



Modelling variability in functional brain networks using embeddings

Rukuang Huang, Chetan Gohil, Mark Woolrich

Oxford Centre for Human Brain Activity (OHBA), Wellcome Centre for Integrative Neuroimaging, Department of Psychiatry, University of Oxford, Oxford, United Kingdom

Corresponding Author: Rukuang Huang (rukuang.huang@psych.ox.ac.uk)

ABSTRACT

Functional neuroimaging techniques allow us to estimate functional networks that underlie cognition. However, these functional networks are often estimated at the group level and do not allow for the discovery of, nor benefit from, subpopulation structure in the data, that is, the fact that some recording sessions may be more similar than others. Here, we propose the use of embedding vectors (c.f. word embedding in Natural Language Processing) to explicitly model individual sessions while inferring networks across a group. This vector is effectively a “fingerprint” for each session, which can cluster sessions with similar functional networks together in a learnt embedding space. We apply this approach to estimate dynamic functional networks using a hierarchical Hidden Markov Model (HMM). We call this approach HIVE (HMM with Integrated Variability Estimation). Using simulated data, we show that HIVE can uncover true subpopulation structure and show improved performance over existing approaches. Using real magnetoencephalography data, we show the learnt embedding vectors (session fingerprints) reflect meaningful sources of variation across a population. Overall, HIVE provides a powerful new approach for modelling individual sessions while leveraging information available across an entire group.

Keywords: functional connectivity, population modelling, unsupervised learning, generative modelling, Bayesian modelling

1. INTRODUCTION

Functional connectivity (FC), which is defined as the temporal correlation between spatially remote regions (Friston, 1994), is a popular tool used to study neuroimaging data. Many previous FC studies have used time-averaged, or static, estimates of FC to identify functional networks in tasks and at rest (Biswal et al., 1995; Brookes et al., 2011; Filippini et al., 2009; Fox & Raichle, 2007; Honey et al., 2009). However, given the dynamic nature of brain activity (Rabinovich et al., 2012) and the fact that a considerable amount of within-subject variation in FC has been observed (Honey et al., 2009; Meindl et al., 2010; Van Dijk et al., 2010), evidence suggests that FC changes with time and under the influence of task (Esposito et al., 2006; Fornito et al., 2012). Consequently,

there is growing interest in the study of dynamic functional networks (Baker et al., 2014; Betti et al., 2013; Brookes et al., 2014; Brovelli et al., 2017; Gohil et al., 2022; Quinn et al., 2018; Vidaurre et al., 2016), which has led to several advancements in the modelling of dynamic FC with methods such as the sliding window approach combined with clustering (Allen et al., 2014; Chang & Glover, 2010), the Hidden Markov Models (the HMMs, Baker et al., 2014; Vidaurre et al., 2016), and Dynamic Network Modes (DyNeMo, Gohil et al., 2022) being proposed.

A key consideration that is often overlooked in the modelling of FC is that there is a considerable amount of inter-subject variability. A common assumption in both static and dynamic FC methods is that the same net-

Received: 31 July 2024 Revision: 1 February 2026 Accepted: 24 February 2026 Available Online: 11 March 2026



The MIT Press

© 2026 The Authors. Published under a Creative Commons Attribution 4.0 International (CC BY 4.0) license.

Imaging Neuroscience, Volume 4, 2026
<https://doi.org/10.1162/IMAG.a.1188>

work, or set of networks, is shared by all subjects, that is, the analysis is done at the group level. This is motivated by the fact that there is limited data per subject. Pooling over as much data as possible (across subjects) improves the estimate of FC by averaging out noise. However, implicitly this means inter-subject variability is considered as noise in the estimate of the group average. Although it is possible to post-hoc estimate the subject-specific FC, for example, dual regression of group independent component analysis (ICA, Beckmann et al., 2009; Nickerson et al., 2017) and dual estimation of the Hidden Markov Model (HMM, Vidaurre et al., 2021), these methods treat each individual independently from other individuals and ignores the relationship with other individuals. Principled hierarchical Bayesian models such as PROFUMO (Farahibozorg et al., 2021; Harrison et al., 2015) improve upon this by jointly inferring group and subject parameters. However, they typically model variability as statistical deviations from a group mean. Our goal is distinct: we aim to explicitly model the latent structure of this variability. By embedding subjects into a low-dimensional space, our approach captures the pairwise relationships and subpopulation structure between individuals, rather than treating them solely as independent samples from a group prior.

Functional neuroimaging data are known to possess significant heterogeneity related to, for instance, intrinsic differences in FC due to demographics (Gohil et al., 2024; Quinn et al., 2024) and systematic differences due to scanner type (e.g., CTF vs Elekta MEG scanners) or site (Yamashita et al., 2019). Disentangling these sources of variability is a key challenge. We wish to identify the particular sources of variability that are of interest, such as the differences in intrinsic FC due to demographics, and remove those that are trivial, such as the scanner type. The intrinsic differences in FC are useful for individualized predictions in downstream tasks such as the development of biomarkers for disease or response to intervention (Fedota & Stein, 2015; Philips et al., 2017; J. J. Taylor et al., 2021; Vidaurre et al., 2021).

Crucially, what is still missing in the current literature is a generative approach that explicitly models the structure of inter-subject variability, rather than treating it solely as statistical deviation from a group average (Farahibozorg et al., 2021; Harrison et al., 2015). While hierarchical models can constrain subject estimates using group priors, they generally do not capture the latent geometric relationships between subjects (such as subpopulations or continuous trends). Our proposed framework in this paper addresses this gap by embedding each subject or session into a shared latent space that captures inter-session/subject similarities and differences. By utilising a hierarchical model to explicitly specify the generative

process of individual-level networks from group-level networks, we enable joint inference of individual variability and group-level structure. We describe the framework and its application to the HMM in detail in Section 2 and illustrate its use in Section 3.

2. METHODS

2.1. Previous methods

2.1.1. Hidden Markov modelling

The basic assumption in the HMM is that the brain transitions between states. Each state has different spatio-temporal characteristics and is represented by a multivariate Gaussian distribution. The covariance matrix of the multivariate Gaussian distribution characterises the interaction between different channels/parcels and encodes information about FC. The generative model (SI Section A.1.1) describes how the time series is generated given a set of model parameters (state means, state covariances, and transition probability matrix). Through training, we infer the parameters that best generate the observed data and partition the time series into a finite number of mutually exclusive states.

2.1.2. Dual estimation

When training the HMM, training data are created by concatenating the time series of different sessions. Therefore, the inferred covariance matrix for each state is shared across all sessions. This means the dynamic FC is estimated at the “group level” and the HMM itself does not provide any estimates of individual session networks. Individual estimates of FC can be obtained by using post-hoc analysis of the HMM—a method called *dual estimation* (Vidaurre et al., 2021), which has a similar rationale as dual regression in Independent Component Analysis (ICA, Beckmann et al., 2009; Nickerson et al., 2017). The idea is for each session, we fix the state probabilities from a trained HMM and re-estimate the covariances with the session’s data only. In this paper, we use HMM-DE (HMM with dual estimation) to refer to the pipeline of training a group-level HMM followed by dual estimation. See SI Section A.1.2 for details of performing dual estimation.

It should be noted that due to structures within a population, some sessions may share more similar FC networks than others. However, dual estimation treats each session independently from other sessions and ignores the relationships between different sessions. As a result, it does not allow the inference of dynamic FC networks to pool information across other sessions with similar networks within the group.

2.1.3. Hierarchical models

An alternative to overcome this is to use hierarchical models that capture variability across sessions. In a hierarchical model, parameters are organised into different levels of hierarchy, for example, group and individual level. Individual-level estimates of functional networks are assumed to be generated by a single probabilistic distribution whose parameters are the group-level estimates of functional networks. Hence, the information of functional networks of different sessions is shared through the group level. For example, PROFUMO (Harrison et al., 2015, 2020) has been successfully used to capture aspects of subject variability in functional modes in fMRI data. However, variability within a population is modelled using very simple parametric distributions (e.g., univariate Gaussian), which fails to capture the rich, multivariate variability expected across sessions.

2.2. Outline

Here, we extend the HMM to hierarchically model session-specific variability in FC networks. We call this extension HIVE (HMM with integrated variability estimation). Instead of having covariances shared across sessions, we assume that data from different sessions are generated by different sets of covariances, which respects the underlying subpopulation structure, that is, data from similar sessions are generated by similar covariances and vice versa. In what follows, we outline the generative model of HIVE and inference of model parameters. We also describe the datasets studied in this work.

2.3. Generative model

Notations

- $[n] = \{1, \dots, n\}$ for $n \in \mathbb{N}$.
- \mathbf{x} is a column vector.
- \mathbf{A} is a matrix.
- $x_{t:n} = \{x_i : i \in [n]\}$.

Let \mathbf{x}_t^i be the observed data at time $t \in [T_i]$ for session $i \in [N]$, where N is the number of sessions, T_i is the total number of time points of session i , and \mathbf{x}_t^i is a vector of length N_c — the number of channels/parcels. It is assumed that the data are generated by

$$\mathbf{x}_t^i | (s_t^i = j) \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_j^i) \quad \forall t \in [T_i], \forall i \in [N], \quad (1)$$

independently. Here, $\boldsymbol{\Sigma}_j^i$ is the covariance matrix that describes the spatio-spectral pattern of session $i \in [N]$ and state $j \in [J]$, where J is the number of states, and the hidden states $s_{t:T_i}^i$ for each session $i \in [N]$ follow a dis-

crete homogeneous Markov process (Norris, 1998), that is,

$$p(s_{t:T_i}^i) = p(s_1^i) \prod_{t=2}^{T_i} p(s_t^i | s_{t-1}^i), \quad (2)$$

where $p(s_t^i = j | s_{t-1}^i = j') = \mathbf{A}_{j'j}^i$ does not depend on time and is the (j', j) -th entry to the transition probability matrix \mathbf{A}^i . In this work, we assume that all sessions have the same transition probability matrix $\mathbf{A}^i = \mathbf{A} \forall i \in [N]$. Next, we hope to model the inter-session relationships between the covariances $\boldsymbol{\Sigma}_j^i$ and this is achieved by the use of embedding vectors.

2.3.1. Embedding vectors

In this paper, we employ embedding vectors, a technique widely used in the Natural Language Processing literature for characterising semantic relationships between words (Mikolov et al., 2013), to explicitly model between-session variability in the basis set of networks used to describe the functional brain activity.

A word embedding is a real-valued vector that encodes the meaning of the word such that words closer in the vector space have similar meanings. In Figure 1a, we show an example of word embeddings. Here, the x-axis encodes whether a word is for animals or non-animals and the y-axis encodes whether the object can fly or not. Similarly, we can assign a vector for each recording session and hope that sessions with similar properties are close to each other in the vector space. In Figure 1b, we show an illustration of possible variability captured by embedding vectors. Here, the x-axis encodes the direction of increasing beta power in motor network whereas the y-axis encodes the direction of increasing peak alpha frequency in visual network.

Embedding vectors have been previously used in computational neuroscience literature to deal with between-subject variability in supervised (Chehab et al., 2021; Csaky et al., 2023) and self-supervised learning (Défossez et al., 2023; Jayalath et al., 2024) contexts. Here, we try to incorporate it into an unsupervised generative model of network activity. Specifically, we propose the *variability encoding block*, which utilises the technique of embedding vectors and hierarchical Bayesian Modelling (Gelman et al., 1995), and apply it to the generative model of the HMM. Conceptually, each session is assigned an embedding vector that acts as a “fingerprint” of this session and contains session-specific information. A function, the variability encoding block, is used to decode the abstract session-specific information hidden in the embedding vectors to the session-specific functional networks. Formally, let $\boldsymbol{\omega}^i$ be the embedding vector

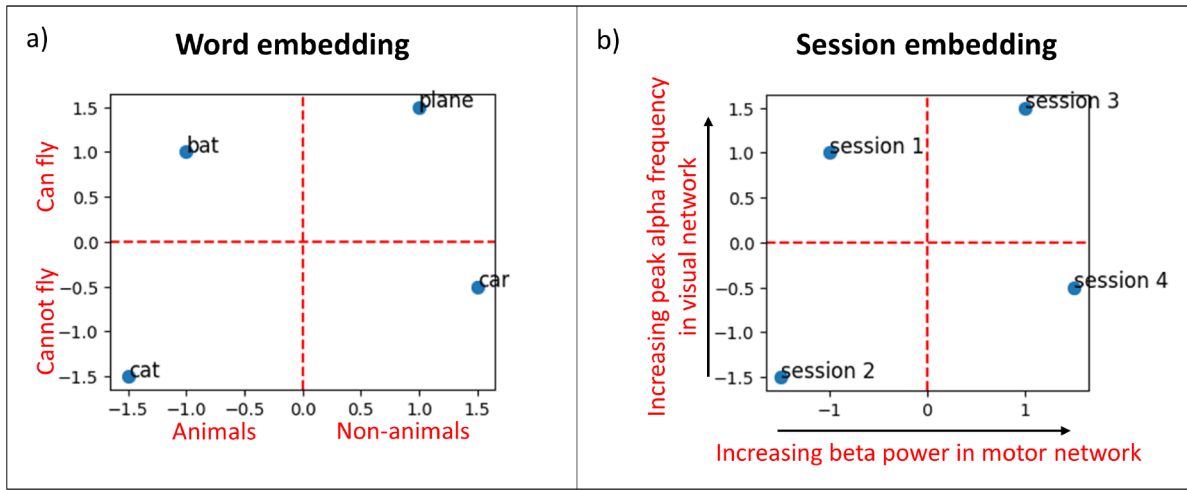


Fig. 1. Examples of embedding vectors. (a) Example of word embedding vectors. The x-axis encodes whether a word is for animals or non-animals, and the y-axis encodes whether the object can fly or not. (b) Illustration of possible variability captured by embedding vectors in brain networks found in electrophysiological data such as M/EEG. The x-axis encodes the direction of increasing beta power in motor network, whereas the y-axis encodes the direction of increasing peak alpha frequency in visual network.

of session i , a vector of length n_ω , where n_ω is called the embedding dimension. Furthermore, let Σ_j be the group-level covariance matrix of state j . Then, the variability encoding block takes both of ω^i and Σ_j as inputs and outputs Σ_j^i . The generative model of HIVE is summarised in Figure 2.

