. For clarity of exposition, any parameter with subscript “1” corresponds to the pre-change-point regime (deemed as Group 1) and the ones with subscript “2” corresponds to the post-change-point (Group 2).

True individual specific ranks are generated from a discrete uniform distribution spanning over \( \{1, \ldots, r^* = 10\} \). The true values of the diagonal matrices \( \{D_{01}\}_{n_1}^{n} \) and \( \{D_{02}\}_{n_1}^{n} \) are generated from uniform(0,0.5) to include a wide range of signal strengths. \( \{\sigma_{01}^2\}_{i}^{n} \) and \( \{\sigma_{02}^2\}_{i}^{n} \) are generated from uniform(0.25,0.50). Using these values, \( \Omega_1 \) and \( \Omega_2 \) are constructed using Equation (9) and we set \( (\phi_1, \phi_2) = (p+1, p+1) \); 100 replicated datasets are then generated from (8).

The MCMC is run for 5000 iterations leaving a burn-in sample of 5000. Subject-specific change point estimates \( \hat{c}_i \) are obtained from the posterior mode of \( c_i \). Since the focus of this section is on correct detection of change-points, we only display the estimated change-points corresponding to the \( n_2 = 40 \) individuals in Figure 8. Our proposed model is successful to recover individual specific change points. The ranks corresponding to the covariance matrices across individuals are also estimated correctly in all the cases as presented in Figure 9.

To demonstrate consistency of the estimate of \( \Omega_j, j = 1, 2; i = 1, \ldots, n \) with an increasing sample size, we consider another simulation setting where \( p \) is fixed at 50 and \( n \) takes values in the range \( \{100, 200, 300, 400, 500\} \) with \( n_2 \in \{40, 80, 120, 160, 200\} \). Figure 10 presents the summary of the variability of the parameters \( (\sigma_{11}, \sigma_{12}, \sigma_{21}, \sigma_{22}) \) appropriately summarized for the \( n \) individuals using the metrics \( d_{11} \) and \( d_{12} \) over 100 simulated replicates. It is evident that on an average \( d_{11} \) (Figure 10a) and \( d_{12} \) (Figure 10c) decrease with a smaller spread with increasing \( n \) and \( n_2 \), respectively. Similarly, the density plots of \( d_{11} \) (Figure 10b) and \( d_{12} \) (Figure 10d) become more concentrated as \( n \) increases. A detailed study of sensitivity and robustness of hierarchical change point model (8) is provided in Section S4 of the Supporting Information.

### 5 REAL DATA ANALYSIS FOR HCP DATASET

In this section, we consider the HCP dataset (Van Essen et al., 2013) as discussed in Section 2. Time series of covariance matrices describing the connectivity were acquired from each subject while they were performing different tasks involving different neural systems, under a resting state or external stimuli. A quick exploratory analysis of the dataset shows the wide variation in the range of values of the covariance matrices. For the change point model (8) to be applicable, we scale each covariance matrix by the lowest singular value of that matrix as a simple variance stabilizing.
transformation. Based on empirical validation from Figure 2 on small effective ranks of the covariance matrices, we applied the hierarchical change point model (8). Task-specific summary of findings is provided below.

5.1 | Case study for motor task

The HCP motor task experiment was set up by Buckner et al. (2011). Participants are presented with visual cues that ask them to either tap their left or right fingers, or squeeze their left or right toes, or move their tongue to map motor areas. In the experiment, there are 13 blocks, with four hand movements, four foot movements, and two tongue movements. In addition, there are three 15-second fixation blocks between different tasks. We identified 10 cortical ROIs related to the motor control around the motor strip area, including left and right postcentral gyrus, precentral gyrus, and central gyrus, and generated a $10 \times 10$ covariance matrix time series with 26 time points. The proposed hierarchical change point detection model is then applied with a fitted rank $= 5$ and change points are observed for 36 individuals. Table S8 in the Supporting Information shows the individuals with change points for different values of the fitted rank. Figure 11a shows the 36 labeled individuals from the first column (fitted rank $= 5$) of Table S8 with their corresponding most dominant change points. The histogram in Figure 11b displaying the pattern of the change points across the individuals shows that most of the individuals have change points at time point 23. In the experiment design, this corresponds to the time point of switching the movement from hand and foot to the tongue. We applied our methodology on the gambling task as well. A discussion on the findings is deferred to the Supporting Information (Section S5).