The generative model describes how data are generated from the model parameters and through model training, and the optimisation process will find the optimal parameters (including embedding vectors) that best describe the data. Ideally, data with similar characteristics are generated from similar embedding vectors and the training process (described in Section 2.4) will group together embedding vectors from similar recording sessions in order to minimise the loss function in Equation (13).

2.3.2. The variability encoding block

In this section, we outline the generative process of session-specific covariances Σ_j^i with the variability encoding block. To ensure legitimate covariance matrices are generated, that is, symmetric and positive definite, we work in the Cholesky space. Let $\mathbf{L}_j, \mathbf{L}_j^i$ be the Cholesky factors (with positive diagonal entries) of Σ_j, Σ_j^i respectively, and furthermore, let $\mathbf{l}_j, \mathbf{l}_j^i$ be the vectors of lower triangular entries of $\mathbf{L}_j, \mathbf{L}_j^i$. Then, there is a bijection between the covariance matrices Σ_j, Σ_j^i and the Cholesky vectors $\mathbf{l}_j, \mathbf{l}_j^i$.

The goal of the variability encoding block is to model the session-specific Cholesky vectors \mathbf{l}_j^i as deviations from the group-level Cholesky vectors \mathbf{l}_j , such that the deviations depend on the session-specific information

provided by the session embedding vectors ω^i as well as the state-specific information provided by the group-level Cholesky vectors \mathbf{l}_j . Formally, for each session i and each state j , we assume

$$\mathbf{l}_j^i = \mathbf{l}_j + d_j^i \mathbf{e}_j^i, \quad (3)$$

where d_j^i is a positive scalar representing the magnitude of the deviation and \mathbf{e}_j^i is a standardised vector representing the pattern of deviations. Notice here the group-level Cholesky vectors \mathbf{l}_j and the session-specific deviations $d_j^i \mathbf{e}_j^i$ are not identifiable, i.e. there are infinitely many combinations that yield the same session-specific Cholesky vectors \mathbf{l}_j^i . Ideally, we prefer the solution where the group-level Cholesky vectors are the ‘‘average’’ of the session-specific Cholesky vectors, that is, we prefer solutions where there is smallest total deviation, while enforcing positivity of this quantity. Hence, we put an exponential prior on d_j^i :

$$d_j^i \sim \text{Exp}(\lambda_j^i). \quad (4)$$

The rate parameter λ_j^i and deviation pattern \mathbf{e}_j^i are generated through

$$\begin{aligned} \lambda_j^i &= \rho\left(f\left(h\left(\xi_j^i\right)\right)\right), \\ \mathbf{e}_j^i &= \eta\left(g\left(h\left(\xi_j^i\right)\right)\right), \end{aligned} \quad (5)$$

where ρ is the softplus function to ensure positivity of λ_j^i and η is a Layer Normalisation layer (Ba et al., 2016) with

HIVE Generative Model

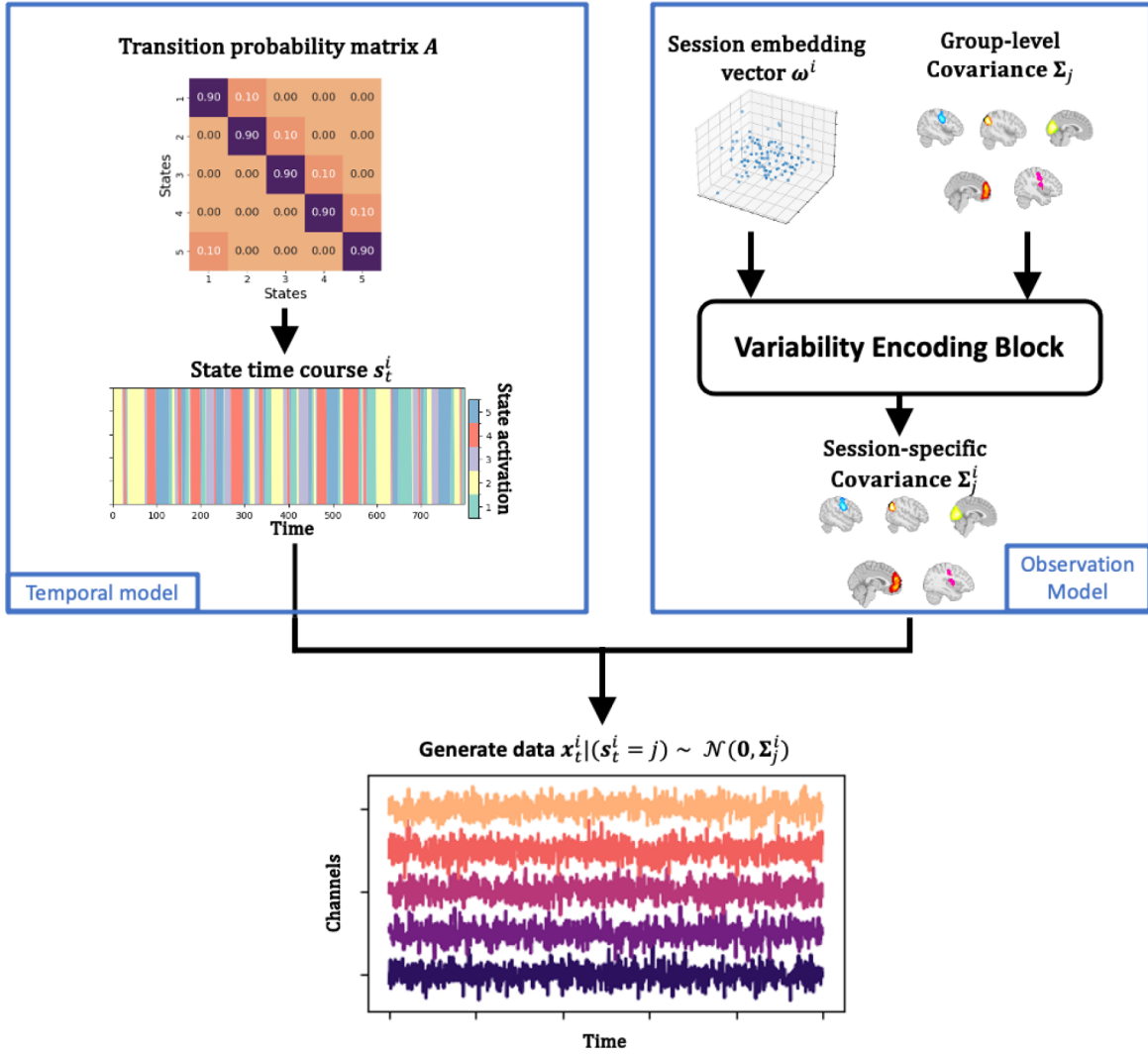


Fig. 2. Generative model of HIVE. State time courses are generated from the group-level transition probability matrix \mathbf{A} through a Markov process. At the same time, session embedding vectors ω^i and group-level covariance Σ_j are passed to the variability encoding block to give session-specific covariance Σ_j^i . Given the state s_t^i at each time, the observed data are generated through the multivariate Gaussian likelihood.

a non-trainable scale parameter of 1 to ensure unit standard deviation of \mathbf{e}_j^i . Affine functions f, g are learnable transformations that extract different information from the hidden state given by the decoder h , a multi-layer perceptron (MLP, Popescu et al., 2009). ξ_j^i is the key object called a concatenated embedding vector that encodes both session and state-specific information. It is formed by concatenating the session embedding vector ω^i and spatial embedding vector I_j^* , which is a lower-dimensional representation of the group-level Cholesky vectors and has length n_j :

$$\xi_j^i = [\omega^i, I_j^*], \quad I_j^* = k(I_j), \quad (6)$$

where k is an affine transformation that serves as an encoder to encode spatial information. As a result, for

each session i and each state j , ξ_j^i is a vector of length $n_\omega + n_j$. The complete generative process (forward pass) of the variability encoding block is summarised in Figure 3.

2.4. Inference

In this section, we outline the process of inferring the model parameters of HIVE. The model parameters of HIVE, ϕ , include

- the transition probability matrix \mathbf{A} ,
- the embedding vectors $\omega = \{\omega^i : i \in [N]\}$,
- the group-level Cholesky vectors $I = \{I_j : j \in [J]\}$,
- the weights of the encoder k , the decoder h , the layer normalisation layer η , and the affine transformations f, g .

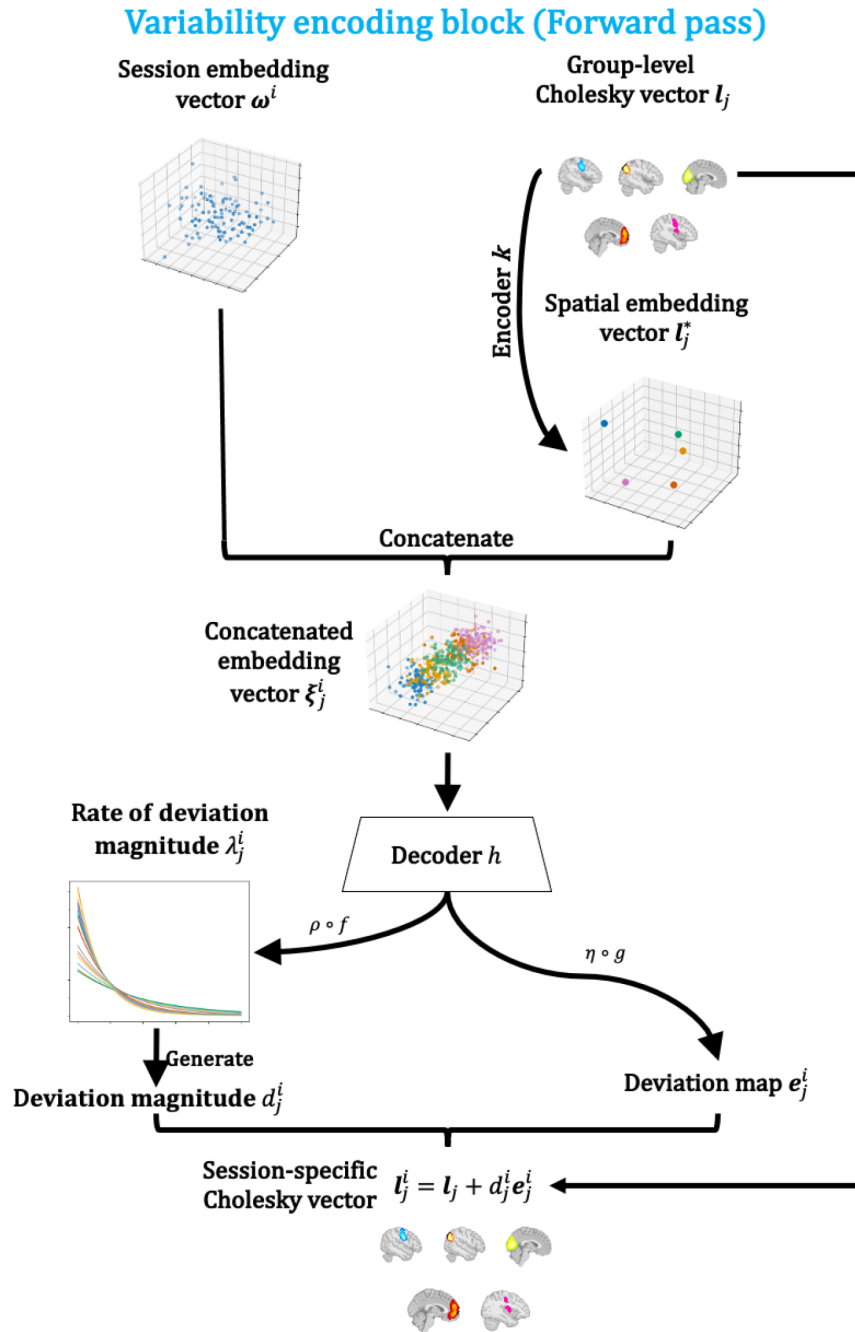


Fig. 3. Forward pass of the variability encoding block. Session-specific information from the session embedding vectors ω^i and state-specific information from the group-level Cholesky vectors l_j are combined to generate session and state-specific Cholesky vectors l_j^i . Here, the encoder k is an affine transformation that condenses state-specific information to a lower-dimensional space. The decoder h is a multi-layer perceptron (MLP) that decodes state and session-specific deviations from the concatenated embedding vectors ξ_j^i . Deviation magnitude d_j^i is generated by an exponential distribution with rate λ_j^i , which is generated by applying an affine transformation f followed by a softplus ρ activation to the output of h . The deviation map e_j^i is generated by applying an affine transformation g and normalisation η to the output of h . Finally, we get the session-specific Cholesky vector l_j^i with $l_j^i = l_j + d_j^i e_j^i$.

To perform inference on these parameters, we employ the EM algorithm (Dempster et al., 1977), and more specifically, a variant of the Baum-Welch algorithm (Baum & Eagon, 1967; Baum & Petrie, 1966).

In practice, due to memory restrictions and scalability, it is infeasible to perform inference based on the entire

sequence of data. Therefore, data are separated into sequences of length T_{seq} , which is much less than $T = \sum_{i=1}^N T_i$ - the total number of data points, so that there are $N_{seq} = \lfloor T / T_{seq} \rfloor$ sequences. For each iteration of the EM algorithm, a random set B of sequences, called a *minibatch*, is used to update the parameters. More

specifically, the sequences are split into $\lfloor N_{seq} / |B| \rfloor$ minibatches and each minibatch is used in turn for updating the model parameters. After all the minibatches have been processed, called an *epoch*, a different randomised set of minibatches is used for the next epoch.

2.4.1. The EM algorithm

During the ‘‘E-step’’, we update the state probabilities (the posterior distribution of the state activations) with the forward-backward algorithm of the Baum-Welch algorithm. During the ‘‘M-step’’, given the state probabilities, we update the transition probability matrix \mathbf{A} with the stochastic update technique used in [Vidaurre et al. \(2018\)](#):

$$\mathbf{A}_{new} = (1 - \zeta) \mathbf{A}_{old} + \zeta \mathbf{A}_{interim}, \quad (7)$$

where \mathbf{A}_{old} and \mathbf{A}_{new} are the transition probability matrices before and after the update, $\mathbf{A}_{interim}$ is the interim update of the transition probability matrix given by the Baum-Welch algorithm, and ζ is the update weight that decreases with training epoch:

$$\zeta(\text{epoch}) = (\text{epoch} + \text{delay})^{-\text{forget}}, \quad (8)$$

where $\text{delay} = 5$ and $\text{forget} = 0.7$ in this work. Next, we need to infer the rest of the parameters in ϕ and notice we do not have access to the exact posterior distribution of the deviation magnitudes $p(\mathbf{d} | \mathbf{x})$, where $\mathbf{d} = \{\mathbf{d}^i : i \in [N]\}$

$= \{d_j^i : i \in [N], j \in [J]\}$ and $\mathbf{x} = \{\mathbf{x}_{t:T_{seq}}^i : i \in [N]\}$ due to the fact that the prior distribution $p(\mathbf{d})$ is parameterised by a complex neural network. Therefore, we approximate $p(\mathbf{d} | \mathbf{x})$ using the variational distribution $q(\mathbf{d})$ with a mean field approximation ([Blei et al., 2017](#)):

$$q(\mathbf{d}) = \prod_{i=1}^N \prod_{j=1}^J q(d_j^i) = \prod_{i=1}^N \prod_{j=1}^J \Gamma(d_j^i | a_j^i, b_j^i). \quad (9)$$

Here, the shape a_j^i and the rate b_j^i parameters of the Gamma distributions Γ are learnable parameters. The reason that Gamma distributions are chosen is that there is an analytic solution for the KL divergence between a Gamma distribution and an Exponential distribution. To sum up, the model parameters ϕ , except the transition probability matrix \mathbf{A} , and the variational parameters $\psi = \{a_j^i, b_j^i : i \in [N], j \in [J]\}$ are inferred jointly by minimising the variational free energy given the state probabilities.

2.4.2. Variational free energy given the state probabilities

While breaking the data into sequences, we ensure that data in each sequence n belong to the same session i_n . Let $\mathbf{x}_{t:T_{seq}}^{(n)}$ be the n -th sequence in the data and similarly for $\mathbf{s}_{t:T_{seq}}^{(n)}$, then the loss function for optimisation for sequence n is the usual variational free energy ([Kingma & Welling, 2013](#)):

$$\begin{aligned} \mathcal{L}_n(\phi, \psi) &= -\text{LL}_n(\phi, \psi) + \text{KL}_n(\phi, \psi) \\ &= -\underbrace{\mathbb{E}_{q(\mathbf{s}_{t:T_{seq}}^{(n)})} \mathbb{E}_{q_\psi(\mathbf{d}^{i_n})} [\log p_\phi(\mathbf{x}_{t:T_{seq}}^{(n)} | \mathbf{s}_{t:T_{seq}}^{(n)}, \mathbf{d}^{i_n})]}_{\text{LL}_n(\phi, \psi)} + \underbrace{\frac{1}{N_{seq}} D_{KL}(q_\psi(\mathbf{d}) \| p_\phi(\mathbf{d}))}_{\text{KL}_n(\phi, \psi)}, \end{aligned} \quad (10)$$

where $q(\mathbf{s}_{t:T_{seq}}^{(n)})$ are the posterior probabilities given by the forward-backward procedure in the Baum-Welch algorithm, D_{KL} is the Kullback-Leibler divergence ([Kullback & Leibler, 1951](#)), and the $\frac{1}{N_{seq}}$ factor is due to the fact that the loss is averaged over sequences in a batch. The $\text{LL}_n(\phi, \psi)$ term acts as a reconstruction loss that describes how well the data are reconstructed by the parameters ϕ, ψ and the $\text{KL}_n(\phi, \psi)$ term acts as a regularisation term that penalises complex variational distributions $q_\psi(\mathbf{d})$ that deviates from the prior $p_\phi(\mathbf{d})$. The $\text{LL}_n(\phi, \psi)$ term can be simplified as

$$\begin{aligned} \text{LL}_n(\phi, \psi) &= \sum_{t=1}^{T_{seq}} \sum_{j=1}^J \mathbb{P}(s_t^{(n)} = j | \mathbf{x}_{t:T_{seq}}^{(n)}) \\ &\quad \mathbb{E}_{q_\psi(\mathbf{d}^{i_n})} \left[\log \mathcal{N}(\mathbf{x}_t^{(n)} | 0, \Sigma_j^{i_n}(\phi, \mathbf{d}_j^{i_n}(\psi))) \right], \end{aligned} \quad (11)$$

where $\mathbb{P}(s_t^{(n)} = j | \mathbf{x}_{t:T_{seq}}^{(n)})$ is the posterior probability that the state at time t of sequence n is j . The $\text{KL}_n(\phi, \psi)$ term can be simplified as

$$\text{KL}_n(\phi, \psi) = \frac{1}{N_{seq}} \sum_{i=1}^N \sum_{j=1}^J D_{KL}(\Gamma(a_j^i, b_j^i) \| \text{Exp}(\lambda_j^i(\phi))). \quad (12)$$

In summary, for a randomly drawn minibatch $B \subseteq [N_{seq}]$, the training involves getting the posterior probabilities of the states with the forward-backward procedure, updating the transition probability matrix \mathbf{A} with the stochastic update technique, and one step of gradient descent type update (in this paper, the ADAM optimiser is used, [Kingma & Ba, 2014](#)) with the loss function

$$\mathcal{L}(\phi, \psi) = \frac{1}{|B|} \sum_{n \in B} \mathcal{L}_n(\phi, \psi), \quad (13)$$

with respect to the parameters ϕ, ψ . As aforementioned, we have the exact formula for the Kullback-Leibler divergence between a Gamma and an Exponential distribution, so that the gradient of $\text{KL}_n(\phi, \psi)$ with respect to ϕ, ψ can be calculated. However, the same cannot be said for the $\text{LL}_n(\phi, \psi)$ term which involves an integral with respect to the variational distribution $q_\psi(\mathbf{d}^i)$. Luckily, we can get an estimate of the gradient with the help of the reparameterisation trick (Kingma & Welling, 2013) for Gamma distributions (Figurnov et al., 2018). The whole training pipeline is implemented in the *osl-dynamics* toolbox (Gohil et al., 2023) with the *Tensorflow* package, which allows trivial calculation of gradients with respect to all the parameters and easy execution of back-propagation. A flowchart summarising the loss calculation is shown in SI section A.1.3 (Fig. A1).

2.4.3. Initialisation of variational parameters

A good initialisation is paramount in training deep neural networks to avoid slow and unstable training. In practice, we found that a good initialisation of the variational parameters ψ is crucial for the training of HIVE, without which the model can either get stuck in local minimum or diverge. We initialise the shape and rate parameters of the Gamma distributions as follows:

$$\begin{aligned} d_{\text{mean}} &= \frac{1}{N} \sum_{i=1}^N |I_{\text{static}}^i - \overline{I_{\text{static}}}| \\ d_{\text{var}} &= \frac{1}{5} d_{\text{mean}} \\ a_j^i &= \frac{d_{\text{mean}}^2}{d_{\text{var}}} \quad \forall i \in [N], j \in [J] \\ b_j^i &= \frac{d_{\text{mean}}}{d_{\text{var}}} \quad \forall i \in [N], j \in [J]. \end{aligned} \quad (14)$$

The idea is to estimate the mean deviation magnitude d_{mean} with the deviation of the Cholesky vectors $\{I_{\text{static}}^i\}_{i=1}^N$ of the static covariance matrices from different sessions and get the parameters so that the resulting Gamma distribution has mean d_{mean} and variance $d_{\text{var}} = \frac{1}{5} d_{\text{mean}}$, which is what we found to be a good initialisation in practice.

2.4.4. Annealing during training

During training, we employed two annealing techniques. The first one is KL annealing (Bowman et al., 2015) where we start the training without the KL term in the loss and gradually increase the contribution of the KL term. We refer the readers to SI Section A.1.4 for more details.

The second one is that we anneal the sampling process of the variational distribution of deviation magnitudes $q(\mathbf{d})$ when applying the reparameterisation trick:

$$\hat{\mathbf{d}} = (1 - \kappa) \bar{\mathbf{d}} + \kappa \mathbf{d}^*, \quad (15)$$

where $\bar{\mathbf{d}}$ is the expectation of $q(\mathbf{d})$, \mathbf{d}^* is a sample from $q(\mathbf{d})$, and κ is the annealing factor that is zero at the beginning of the training and gradually increases to one. We find this significantly improves the convergence of the model and we believe this serves as an exploration-exploitation mechanism, which is extensively studied in the fields of reinforcement learning (Wang et al., 2018) and Bayesian Optimisation (Jalali et al., 2012). At the beginning of training, only the expectation is used and the gradient will intend to push the mean deviations to the correct position. This is the exploration phase. As training progresses, we gradually take into account the variance and higher moments of the variational distribution, which allows the training to fine tune the variational parameters. This is the exploitation phase.

2.5. Choosing the embedding dimension

We need to pre-define the length n_ω of the embedding vectors ω^i . A smaller n_ω will lead to a more parsimonious model but there could be a risk of over-regularisation. On the other hand, a larger n_ω will lead to a more flexible model but could lead to overfitting. In practice, we use the loss function and follow the following procedure. Firstly, we define a set of candidate values for n_ω , say $\{n_\omega^1, \dots, n_\omega^M\}$ where $n_\omega^k < n_\omega^{k+1}, \forall k \in [M-1]$. Then, we train the model with each of these candidates of embedding dimension a number of times independently and for each run, we get the training loss (variational free energy on the training data) of the model. Starting from $k = 1$, we test whether the training loss of the model with embedding dimension n_ω^{k+1} is significantly lower than that with embedding dimension n_ω^k . If it is, we continue to test for n_ω^{k+1} and n_ω^{k+2} . If it is not, we choose n_ω^k as the embedding dimension. The test for significance can be done with a one sided t-test or a non-parametric permutation test. Formally, the procedure is summarised in Algorithm 1.

Algorithm 1: Choosing the embedding dimension

Input: Embedding dimension candidates $\{n_\omega^1, \dots, n_\omega^M\}$

Output: Chosen embedding dimension n_ω^*

Procedure:

for $k \in [M-1]$ **do**

Test the null hypothesis that the training loss of model with n_ω^k is not significantly lower than that with n_ω^{k+1}

if Null hypothesis is not rejected **then**

return n_ω^k

return n_ω^M

2.6. Datasets

2.6.1. Simulated datasets

In the simulation studies, we simulate data with an HMM generative model. To do this, we specify a transition probability matrix and randomly simulate covariance matrices for each state (sessions can have different covariances). Subsequently, a Markov chain is simulated with the pre-specified transition probability matrix. Lastly, at each time point, data are simulated using a multivariate Gaussian distribution with the state covariance matrix which is active at the given time. Gaussian noise is also added to the final time series data (see [SI Section A.1.5](#) for details). The simulated data for the 3 simulation studies are described below.

- **Simulation 1.** We simulate data with $N = 10$, $T_i = 25,600$, $\forall i \in [N]$, $N_c = 11$, $J = 5$. Here, the state covariance matrices of sessions 1 and 2 are altered and all other sessions have the same unaltered group-level covariances—session 1 has increased variance in channel 1 and decreased variance in channel 2 while session 2 has decreased variance in channel 1 and increased variance in channel 2. For both session 1 and 2, the increased variances are 5 times and the decreased variances are 20% of the group-level variance.
- **Simulation 2.** Here, $N = 100$, $T_i = 3,000$, $\forall i \in [N]$, $N_c = 40$, $J = 5$. Sessions are assigned into 3 groups, and an embedding vector for each session is simulated according to the session's assigned group. Session-specific covariances are simulated based on the simulated embedding vectors (see [SI Section A.1.5.3](#) for details). Notice, this is a generalisation of the data in Simulation 1, where one group contains 8 sessions and each of the other 2 groups contain only 1 session, that is, there is no variance within these groups.
- **Simulation 3.** In this study, 100 datasets are simulated, where 10 datasets are simulated for each $N \in \{10k : k = 1, \dots, 10\}$ and $T_i = 3000$, $\forall i \in [N]$, $N_c = 40$, $J = 5$. Session-specific covariances are simulated in the same way as in Simulation 2.

2.6.2. Real MEG data

We demonstrate the use cases of the proposed model with 3 publicly available MEG datasets, including two resting-state and one visual task dataset. The datasets are source-reconstructed and parcellated to 38 regions of interest. We describe the steps of data processing before model training below.