One obvious limitation of (8) is that it can only account for the most dominant change point. It is possible that there exist more than one change point for a specific individual under a certain task. In the following, we extended the methodology to enable detection of multiple change points.
Multiple change point analysis

Our hierarchical change point model detects the most dominant change points along the time frame. We adapted a standard sliding window approach to detect multiple change points for different individuals. Denote by \( c_i(1 < c_i < T) \) the first most dominant change point in the interval \( \{1, \ldots, T\} \) for individual \( i \), which is detected through the hierarchical change point model. We slide our time window before and after the most dominant change points to capture any additional changes.

**FIGURE 10** Left panel: Boxplot of \( d_{\Omega_1} \) (upper panel) and \( d_{\Omega_2} \) (lower panel) across simulated replicates for increasing \( n \) and \( n_2 \) respectively. Right panel: Density plots of \( d_{\sigma_1} \) (upper panel) and \( d_{\sigma_2} \) (lower panel) w.r.t. varying \( n \) and \( n_2 \), respectively.

**FIGURE 11** (a) Heat map of a binary matrix consisting of 1 to \((i,j)\)th position that corresponds to the \( i \)th labeled individual and the \( j \)th (\( j = 1, \ldots, 26 \)) time point, which is a change point for the corresponding individual and 0 otherwise. The heat map was made with individuals consisting of change points under the motor task. Individuals are labeled on both sides of the y-axis. (b) Histogram of change points under motor task, which indicates most of the individuals have change points at 23.

### 5.2 Multiple change point analysis

Our hierarchical change point model detects the most dominant change points along the time frame. We adapted a standard sliding window approach to detect multiple change points for different individuals. Denote by \( c_i(1 < c_i < T) \) the first most dominant change point in the interval \( \{1, \ldots, T\} \) for individual \( i \), which is detected through the hierarchical change point model. We slide our time window before and after the most dominant change points to capture any additional changes.
We note here that applying our change-point model over a time window containing $c_i$ recovers the $c_i$ as the most dominant change point. Hence we consider the windows $\{1,...,c_i-1\} \cup \{c_i+1,...,T\}$ for further detection of the next dominant change point. Suppose, there is a change point $c_i^*$ in the interval $\{1,...,c_i-1\}$. Then we again split the time window into $\{1,...,c_i^*-1\} \cup \{c_i^*+1,...,c_i-1\}$ and apply the change point detection method to the two intervals separately. The same procedure is followed on the time window $\{c_i+1,...,T\}$. Figure 12 shows the individuals specific multiple change points under the motor task where we considered individuals with at least two change points. We detected 21 individuals with multiple change points under the motor task, which is shown in Figure 12. There is no individual under the motor task with more than four change points.

6 | DISCUSSION

To discover patterns within the connectivity matrix of human brain as subjects perform specific tasks, we start with a simple Wishart distribution with an approximate low rank structure on the mean for modeling the covariance valued data. The methodology allows straightforward extension to a hierarchical model of multiple subjects where covariance valued time series is available for each subject. Another important extension is to develop a method for detecting a single change point in the covariance time series. Applying the methodology to the HCP data for the motor task reveals that the change point is associated with a particular regime switch of the experimental design. Also, the application to the resting state individuals in the ADNI study does not reveal any change point, which is in accordance with the expert opinions.

Another interesting application related to the HCP dataset is where the subjects are performing psychometric tasks and the goal is to understand how the connectivity evolves over time and whether a particular pattern in the time series motif is associated with the subjects’ “intelligence” or mental ability. In this case, the goal is to understand how the connectivity changes with time and it is important to allow more complex time varying structure in the evolution of the covariance matrix. Such applications also call for development of the joint model of the mental ability scores and the connectivity matrices and is an interesting topic for future research.

For simplicity, we focused on a single Wishart distribution as a model for the covariance value data. A more flexible alternative beyond the Wishart family is to consider a mixture of Wishart distributions, particularly to allow for departures that are not captured by a single scale parameter. However, this comes with an additional burden of identifying and interpreting the component specific mean parameters that are required to be properly regularized to get a meaningful inference.

CONFLICT OF INTEREST STATEMENT

The authors declare no potential conflict of interests.

DATA AVAILABILITY STATEMENT

The data was available from a reference paper (Dai, Zhang, & Srivastava 2017).
Lin, F., Ren, P., Mapstone, M., Meyers, S. P., Porsteinsson, A., Baran, T. M., & Initiative, A. D. N. (2017). The cingulate cortex of older adults with excellent memory capacity. *Cortex, 86*, 83–92.

Lindquist, M. A., Xu, Y., Niebel, M. B., & Caffo, B. S. (2014). Evaluating dynamic bivariate correlations in resting-state FMRI: A comparison study and a new approach. *NeuroImage, 101*, 541–546.