- **Raw data.** The first resting-state dataset ([J. R. Taylor et al., 2017](#), we refer to this dataset as the Cam-CAN dataset) contains eyes-closed data from 612 healthy participants. These data were collected using an Elekta Neuromag Vectorview 306 scanner at a sampling frequency of 1 kHz. A highpass filter of 0.03 Hz and MaxFilter were applied. In the visual task MEG dataset ([Wakeman & Henson, 2015](#), we refer to this dataset as the Wakeman-Henson dataset), each of the 19 health participants were scanned 6 times, during which 3 types of visual stimuli were shown to the participants. The data were also collected using an Elekta Neuromag Vectorview 306 scanner. The second resting-state dataset was collected using a 275-channel CTF scanner. This dataset (we refer to this dataset as the Nottingham dataset) contains eyes-closed data from 64 healthy participants, collected at Nottingham University, UK as part of the MEGUK partnership.
- **Preprocessing.** The Cam-CAN and Nottingham datasets were preprocessed with the same pipeline using the *osl-ephys* package ([van Es et al., 2024](#)). The data were band-pass filtered between 0.5 Hz and 125 Hz to remove high-frequency noise and low-frequency drifts. This is followed by a notch filter at 50 Hz and 100 Hz to remove a known artefact due to power line. Then, the data were downsampled to 250 Hz to reduce computational load. Additionally, automated bad segment and bad channel detection, with the generalised ESD procedure ([Rosner, 1983](#)), were applied to remove abnormally noisy segments and channels of the recording. Finally, an independent component analysis (ICA) with 64 components was applied. Components with high correlation (0.35) with ECG and EoG recordings are removed. The preprocessing of the Wakeman-Henson dataset is the same as above except that in the end, ICA with 40 components was applied.
- **Source reconstruction and parcellation.** Coregistration and source reconstruction were done using OSL. Structural data were coregistered with the MEG data using an iterative close-point algorithm and digitised head points acquired with a Polhemus pen were matched to individual subject's scalp surfaces extracted with FSL's BET tool ([Jenkinson et al., 2005](#); [Smith, 2002](#)). The nose was not included in the coregistration as the structural MRI images were defaced. Preprocessed sensor data were source reconstructed onto an 8 mm isotropic grid using a linearly constrained minimum variance beamformer ([Van Veen & Buckley, 1988](#)).

Voxels were then parcellated into 38 anatomically defined regions of interest, before the symmetric spatial leakage correction described in Colclough et al. (2015) was applied.

- **Data preparation.** Before model training, we follow the preparation steps described in Gohil et al. (2022). The data were time-delay embedded with ± 7 lags. Then, principal component analysis (PCA) was used to reduce the dimensionality to 80 channels before a standardisation step (z-transform) was applied to make sure each channel has zero mean and variance of one.

3. RESULTS

In this section, we validate HIVE, show its use cases, and compare to HMM-DE on both simulated and real data. We have included model specifications and details on hyperparameters for the models trained on different datasets in SI Section A.1.6. Scripts for reproducing the results are also publicly available on github.com/OHBA-analysis/Huang2025_ModelVariabilityWithEmbeddings.

3.1. Simulation

Here, we show the results of training HIVE on the 3 simulated datasets described in Section 2.6.1 and compare these results with those from HMM-DE. We wish to highlight the advantages of HIVE in the following aspects:

- We wish to demonstrate HIVE embeddings reflect inter-subject variability and similarity: Simulation 1 and 2 do this.
- We wish to demonstrate correct state inference with a known ground truth: Simulation 2 does this.
- We wish to demonstrate HIVE infers better individualised networks (state covariances) compared to existing methods (HMM-DE): Simulation 3 does this.

3.1.1. Simulation 1: Variability encoding block learns session-specific covariance deviations

We aim to test if patterns of deviations across multiple channels can be learnt by the generative model, in particular, the variability encoding block, during inference. This is an important feature in the sense that we want the generative model to act as a prior that regularises how each session can deviate from the group. Firstly, we can see from Figure 4a that both HMM-DE and HIVE can infer the group-level covariances accurately.

What HIVE offers in addition is shown in Figure 4b where the inferred session embedding vectors are plot-

ted. We can clearly see that the embedding vectors for sessions 1 and 2 are far away from the cluster formed by other sessions which have unaltered covariances, which matches well with the ground truth. The model's capability to encode deviation pattern across multiple channels in the generative model is illustrated in Figure 4c, in which the simulated patterns of deviation for sessions 1 and 2 can be generated from the trained variability encoding block. Although HMM-DE can infer session-specific covariances via dual estimation, it cannot generate data with patterns that deviate from the group-level average. This is because intrinsically HMM-DE is a group-level model.

3.1.2. Simulation 2: HIVE correctly infers the similarity between sessions

In this simulation study, we aim to show that HIVE can recover the ground truth underlying subpopulation structure (simulated embedding vectors of sessions). This is demonstrated in Figure 5a, where the ground-truth grouping of sessions is recovered by the inferred session embedding vectors. In particular, sessions which are close together in the simulated space (e.g., sessions 7 and 45) stay close in the inferred space. Figure 5b shows both HMM-DE and HIVE can infer the state time courses perfectly. Furthermore, we can see from Figure 5c that both approaches can recover the pairwise session relationship between session-specific covariances, though HMM-DE overestimates the pairwise distance due to noise added to the data, whereas HIVE has the preferable behaviour of underestimating the pairwise distance due to the regularising effect of the prior on the deviation magnitude. This prior makes the inferred session-specific covariances more similar to the group average, when there is insufficient evidence available in the data to do otherwise.

3.1.3. Simulation 3: HIVE improves the estimation of session variability

Now, we focus on comparing HIVE and HMM-DE in terms of accuracy of inferred session-specific covariances. Both HIVE and HMM-DE are trained on each of the 100 simulated datasets. In Figure 6a, the accuracy (correlation with ground-truth session-specific covariances) of the inferred session-specific covariance for each state and each session is plotted against the number of sessions in the dataset. We can see qualitatively HIVE achieves better performance compared to HMM-DE. In particular, HIVE always achieves higher mean and median accuracy than HMM-DE across all number of sessions N . Notably, we can observe that a significant mass of the accuracies from

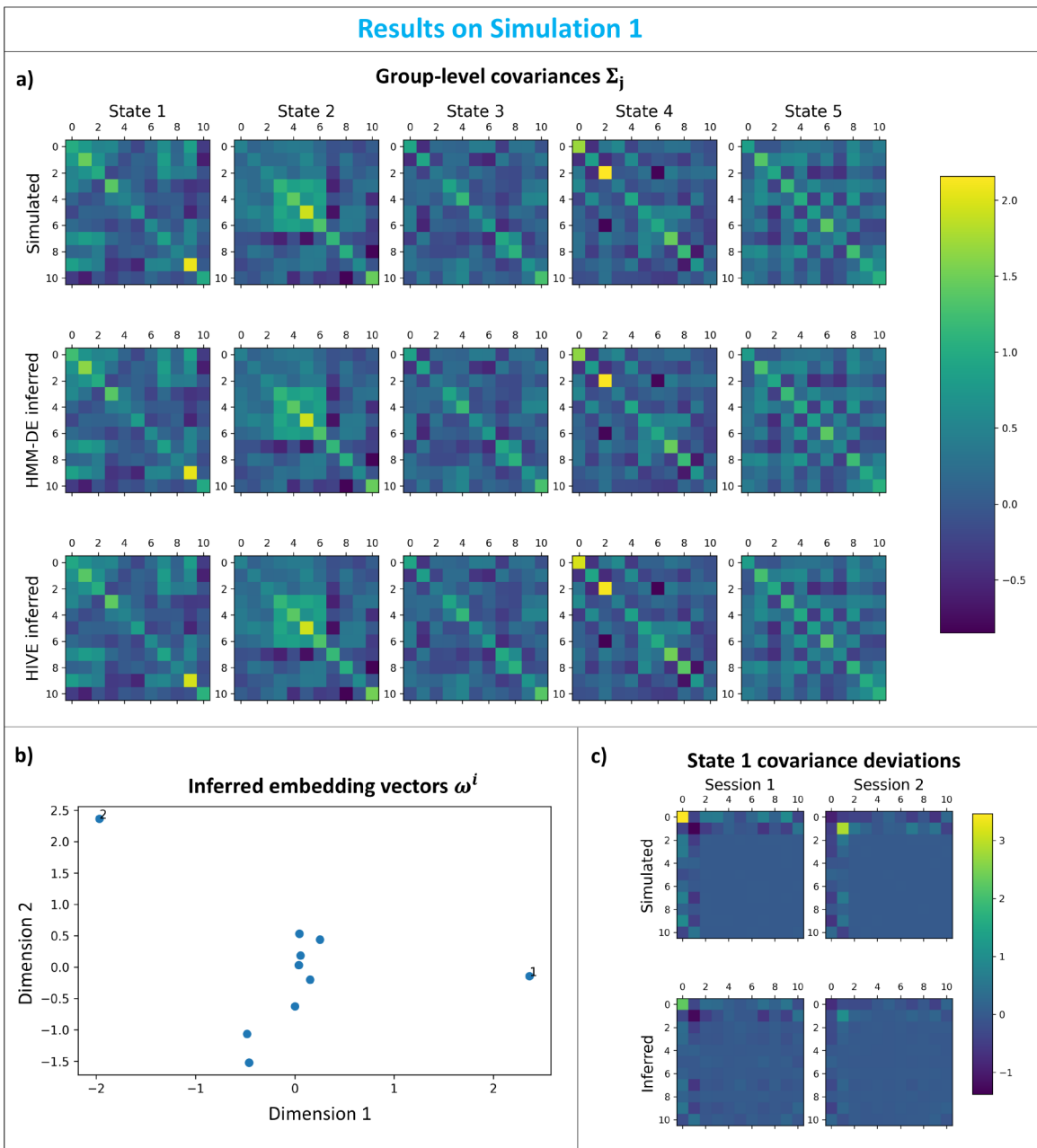


Fig. 4. Simulation 1: Variability encoding block learns session-specific covariance deviations. (a) Simulated (top row), HMM-DE inferred (middle row), and HIVE inferred (bottom row) group-level covariances. The columns show covariances of different states. (b) Inferred session embedding vectors. Sessions 1 and 2 are annotated. (c) Simulated deviations (top row) and deviations from trained generative model of HIVE (bottom row) for state 1 for session 1 (left column) and for session 2 (right column).

HMM-DE is concentrated at low values (< 0.3), whereas there are less extremely low accuracies from HIVE. This again shows HIVE is more robust to noise in the data.

In order to more rigorously investigate the advantages of HIVE over HMM-DE, we regress the accuracies on whether HIVE is used, the number of sessions (N), and the interaction between the two:

$$Accuracy = \beta_0 + \beta_2 \times \mathbb{1}\{\text{HIVE}\} + \beta_2 \times N + \beta_3 \times \mathbb{1}\{\text{HIVE}\} \times N. \tag{16}$$

The results of this fitted regression are summarised in [Figure 6b](#). In particular, the baseline accuracy of HIVE is 0.108 higher than HMM-DE and this effect is significant. Furthermore, the accuracy of HMM-DE increases by 0.00045 per number of session, which translates to 0.0450 per increase of 100 sessions, and this effect is also significant. More importantly, the interaction term β_3 has a significantly positive effect, meaning for each increase of 100 sessions, the increase in accuracy when using HIVE is 0.0458 higher than using HMM-DE (i.e.,

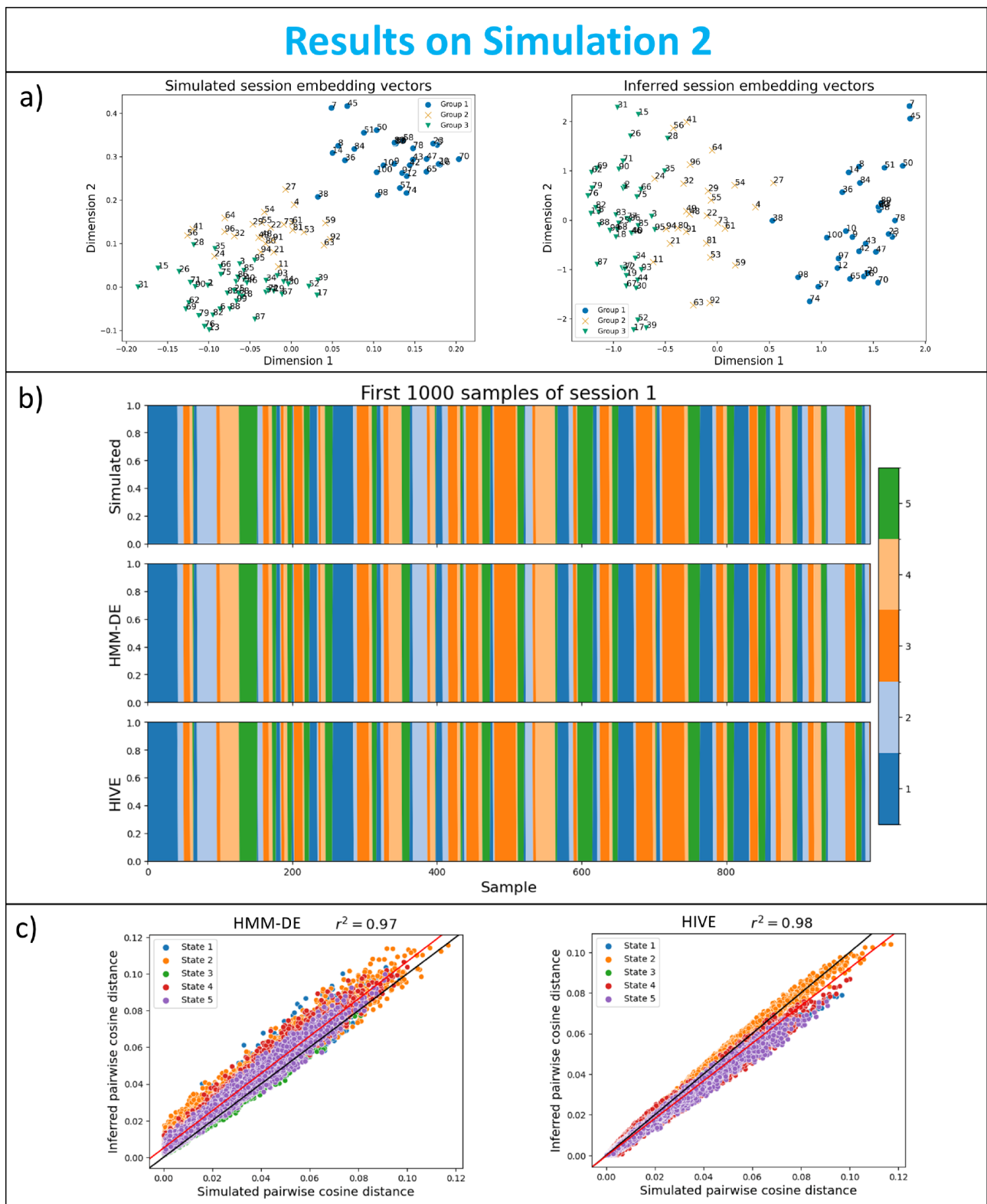


Fig. 5. Simulation 2: HIVE correctly infers the similarity between sessions. (a) Simulated (left) and LDA-projected inferred (right) session embedding vectors. Each point is marked and coloured by the ground-truth group assignment, and is annotated by the session number. (b) Simulated (top), HMM-DE inferred (middle), and HIVE inferred (bottom) state time courses. (c) Session-pairwise cosine distance of simulated (x-axis) against inferred (y-axis) covariances from HMM-DE (left) and HIVE (right). The black line shows the $y = x$ line, which corresponds to optimal performance, and the red line is a fitted line through the points, with the coefficient of determinant r^2 displayed in the title.

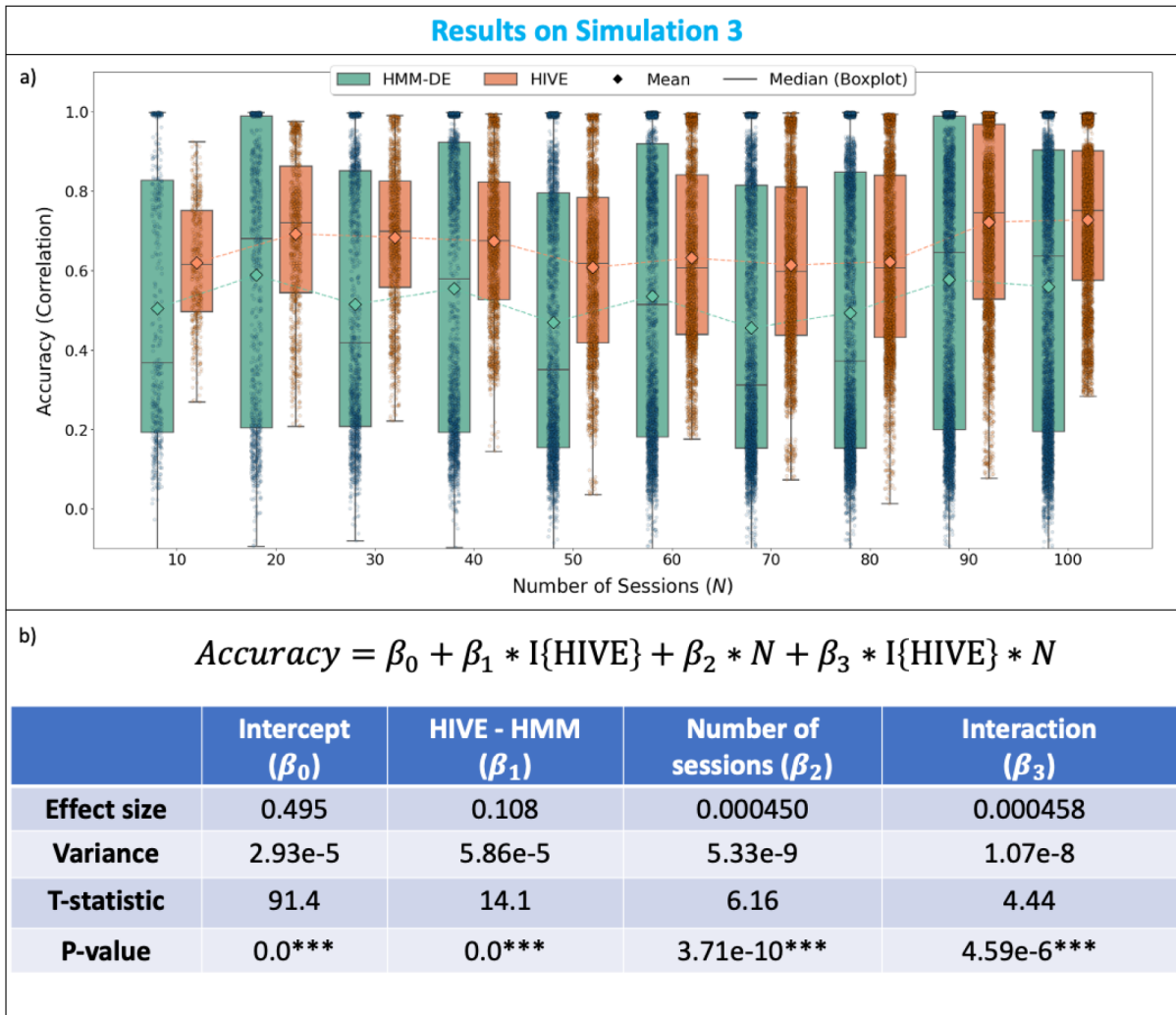


Fig. 6. Simulation 3: HIVE improves the estimation of session variability. (a) Accuracy (correlation with ground-truth session-specific covariances) is plotted against the number of sessions in the dataset. Each dot corresponds to the accuracy of inferred session-specific covariances for each state and each session. (b) A regression is fitted to explore advantages of HIVE over HMM-DE. Values are shown in 3 significant figures.

approximately doubled the increase in accuracy). This shows the advantage of the variability encoding block in making use of data of heterogeneous sessions to help infer on every single session.

3.2. Real MEG data

3.2.1. Wakeman Henson: HIVE reveals similarities and differences between MEG recordings

In this section, we study the Wakeman-Henson dataset (see Section 2.6.2). This dataset contains 6 sessions for each of the 19 subjects, totalling 114 recording sessions. Ideally, we would expect the sessions for a subject to be more similar than sessions for different subjects. In this study, we assign each session an embedding vector and train HIVE on these data with $n_{\omega} = 10$ (see SI Section A.1.7). The session-pairwise cosine distances of embedding

vectors are plotted in Figure 7a, which shows clear block diagonal structure—session embedding vectors from the same subject are closer together than those from different subjects. This shows the model is able to identify certain recordings which have similar deviations from the group, despite HIVE being trained in an unsupervised manner with no knowledge of which sessions belong to which subjects.

To compare with the traditional approaches, we train both HIVE and HMM-DE on this dataset. In Figure 7b, the session-pairwise L2 distances of inferred covariances from both approaches are plotted. Although the same block diagonal structure can be seen in both approaches, it is clearer in the case of HIVE. One can observe there is a particular session (session 3 of subject 8) that has much higher distance with all other sessions, especially with HMM-DE. This is related to the fact that this particular

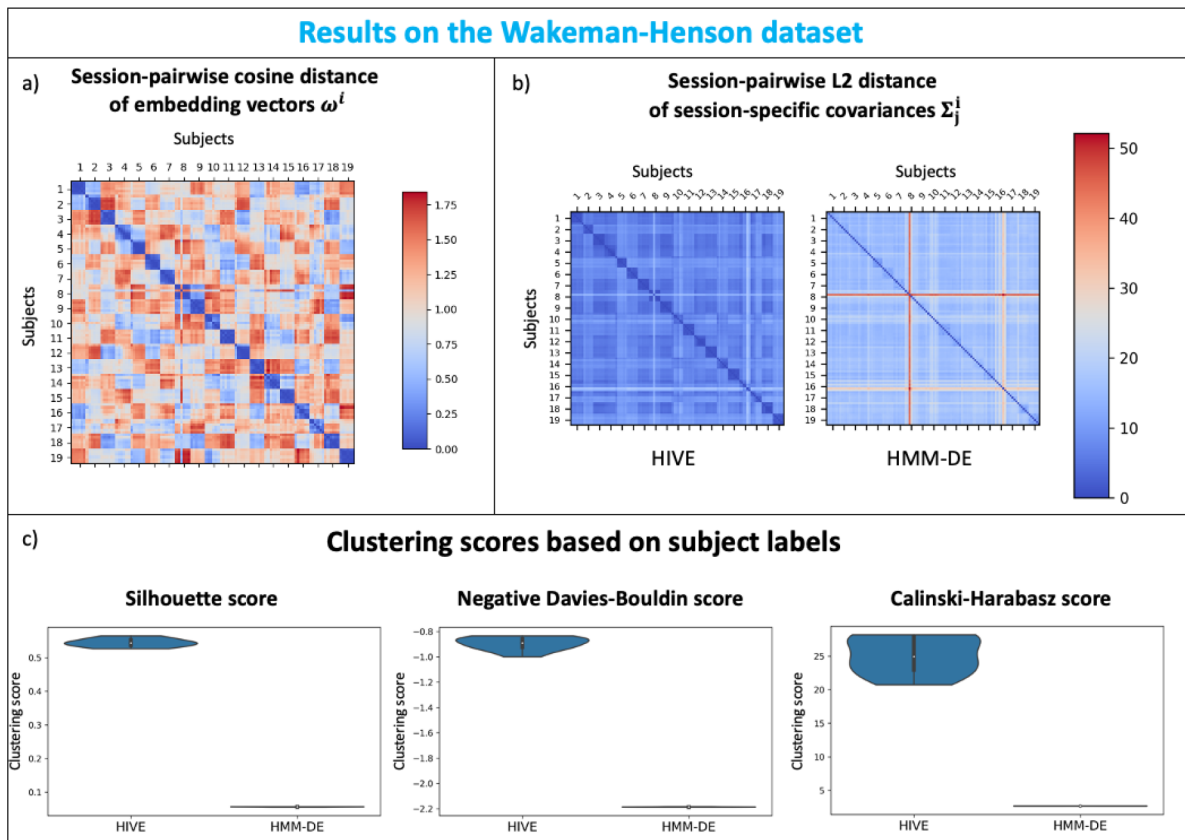


Fig. 7. Wakeman Henson: HIVE reveals similarities and differences between MEG recordings. (a) Session-pairwise cosine distance of inferred embedding vectors. Embedding vectors for the same subject are grouped together and have smaller distance between them. (b) Session-pairwise L2 distance of inferred covariances from HIVE (left) and HMM-DE (right). (c) Clustering metrics - Silhouette score (left), negative Davies-Bouldin score (middle), Calinski-Harabasz score (right), based on subject labels for 10 independent runs of both approaches. Higher values for these metrics indicate better clustering.

session has very different oscillatory activity compared to other sessions (see [SI Section A.1.8](#)). In order to quantify the advantages with HIVE, we employ 3 different metrics—Silhouette score ([Rousseeuw, 1987](#)), Davies-Bouldin score ([Davies & Bouldin, 1979](#)), and Calinski-Harabasz score ([Caliński & Harabasz, 1974](#)) for assessing how well the inferred covariances form distinct and well-separated subject clusters. [Figure 7c](#) shows that with all three metrics, HIVE inferred covariances form tighter and more distinct clusters than dual estimated covariances from HMM-DE.

3.2.2. Combined dataset: HIVE reveals systematic variability across different scanners

In this study, we want to test if our model is able to differentiate data acquired by different scanners. We train on two different resting-state MEG datasets (Nottingham and Cam-CAN) described in [Section 2.6.2](#). There are 128 subjects in total and 64 subjects from each dataset. To avoid the possibility that the model is biased towards either of the datasets, we match the age and sex profiles.

Here, we choose $n_\omega = 20$ (see [SI Section A.1.7](#)). We can clearly see two clusters of embedding vectors inferred in [Figure 8a](#), and at the same time an age gradient in the embedding space. This means scanner type and age information are simultaneously encoded (in different directions in the embedding space) by the embedding vectors, despite the fact that HIVE is trained unsupervised with no knowledge that the sessions were from different scanner types. We can also see a block diagonal structure in [Figure 8b](#) where recording sessions scanned by the same scanner have smaller pairwise cosine distances of their inferred embedding vectors. With the clustering metrics described in [Section 3.2.1](#), we see from [Figure A10](#) that subject-specific covariances from HIVE form better-defined clusters based on scanners/sites.