Lopes, H., & West, M. (2004). Bayesian model assessment in factor analysis. *Statistica Sinica, 14*, 41–67.

Miller, K. L., Alfaro-Almagro, F., Bangerter, N. K., Thomas, D. L., Yacoub, E., Xu, J., Bartsch, A. J., Jbabdi, S., Sotiropoulos, S. N., Andersson, J. L. R., & Griffant, L. (2016). Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nature neuroscience, 19*(11), 1523.

Monti, R. P., Hellyer, P., Sharpe, D., Leech, R., Anagnostopoulou, C., & Montana, G. (2014). Estimating time-varying brain connectivity networks from functional MRI time series. *NeuroImage, 103*, 427–443.

Park, H.-J., & Friston, K. (2013). Structural and functional brain networks: From connections to cognition. *Science, 342*(6158), 1238411.

Pati, D., Bhattacharya, A., Pillai, N. S., & Dunson, D. (2014). Posterior contraction in sparse Bayesian factor models for massive covariance matrices. *The Annals of Statistics, 42*(3), 1102–1130.

Polson, N. G., & Scott, J. G. (2012). On the half-Cauchy prior for a global scale parameter. *Bayesian Analysis, 7*(4), 887–902.

Pourahmadi, M. (2011). Covariance estimation: The GLM and regularization perspectives. *Statistical Science, 26*(3), 369–387.

Pourahmadi, M., Daniels, M. J., & Park, T. (2007). Simultaneous modelling of the Cholesky decomposition of several covariance matrices. *Journal of Multivariate Analysis, 98*(3), 568–587.

Schott, J. R. (1999). A test for proportional covariance matrices. *Computational statistics & data analysis, 32*(2), 135–146.

Schott, J. R. (2001). Some tests for the equality of covariance matrices. *Journal of statistical planning and inference, 94*(1), 25–36.

Smith, S. M., Nichols, T. E., Vidaurre, D., Winkler, A. M., Behrens, T. E. J., Glasser, M. F., Ugurbil, K., Barch, D. M., Van Essen, D. C., & Miller, K. L. (2015). A positive-negative mode of population covariation links brain connectivity, demographics and behavior. *Nature Neuroscience, 18*(11), 1565–1567.

Stingo, F. C., Guindani, M., Vannucci, M., & Calhoun, V. D. (2013). An integrative Bayesian modeling approach to imaging genetics. *Journal of the American Statistical Association, 108*(503), 876–891.

Turner, R. (2016). Uses, misuses, new uses and fundamental limitations of magnetic resonance imaging in cognitive science. *Philosophical Transactions of the Royal Society B: Biological Sciences, 371*(1705), 20150349.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., & Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage, 15*(1), 273–289.

Van der Pas, S. L., Kleijn, B. J. K., & van der Vaart, A. W. (2014). The horseshoe estimator: Posterior concentration around nearly black vectors. *Electronic Journal of Statistics, 8*, 2585–2618.

Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E. J., Yacoub, E., Ugurbil, K., & Consortium, W-MHCP (2013). The WU-Minn human connectome project: An overview. *Neuroimage, 80*, 62–79.

Vershynin, R. (2012). Introduction to the non-asymptotic analysis of random matrices. In Y. Eldar, & G. Kutyniok (Eds.), *Compressed Sensing, Theory and Applications* (pp. 210–268). Cambridge University Press.

Warnick, R., Guindani, M., Erhardt, E., Allen, E., Calhoun, V., & 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.

Weiner, M. W., Aisen, P. S., Jack Jr, C. R., Jagust, W. J., Trojanowski, J. Q., Shaw, L., Saykin, A. J., Morris, J. C., Cairns, N., Beckett, L. A., & Toga, A. (2010). The Alzheimer’s disease neuroimaging initiative: Progress report and future plans. *Alzheimer’s & Dementia, 6*(3), 202–211.

Yu, P. L. H., Li, W. K., & Ng, F. C. (2017). The generalized conditional autoregressive Wishart model for multivariate realized volatility. *Journal of Business & Economic Statistics, 35*(4), 513–527.

Zhang, Z., Allen, G., Zhu, H., & Dunson, D. (2018). Relationships between human brain structural connectomes and traits. bioRxiv 256933.

**SUPPORTING INFORMATION**

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Acharyya, S., Zhang, Z., Bhattacharya, A., & Pati, D. (2023). Bayesian hierarchical modeling on covariance valued data. *Stat, 12*(1), e534. [https://doi.org/10.1002/sta4.534]