HIVE also provides a way to summarise the differences in state-specific spectral content between scanner types. For both of the datasets, the centroids of the embedding vectors are computed (shown as black stars in the left panel of [Fig. 8a](#)). For each of the centroids, we select ten nearest neighbours in the embedding space, whose spectral content are averaged to give a represen-

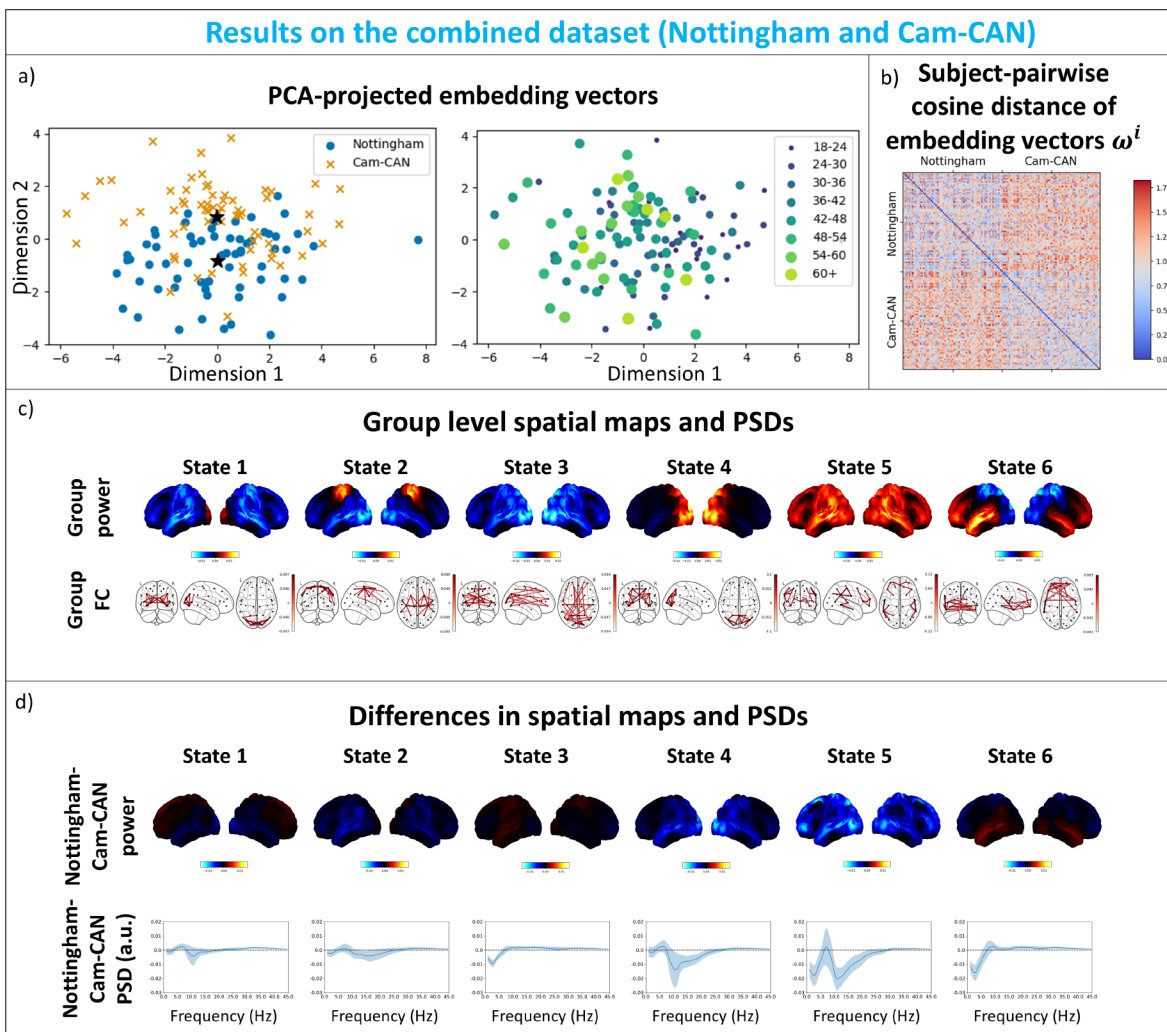


Fig. 8. Combined dataset: HIVE reveals systematic variability across different scanners. (a) PCA-projected embedding vectors show axes that reflect site and age. They are coloured and marked with different datasets (left, black stars indicate centroids of embedding vectors from the two datasets) and different age groups (right, older subjects are coloured with lighter colours and larger dots). (b) Inter-subject differences reflect site differences. Subject-pairwise cosine distance of embedding vectors. (c) Group-level power maps and networks. 6 states are inferred with HIVE. The top and bottom rows show the group-level power (red areas show above average and blue areas show below average power across states) and FC maps (top 3% edges are plotted). (d) Differences between sites in power. The top row shows power difference (red areas show higher and blue areas show lower power in Nottingham subjects than Cam-CAN subjects) for each state between centroids of both datasets. The bottom row shows the difference in PSDs (solid lines show the means, and shaded areas show one standard deviation across the parcels) across frequencies between datasets.

tation of spectral contents of each of the datasets. In Figure 8d, we can see these differences between the “centroids” of the datasets. In particular, subjects from the Nottingham dataset generally have lower power than those from Cam-CAN, especially in posterior regions and alpha band.

3.2.3. Cam-CAN: HIVE reveals individual variability across age

Here, we use the Cam-CAN dataset described in Section 2.6.2, that consists of 612 healthy subjects aged between 18–88 and choose $n_{\omega} = 50$ (see SI Section A.1.7).

From Figure 9a, we can clearly see an age gradient from darker, smaller to lighter, larger dots. This means after training, age information is encoded by the embedding vectors of the subjects, despite the fact that HIVE is trained unsupervised with no knowledge of the subjects’ ages. In order to see if the learnt hidden representation given by the embedding vectors helps improve the model’s ability to distinguish between subject demographics, we try to predict age with the inferred subject-specific covariances. Here, HMM-DE and HIVE with increasing embedding dimensions are trained on this dataset. Shown in Figure 9b are the distributions of accuracy (given by different folds of cross-validation) of predicting

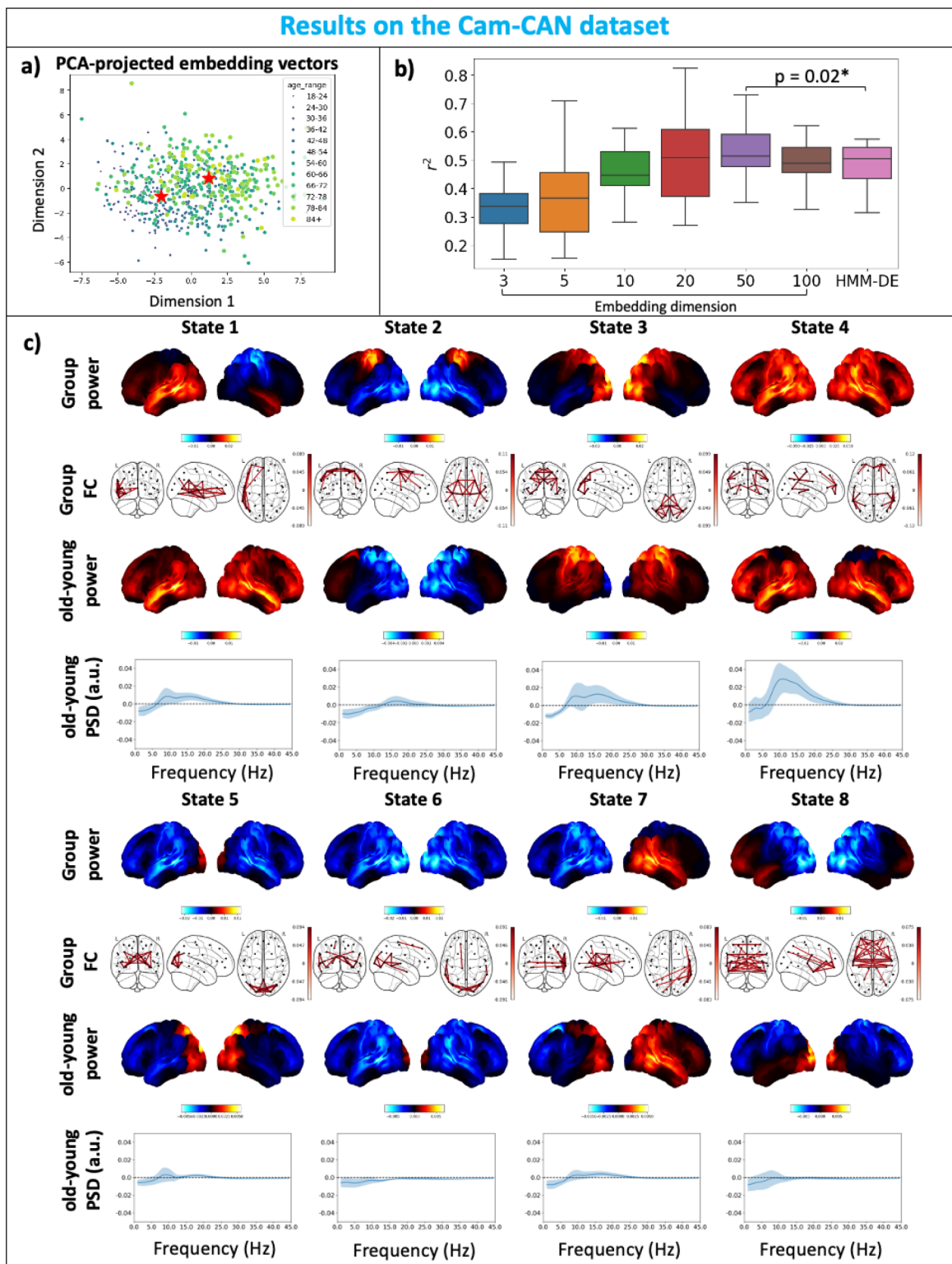


Fig. 9. Cam-CAN: HIVE reveals individual variability across age. (a) Inferred embedding vectors projected to 2 dimensions with PCA. Darker-coloured and smaller dots are younger participants. Lighter-coloured and larger dots are older participants. The two red stars show centroids of subjects with age < 36 and > 72. (b) 20-fold cross-validated age prediction accuracy for HIVE with different embedding dimensions and HMM-DE. (c) 8 states are inferred by HIVE. Group-level power (red areas show above average and blue areas show below average power across states) and FC maps (top 3% edges are plotted) are shown as well as the differences in power (red areas show higher and blue areas show lower power in older subjects than younger subjects) and PSDs (solid lines show the means, and shaded areas show one standard deviation across the parcels) between old and young subjects.

age with inferred subject-specific covariances (see [SI Section A.1.9](#) for details). The coefficient of determinant (r^2) is used as a measure of prediction accuracy and it gradually increases with the embedding dimension until there is a significant improvement over HMM-DE when using an embedding dimension of 50. There is a slight drop in accuracy when using an embedding dimension of 100, which could be due to overfitting. Notice that when selecting the embedding dimension, we do not see a significant decrease in variational free energy if we increase n_{ω} from 50 to 100 ([Fig. A3](#)).

Centroids in the embedding space of subjects of age groups <36 and >72 are used as representatives of young and old subjects. Similar to the analysis in [Section 3.2.2](#), we can find the nearest neighbours (here we choose 20 subjects) for both representatives and get the power maps, PSDs for young and old subjects. We show the results in [Figure 9c](#). We can conclude that HIVE can discover meaningful age patterns in the data.

4. DISCUSSION

In this paper, we propose the use of embedding vectors as a means of characterising functional networks in different sessions/subjects, similar to how word embedding vectors characterise semantic differences between words in a dictionary. These embedding vectors are incorporated into a generative model of FC that uses covariance matrices to describe network activity. This can be potentially used in many different variants of brain network models, including static (time-averaged) approaches and other dynamic network models (e.g., DyNeMo, [Gohil et al., 2022](#)).

In HIVE, similar to the approach taken in PROFUMO ([Harrison et al., 2015, 2020](#)), session deviations from the group-level estimates are generated through a Bayesian prior. The additional feature of embedding vectors allows the model to find subpopulations in the group. By training HIVE on the Wakeman-Henson dataset, we observe that inter-subject variability is much greater than inter-session variability ([Section 3.2.1](#)). This was also found in fMRI literature ([Gratton et al., 2018](#)).

There might be concern over potential over-regularising the variability by the exponential prior on deviation magnitude based on results shown with Simulation 2 ([Fig. 5c](#)) and with the Wakeman-Henson dataset ([3.4b](#)). However, the amount of regularisation provided by the exponential prior depends on the learnable MLP decoder and embedding vectors, that is, the model can learn to automatically adjust the amount of regularisation through training. We believe the shrinkage effect when there is not enough data and excess of noise is the preferred behaviour. Results in [Sections 3.2.1](#) and [3.2.2](#) that show HIVE

inferred covariances form better-separated clusters with either subject labels or scanner types than dual estimated covariances, as well as the fact that HIVE inferred covariances have greater prediction power of age over dual estimated covariances ([Fig. 9b](#)), should help mitigate this concern. Readers might also be concerned about the performance of HIVE trained with different data quality; we have provided the results of training HMM-DE and HIVE on simulated data with different signal-to-noise ratio and it shows HIVE can perform very well with SNR as low as 0.5 ([SI Section A.1.13](#)).

Using simulations, we show that during inference the variability encoding block learns multivariate session-specific deviations ([Section 3.1.1](#)). This is an important feature because multivariate session-specific deviations are expected in real data and this model is capable of learning these. Additionally, we see that the model can accurately infer the pairwise relationships between sessions, that is, discover subpopulation structure through the embedding vectors. We also demonstrated this through three real data studies where we show the structure in the space of embedding vectors is a manifestation of subject ([Section 3.2.1](#)), scanner-type ([Section 3.2.2](#)) and age ([Sections 3.2.2, 3.2.3](#)) differences. These results have profound implications in normative modelling. For instance, the embedding vectors of subjects can be seen as reference points in a population, extracted in a data-driven way. The effects of ageing, deviations from the group, disease progression, etc, can be studied in the space of embedding vectors. The embedding vectors might seem to be abstract representations of the data itself and could be hard to interpret. But thanks to how we formulate the generative model, we can generate the networks of any point, including those not in the training set, in the space of embedding vectors by passing the embedding vector through a trained variability encoding block. Another approach, which is taken in this work, is to aggregate networks from the nearest neighbours ([Sections 3.2.2, 3.2.3](#)).

Potential application of the proposed framework also extends to data harmonisation. For example, suppose we have data from multiple datasets, potentially scanned with different scanners. We can assign different embedding vectors for recordings from different datasets and the same embedding vectors for recordings from the same dataset. By doing this, we can remove the dataset effect on the group-level estimates of networks and retain variability between recordings from the same dataset. Moreover, the proposed framework provides a natural way to isolate different sources of variation (i.e., different sets of embedding vectors for different sources of variation). In the example above, aside from embedding vectors for different datasets, we can also assign embedding

vectors for different age groups, sexes, etc. Hence, the effect of each source of variation can be studied independently in their own space of embedding vectors.

Similarly, apart from differences in data sources, different choices in preprocessing pipeline, source reconstruction/parcellation can affect the individual networks/power maps and it is possible to find different inter-subject similarities/relationships for different preprocessing/source reconstruction/parcellation choices. Studying these effects will be interesting future work and HIVE can provide a good basis for studying the effects of these different combinations of data processing choices.

Transfer learning is also a potential use case of the proposed model. It is often the case that studies for specific demographics or diseases have a small amount of data (we call these “boutique” studies), which could result in a lack of statistical power for discovery. Given the evidence in this work that HIVE inferred networks form better separated clusters (Sections 3.2.1, 3.2.2) and provide more prediction power (Section 3.2.3) than HMM-DE, an important next step for this model would be to train the model on large-scale datasets, and either apply or fine-tune the trained model on the boutique dataset to see if statistical power can be improved.

In this paper, we focused on applying HIVE to MEG data. MEG is an interesting application of HIVE, since, as well as capturing variability over sessions in spatial maps of power, it can also capture variability in the auto and cross-spectra, which have been shown to have strong predictive power (Stier et al., 2025).

However, there is no reason why the proposed model cannot be applied to fMRI data where the number of subjects is much larger, especially given the observation in Figure 6, that HIVE provides significantly more improvement in accuracy compared to HMM-DE with increasing numbers of sessions. However, we acknowledge that the number of samples per subject is much lower in fMRI and the linear trend shown in the left panel of Figure 6 is unlikely to hold if we have an extremely high number of subjects (e.g., in UK Biobank). A comprehensive study on fMRI data is beyond the scope of the current work, but is no doubt, in our opinion, an important direction to explore.

In Section 2.3, we mentioned that we assume all sessions share the same transition probability matrix. For our experiments we found as long as we have enough data per subject, this assumption has a tiny effect on the results. After all, the transition probability matrix is part of the prior distribution and will be overridden by the likelihood when there is enough data. However, a formal investigation is needed when the model is applied to datasets where the amount of data per session is relatively small, for example, as is typically the case in fMRI.

It should be noted that if we choose to colour the PCA-projected embedding vectors in Figure 9a with, for instance, sex, we cannot see clear clustering according to sex (Fig. A6). The reason for this is two-fold. Firstly, PCA is used to visualise embedding vectors, which means only the two directions of biggest variation of the embedding vectors are visualised. This is the reason why we chose to present differences in age and scanner-types. However, we could use techniques like Linear Discriminant Analysis (LDA, Fisher, 1936) to find the linear transformation of the session embedding vectors that best separates a specific source of variation. Secondly, during optimisation of the loss function in Equation (13), it might not be as rewarding to group embedding vectors according to sources of variation that have a smaller effect than those that have a larger effect on the variability of the data. To solve this, a potential solution will be to assign embedding vectors for different sources of variation so that the effects of each source of variation are separated.

Due to the additional complexity of the model, HIVE has more hyper-parameters than the HMM, including the number of layers, number of neurons per layer in the decoder of the variability encoding block. However, in practice, we find that results are very robust to the choice of these hyper-parameters and we have used the same set of hyper-parameters for all simulation and real data studies (see SI Section A.1.6). We suggest users of this model to use the same set of hyper-parameters used in this paper. Furthermore, HIVE is remarkably stable over independent runs of different initialisations of model parameters. We show this in SI Section A.1.12. Although we have shown in Figure 9b that prediction performance changes with the number of embedding dimensions, the conclusion that an age gradient can be observed in the embedding space is consistently found across a wide range of choices for the number of embedding dimensions, and the same can be said about the differences in scanner types shown in Figure 8a. Moreover, when we do not see a significant decrease in the variational free energy when increasing the embedding dimension, we see a drop of prediction power of age with HIVE inferred covariances (from $n_{\omega} = 50$ to $n_{\omega} = 100$, see Fig. 9b and Fig. A3). Nevertheless, we acknowledge that finding a principled way that does not depend on heuristics or post-hoc analysis is an important future direction to explore.

Choosing the number of states in the state-based models, including HMM and HIVE, is a long standing challenge. This work does not attempt to address this issue and we make sure all the conclusions we made in this paper are independent of the choice of number of states. Previous studies have shown the variational free

energy keeps decreasing with increasing number of states (Baker et al., 2014), similarly for other dynamic network models (Gohil et al., 2022). We hypothesise that this could be due to the inter-session variability in the data (see Fig. A5) and modelling variability in the data could be a potential solution to this problem. In this work, we have provided a proof of concept on the possibility of modelling variability with embedding vectors. However, this body of work is by no means perfect. Future work is needed to validate the model further.

5. CONCLUSION

We proposed the use of embedding vectors to model individual functional neuroimaging sessions and applied this approach to extend the HMM, giving us HIVE. The variability encoding block explicitly models variability within a population in a principled way. We provide a way to perform efficient inference on the model parameters and the algorithm is readily scalable to large amount of data. With a Bayesian prior, the model pools information across individuals for how they may deviate from the group mean. The embedding vectors allow the model to group together similar data and help the interpretation of sources of variation in a population. This is an important step towards making use of the numerous large-scale datasets collected with different protocols. We believe the above results demonstrate that the proposed model provides a novel perspective in population modelling and in the inference of functional networks.

ETHICS

The study that collected the CamCAN dataset was conducted in compliance with the Helsinki Declaration, and had been approved by the local ethics committee, Cambridgeshire 2 Research Ethics Committee. Written informed consent was given by participants. See Shafto et al. (2014) for details regarding protocols. The Wakeman-Henson dataset (Wakeman & Henson, 2015) was approved by Cambridge University Psychological Ethics Committee. Written informed consent was obtained from participants. As part of the UK MEG Partnership, the Nottingham dataset was collected at the University of Nottingham. All participants gave written informed consent, and ethical approval was granted by the University of Nottingham Medical School Research Ethics Committee.

DATA AND CODE AVAILABILITY

Data used are publicly available. Availability of the Nottingham dataset is at the official MEGUK site: <https://meguk.ac.uk/database/>. For the Wakeman-Henson data-

set, we refer the readers to the original paper (Wakeman & Henson, 2015). For the Cam-CAN dataset, we refer the readers to the original paper (J. R. Taylor et al., 2017). Source code for HIVE is available in the *osl-dynamics* toolbox (Gohil et al., 2023) and scripts to reproduce results in this paper are available here:

github.com/OHBA-analysis/Huang2025_ModelVariabilityWithEmbeddings.

AUTHOR CONTRIBUTIONS

R.H.: Conceptualisation, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing—original draft, Writing—review and editing, and Visualisation. C.G.: Conceptualisation, Methodology, Software, Data curation, and Writing—review and editing. M.W.: Conceptualisation, Methodology, Data curation, Writing—review and editing, and Supervision.

DECLARATION OF COMPETING INTEREST

No competing interests.

ACKNOWLEDGEMENTS

This research was supported by the National Institute for Health Research (NIHR) Oxford Health Biomedical Research Centre. The Wellcome Centre for Integrative Neuroimaging is supported by core funding from the Wellcome Trust (203139/Z/16/Z). R.H. is supported by the EPSRC Centre for Doctoral Training in Health Data Science (EP/S02428X/1). C.G. is supported by the Wellcome Trust (215573/Z/19/Z). M.W. is supported by the Wellcome Trust (106183/Z/14/Z, 215573/Z/19/Z), the New Therapeutics in Alzheimer's Diseases (NTAD) study supported by UK MRC, the Dementia Platform UK (RG94383/RG89702), and the NIHR Oxford Health Biomedical Research Centre (NIHR203316). The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available with the online version here: <https://doi.org/10.1162/IMAG.a.1188#supplementary-data>

REFERENCES

- Allen, E. A., Damaraju, E., Plis, S. M., Erhardt, E. B., Eichele, T., & Calhoun, V. D. (2014). Tracking whole-brain connectivity dynamics in the resting state. *Cerebral Cortex*, 24(3), 663–676. <https://doi.org/10.1093/cercor/bhs352>

- Ba, J. L., Kiros, J. R., & Hinton, G. E. (2016). Layer normalization. *arXiv preprint arXiv:1607.06450*. <https://doi.org/10.20944/preprints202411.2377.v1>
- Baker, A. P., Brookes, M. J., Rezek, I. A., Smith, S. M., Behrens, T., Probert Smith, P. J., & Woolrich, M. (2014). Fast transient networks in spontaneous human brain activity. *elife*, 3, e01867. <https://doi.org/10.7554/elife.01867>
- Baum, L. E., & Eagon, J. A. (1967). An inequality with applications to statistical estimation for probabilistic functions of Markov processes and to a model for ecology. *Bulletin of the American Mathematical Society*, 73(3), 360–363. <https://doi.org/10.1090/s0002-9904-1967-11751-8>
- Baum, L. E., & Petrie, T. (1966). Statistical inference for probabilistic functions of finite state Markov chains. *The Annals of Mathematical Statistics*, 37(6), 1554–1563. <https://doi.org/10.1214/aoms/1177699147>
- Beckmann, C. F., Mackay, C. E., Filippini, N., & Smith, S. M. (2009). Group comparison of resting-state fMRI data using multi-subject ICA and dual regression. *Neuroimage*, 47(Suppl 1), S148. [https://doi.org/10.1016/s1053-8119\(09\)71511-3](https://doi.org/10.1016/s1053-8119(09)71511-3)
- Betti, V., Della Penna, S., De Pasquale, F., Mantini, D., Marzetti, L., Romani, G. L., & Corbetta, M. (2013). Natural scenes viewing alters the dynamics of functional connectivity in the human brain. *Neuron*, 79(4), 782–797. <https://doi.org/10.1016/j.neuron.2013.06.022>
- Biswal, B., Zerrin Yetkin, F., Haughton, V. M., & Hyde, J. S. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magnetic Resonance in Medicine*, 34(4), 537–541. <https://doi.org/10.1002/mrm.1910340409>
- Blei, D. M., Kucukelbir, A., & McAuliffe, J. D. (2017). Variational inference: A review for statisticians. *Journal of the American Statistical Association*, 112(518), 859–877. <https://doi.org/10.1080/01621459.2017.1285773>
- Bowman, S. R., Vilnis, L., Vinyals, O., Dai, A. M., Jozefowicz, R., & Bengio, S. (2015). Generating sentences from a continuous space. *arXiv preprint arXiv:1511.06349*. <https://doi.org/10.18653/v1/k16-1002>
- Brookes, M. J., O'Neill, G. C., Hall, E. L., Woolrich, M. W., Baker, A., Corner, S. P., Robson, S. E., Morris, P. G., & Barnes, G. R. (2014). Measuring temporal, spectral and spatial changes in electrophysiological brain network connectivity. *Neuroimage*, 91, 282–299. <https://doi.org/10.1016/j.neuroimage.2013.12.066>
- Brookes, M. J., Woolrich, M., Luckhoo, H., Price, D., Hale, J. R., Stephenson, M. C., Barnes, G. R., Smith, S. M., & Morris, P. G. (2011). Investigating the electrophysiological basis of resting state networks using magnetoencephalography. *Proceedings of the National Academy of Sciences*, 108(40), 16783–16788. <https://doi.org/10.1073/pnas.1112685108>
- Brovelli, A., Badier, J.-M., Bonini, F., Bartolomei, F., Coulon, O., & Auzias, G. (2017). Dynamic reconfiguration of visuomotor-related functional connectivity networks. *Journal of Neuroscience*, 37(4), 839–853. <https://doi.org/10.1523/jneurosci.1672-16.2016>
- Caliński, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics-Theory and Methods*, 3(1), 1–27. <https://doi.org/10.1080/03610927408827101>
- Chang, C., & Glover, G. H. (2010). Time–frequency dynamics of resting-state brain connectivity measured with fMRI. *Neuroimage*, 50(1), 81–98. <https://doi.org/10.1016/j.neuroimage.2009.12.011>
- Chehab, O., Defossez, A., Loiseau, J.-C., Gramfort, A., & King, J.-R. (2021). Deep recurrent encoder: A scalable end-to-end network to model brain signals. *arXiv preprint arXiv:2103.02339*. <https://doi.org/10.51628/001c.38668>
- Colclough, G. L., Brookes, M. J., Smith, S. M., & Woolrich, M. W. (2015). A symmetric multivariate leakage correction for MEG connectomes. *Neuroimage*, 117, 439–448. <https://doi.org/10.1016/j.neuroimage.2015.03.071>
- Csaky, R., Van Es, M. W., Parker Jones, O., & Woolrich, M. (2023). Group-level brain decoding with deep learning. *Human Brain Mapping*, 44(17), 6105–6119. <https://doi.org/10.1002/hbm.26500>
- Davies, D. L., & Bouldin, D. W. (1979). A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-1(2), 224–227. <https://doi.org/10.1109/tpami.1979.4766909>
- Défossez, A., Caucheteux, C., Rapin, J., Kabeli, O., & King, J.-R. (2023). Decoding speech perception from non-invasive brain recordings. *Nature Machine Intelligence*, 5(10), 1097–1107. <https://doi.org/10.1038/s42256-023-00714-5>
- Dempster, A. P., Laird, N. M., & 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. <https://doi.org/10.1111/j.2517-6161.1977.tb01600.x>
- Esposito, F., Bertolino, A., Scarabino, T., Latorre, V., Blasi, G., Popolizio, T., Tedeschi, G., Cirillo, S., Goebel, R., & Di Salle, F. (2006). Independent component model of the default-mode brain function: Assessing the impact of active thinking. *Brain Research Bulletin*, 70(4-6), 263–269. <https://doi.org/10.1016/j.brainresbull.2006.06.012>
- Farahibozorg, S.-R., Bijsterbosch, J. D., Gong, W., Jbabdi, S., Smith, S. M., Harrison, S. J., & Woolrich, M. W. (2021). Hierarchical modelling of functional brain networks in population and individuals from big fMRI data. *NeuroImage*, 243, 118513. <https://doi.org/10.1016/j.neuroimage.2021.118513>
- Fedota, J. R., & Stein, E. A. (2015). Resting-state functional connectivity and nicotine addiction: Prospects for biomarker development. *Annals of the New York Academy of Sciences*, 1349(1), 64–82. <https://doi.org/10.1111/nyas.12882>
- Furnov, M., Mohamed, S., & Mnih, A. (2018). Implicit reparameterization gradients. *Advances in Neural Information Processing Systems*, 31. <https://doi.org/10.52202/079017-3285>
- Filippini, N., MacIntosh, B. J., Hough, M. G., Goodwin, G. M., Frisoni, G. B., Smith, S. M., Matthews, P. M., Beckmann, C. F., & Mackay, C. E. (2009). Distinct patterns of brain activity in young carriers of the APOE-ε4 allele. *Proceedings of the National Academy of Sciences*, 106(17), 7209–7214. <https://doi.org/10.1073/pnas.0811879106>
- Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7(2), 179–188. <https://doi.org/10.1111/j.1469-1809.1936.tb02137.x>
- Fornito, A., Harrison, B. J., Zalesky, A., & Simons, J. S. (2012). Competitive and cooperative dynamics of large-scale brain functional networks supporting recollection. *Proceedings of the National Academy of Sciences*, 109(31), 12788–12793. <https://doi.org/10.1073/pnas.1204185109>
- Fox, M. D., & Raichle, M. E. (2007). Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nature Reviews Neuroscience*, 8(9), 700–711. <https://doi.org/10.1038/nrn2201>
- Friston, K. J. (1994). Functional and effective connectivity in neuroimaging: A synthesis. *Human Brain Mapping*, 2(1-2), 56–78. <https://doi.org/10.1002/hbm.460020107>

- Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (1995). *Bayesian data analysis*. Chapman; Hall/CRC. <https://doi.org/10.1201/9780429258411>
- Gohil, C., Huang, R., Roberts, E., van Es, M., Quinn, A., Vidaurre, D., & Woolrich, M. (2023). osl-dynamics: A toolbox for modelling fast dynamic brain activity. *eLife*, 12. <https://doi.org/10.7554/elife.91949.1>
- Gohil, C., Kohl, O., Pitt, J., van Es, M. W., Quinn, A. J., Vidaurre, D., Turner, M. R., Nobre, A. C., & Woolrich, M. W. (2024). Effects of age on resting-state cortical networks. *bioRxiv*, 2024-09. <https://doi.org/10.1101/2024.09.23.614004>
- Gohil, C., Roberts, E., Timms, R., Skates, A., Higgins, C., Quinn, A., Pervaiz, U., van Amersfoort, J., Notin, P., Gal, Y., Adaszewski, S., & Woolrich, M. (2022). Mixtures of large-scale dynamic functional brain network modes. *NeuroImage*, 263, 119595. <https://doi.org/10.1016/j.neuroimage.2022.119595>
- Gratton, C., Laumann, T. O., Nielsen, A. N., Greene, D. J., Gordon, E. M., Gilmore, A. W., Nelson, S. M., Coalson, R. S., Snyder, A. Z., Schlaggar, B. L., Dosenbach, N. U. F., & Petersen, S. E. (2018). Functional brain networks are dominated by stable group and individual factors, not cognitive or daily variation. *Neuron*, 98(2), 439–452. <https://doi.org/10.1016/j.neuron.2018.03.035>
- Harrison, S. J., Bijsterbosch, J. D., Segerdahl, A. R., Fitzgibbon, S. P., Farahibozorg, S.-R., Duff, E. P., Smith, S. M., & Woolrich, M. W. (2020). Modelling subject variability in the spatial and temporal characteristics of functional modes. *NeuroImage*, 222, 117226. <https://doi.org/10.1016/j.neuroimage.2020.117226>
- Harrison, S. J., Woolrich, M. W., Robinson, E. C., Glasser, M. F., Beckmann, C. F., Jenkinson, M., & Smith, S. M. (2015). Large-scale probabilistic functional modes from resting state fMRI. *NeuroImage*, 109, 217–231. <https://doi.org/10.1016/j.neuroimage.2015.01.013>
- Honey, C. J., Sporns, O., Cammoun, L., Gigandet, X., Thiran, J.-P., Meuli, R., & Hagmann, P. (2009). Predicting human resting-state functional connectivity from structural connectivity. *Proceedings of the National Academy of Sciences*, 106(6), 2035–2040. <https://doi.org/10.1073/pnas.0811168106>
- Jalali, A., Azimi, J., & Fern, X. (2012). Exploration vs exploitation in Bayesian optimization. *CoRR*. https://doi.org/10.1007/978-3-642-40988-2_14
- Jayalath, D., Landau, G., Shillingford, B., Woolrich, M., & Parker Jones, O. (2024). The Brain's bitter lesson: Scaling speech decoding with self-supervised learning. *arXiv preprint arXiv:2406.04328*. <https://doi.org/10.20944/preprints202411.2377.v1>
- Jenkinson, M., Pechaud, M., & Smith, S. (2005). BET2: MR-based estimation of brain, skull and scalp surfaces. *Eleventh Annual Meeting of the Organization for Human Brain Mapping*, 17(3), 167. [https://doi.org/10.1016/s1053-8119\(08\)70003-x](https://doi.org/10.1016/s1053-8119(08)70003-x)
- Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*. <https://doi.org/10.1063/pt.5.028530>
- Kingma, D. P., & Welling, M. (2013). Auto-encoding variational Bayes. *arXiv preprint arXiv:1312.6114*. <https://doi.org/10.20944/preprints202411.2377.v1>
- Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1), 79–86. <https://doi.org/10.1214/aoms/117729694>
- Meindl, T., Teipel, S., Elmouden, R., Mueller, S., Koch, W., Dietrich, O., Coates, U., Reiser, M., & Glaser, C. (2010). Test-retest reproducibility of the default-mode network in healthy individuals. *Human Brain Mapping*, 31(2), 237–246. <https://doi.org/10.1002/hbm.20860>
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*. <https://doi.org/10.20944/preprints202411.2377.v1>
- Nickerson, L. D., Smith, S. M., Öngür, D., & Beckmann, C. F. (2017). Using dual regression to investigate network shape and amplitude in functional connectivity analyses. *Frontiers in Neuroscience*, 11, 115. <https://doi.org/10.3389/fnins.2017.00115>
- Norris, J. R. (1998). *Markov chains*. Cambridge University Press. <https://doi.org/10.2307/2585724>
- Philips, G. R., Daly, J. J., & Principe, J. C. (2017). Topographical measures of functional connectivity as biomarkers for post-stroke motor recovery. *Journal of Neuroengineering and Rehabilitation*, 14, 1–16. <https://doi.org/10.1186/s12984-017-0277-3>
- Popescu, M.-C., Balas, V. E., Perescu-Popescu, L., & Mastorakis, N. (2009). Multilayer perceptron and neural networks. *WSEAS Transactions on Circuits and Systems*, 8(7), 579–588. <https://doi.org/10.1109/iscas.1996.541648>
- Quinn, A. J., Atkinson, L. Z., Gohil, C., Kohl, O., Pitt, J., Zich, C., Nobre, A. C., & Woolrich, M. W. (2024). The GLM-spectrum: A multilevel framework for spectrum analysis with covariate and confound modelling. *Imaging Neuroscience*, 2, 1–26. https://doi.org/10.1162/imag_a_00082
- Quinn, A. J., Vidaurre, D., Abeysuriya, R., Becker, R., Nobre, A. C., & Woolrich, M. W. (2018). Task-evoked dynamic network analysis through hidden Markov modeling. *Frontiers in Neuroscience*, 12, 603. <https://doi.org/10.3389/fnins.2018.00603>
- Rabinovich, M. I., Friston, K. J., & Varona, P. (2012). *Principles of brain dynamics*. MIT Press Cambridge, MA. <https://doi.org/10.7551/mitpress/9108.001.0001>
- Rosner, B. (1983). Percentage points for a generalized ESD many-outlier procedure. *Technometrics*, 25(2), 165–172. <https://doi.org/10.1080/00401706.1983.10487848>
- Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53–65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)
- Shafto, M. A., Tyler, L. K., Dixon, M., Taylor, J. R., Rowe, J. B., Cusack, R., Calder, A. J., Marslen-Wilson, W. D., Duncan, J., Dalgleish, T., Henson, R. N., Brayne, C., & Matthews, F. E.; Cam-CAN. (2014). The Cambridge Centre for Ageing and Neuroscience (Cam-CAN) study protocol: A cross-sectional, lifespan, multidisciplinary examination of healthy cognitive ageing. *BMC Neurology*, 14, 1–25. <https://doi.org/10.1186/s12883-014-0204-1>
- Smith, S. M. (2002). Fast robust automated brain extraction. *Human Brain Mapping*, 17(3), 143–155. <https://doi.org/10.1002/hbm.10062>
- Stier, C., Balestrieri, E., Fehring, J., Focke, N. K., Wollbrink, A., Dannlowski, U., & Gross, J. (2025). Temporal autocorrelation is predictive of age—An extensive MEG time-series analysis. *Proceedings of the National Academy of Sciences*, 122(8), e2411098122. <https://doi.org/10.1073/pnas.2411098122>
- Taylor, J. R., Williams, N., Cusack, R., Auer, T., Shafto, M. A., Dixon, M., Tyler, L. K., Cam-CAN, & Henson, R. N. (2017). The Cambridge Centre for Ageing and Neuroscience (Cam-CAN) data repository: Structural and functional MRI, MEG, and cognitive data from a cross-sectional adult lifespan sample. *NeuroImage*, 144, 262–269. <https://doi.org/10.1016/j.neuroimage.2015.09.018>

- Taylor, J. J., Kurt, H. G., & Anand, A. (2021). Resting state functional connectivity biomarkers of treatment response in mood disorders: A review. *Frontiers in Psychiatry*, *12*, 565136. <https://doi.org/10.3389/fpsy.2021.565136>
- Van Dijk, K. R., Hedden, T., Venkataraman, A., Evans, K. C., Lazar, S. W., & Buckner, R. L. (2010). Intrinsic functional connectivity as a tool for human connectomics: Theory, properties, and optimization. *Journal of Neurophysiology*, *103*(1), 297–321. <https://doi.org/10.1152/jn.00783.2009>
- Van Veen, B. D., & Buckley, K. M. (1988). Beamforming: A versatile approach to spatial filtering. *IEEE ASSP Magazine*, *5*(2), 4–24. <https://doi.org/10.1109/53.665>
- van Es, M. W., Gohil, C., Quinn, A. J., & Woolrich, M. W. (2024). osl-ephys: A Python toolbox for the analysis of electrophysiology data. *Frontiers in Neuroscience*, *19*, 1522675. <https://doi.org/10.3389/fnins.2025.1522675>
- Vidaurre, D., Abeyesuriya, R., Becker, R., Quinn, A. J., Alfaro-Almagro, F., Smith, S. M., & Woolrich, M. W. (2018). Discovering dynamic brain networks from big data in rest and task. *NeuroImage*, *180*, 646–656. <https://doi.org/10.1016/j.neuroimage.2017.06.077>
- Vidaurre, D., Llera, A., Smith, S. M., & Woolrich, M. W. (2021). Behavioural relevance of spontaneous, transient brain network interactions in fMRI. *Neuroimage*, *229*, 117713. <https://doi.org/10.1016/j.neuroimage.2020.117713>
- Vidaurre, D., Quinn, A. J., Baker, A. P., Dupret, D., Tejero-Cantero, A., & Woolrich, M. W. (2016). Spectrally resolved fast transient brain states in electrophysiological data. *Neuroimage*, *126*, 81–95. <https://doi.org/10.1016/j.neuroimage.2015.11.047>
- Wakeman, D. G., & Henson, R. N. (2015). A multi-subject, multi-modal human neuroimaging dataset. *Scientific Data*, *2*(1), 1–10. <https://doi.org/10.1038/sdata.2015.1>
- Wang, H., Zariphopoulou, T., & Zhou, X. (2018). Exploration versus exploitation in reinforcement learning: A stochastic control approach. *arXiv preprint arXiv:1812.01552*. <https://doi.org/10.2139/ssrn.3316387>
- Yamashita, A., Yahata, N., Itahashi, T., Lisi, G., Yamada, T., Ichikawa, N., Takamura, M., Yoshihara, Y., Kunitatsu, A., Okada, N., Yamagata, H., Matsuo, K., Hashimoto, R., Okada, G., Sakai, Y., Morimoto, J., Narumoto, J., Shimada, Y., Kasai, K., ... Imamizu, H. (2019). Harmonization of resting-state functional MRI data across multiple imaging sites via the separation of site differences into sampling bias and measurement bias. *PLoS Biology*, *17*(4), e3000042. <https://doi.org/10.1371/journal.pbio.3000042>