



# Multi-dynamic modelling reveals strongly time-varying resting fMRI correlations

Usama Pervaiz<sup>a,\*</sup>, Diego Vidaurre<sup>b,c</sup>, Chetan Gohil<sup>b</sup>, Stephen M. Smith<sup>a</sup>, Mark W. Woolrich<sup>b</sup>

<sup>a</sup> Oxford Centre for Functional MRI of the Brain (FMRIB), Wellcome Centre for Integrative Neuroimaging, Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford OX3 9DU, United Kingdom

<sup>b</sup> Oxford Centre for Human Brain Activity (OHBA), Wellcome Centre for Integrative Neuroimaging, Department of Psychiatry, University of Oxford, Oxford, OX3 7JX, United Kingdom

<sup>c</sup> Department of Clinical Medicine, Aarhus University, Denmark

## ARTICLE INFO

### Article history:

Received 23 June 2021

Revised 29 October 2021

Accepted 10 January 2022

Available online 29 January 2022

### Keywords:

Functional connectivity

Time-Varying functional connectivity

Dynamic functional connectivity

Deep learning

Adversarial learning

Transient brain networks

Hidden Markov model

LSTM

RNNs

## ABSTRACT

The activity of functional brain networks is responsible for the emergence of time-varying cognition and behaviour. Accordingly, time-varying correlations (Functional Connectivity) in resting fMRI have been shown to be predictive of behavioural traits, and psychiatric and neurological conditions. Typically, methods that measure time varying Functional Connectivity (FC), such as sliding windows approaches, do not separately model when changes occur in the mean activity levels from when changes occur in the FC, therefore conflating these two distinct types of modulation. We show that this can bias the estimation of time-varying FC to appear more stable over time than it actually is. Here, we propose an alternative approach that models changes in the mean brain activity and in the FC as being able to occur at different times to each other. We refer to this method as the Multi-dynamic Adversarial Generator Encoder (MAGE) model, which includes a model of the network dynamics that captures long-range time dependencies, and is estimated on fMRI data using principles of Generative Adversarial Networks. We evaluated the approach across several simulation studies and resting fMRI data from the Human Connectome Project (1003 subjects), as well as from UK Biobank (13301 subjects). Importantly, we find that separating fluctuations in the mean activity levels from those in the FC reveals much stronger changes in FC over time, and is a better predictor of individual behavioural variability.

© 2022 The Authors. Published by Elsevier B.V.

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

## 1. Introduction

Large-scale networks of brain activity can be detected as fluctuations of blood oxygenation levels in functional magnetic resonance imaging (fMRI) (Biswal et al., 1995; Fox et al., 2005). These functional networks can be detected in the presence of an experimental paradigm and in spontaneous resting-state activity (De Luca et al., 2006; Tavor et al., 2016; Fox and Raichle, 2007). Many studies estimate FC by calculating the average correlation across an entire fMRI scanning session data (Rogers et al., 2007; O'Reilly et al., 2012; Smith et al., 2013; Smith, 2012; Pervaiz et al., 2020), and these time-averaged measures of FC have been widely

associated with cognitive phenotypes and neuropsychiatric disorders in the literature (Greicius et al., 2004; Anand et al., 2005; Hahn et al., 2011; Hacker et al., 2012; Lynall et al., 2010).

There is now increasing interest in studying the time-varying nature of FC in fMRI. Significant within-session fluctuations in FC during rest or task fMRI have been demonstrated in a number of studies (Lurie et al., 2020; Vidaurre et al., 2017; Hutchison et al., 2013; Allen et al., 2014; Parr et al., 2018) and are also well-established in electrophysiological data (Baker et al., 2014; Vidaurre et al., 2018b; De Pasquale et al., 2010; 2012). Further, a number of research studies have shown the utility of time-varying FC (TVFC) in identifying many psychiatric and neurological conditions e.g., multiple sclerosis (Van Schependom et al., 2019; d'Ambrosio et al., 2020; Huang et al., 2019), schizophrenia (Abrol et al., 2017), Parkinson's (Cai et al., 2018), stroke (Chen et al., 2018), autism spectrum disorder (Guo et al., 2019), epilepsy (Klugah-Brown et al., 2019) and major depressive disorder (Zhi et al., 2018); and the

\* Corresponding author.

E-mail addresses: [usama.pervaiz@ndcn.ox.ac.uk](mailto:usama.pervaiz@ndcn.ox.ac.uk), [12beepervaiz@seecs.edu.pk](mailto:12beepervaiz@seecs.edu.pk) (U. Pervaiz).

dynamics of TVFC have been shown to coactivate with spontaneous replay of recently acquired information (Higgins et al., 2021). Moreover, inter-subject differences in TVFC have also been associated with a number of behavioural and cognitive traits (Liégeois et al., 2019; Vidaurre et al., 2017), with some aspects of intelligence only being explained by TVFC (Vidaurre et al., 2021).

Despite these successes in the estimation and use of TVFC, it is striking how homogeneous FC actually appears to be in fMRI data over time. Indeed, there are a number of studies reporting that there is even insufficient evidence of within-session fluctuations in FC in resting fMRI data (Laumann et al., 2017; Hindriks et al., 2016; Leonardi and Van De Ville, 2015). While it is possible that close to homogeneous FC in fMRI is a genuine phenomenon, it is also possible that it is instead caused by limitations in the methods used to compute TVFC.

Most previous work is based on the sliding window correlation (SWC) approach (Sakoğlu et al., 2010). There are many variants of the SWC method (Allen et al., 2014; Lindquist et al., 2014; Liégeois et al., 2016; Leonardi and Van De Ville, 2015), which are typically limited by using sliding windows of a specified, fixed width. This means that they fail to adapt to the intrinsic time-point by time-point variability of the FC, resulting in changes in FC getting smeared together, thereby making FC look more homogeneous over time. Some of these limitations can be overcome by methods that adapt automatically to time periods of distinct network activity, such as those based on Hidden Markov modelling (HMM) (Vidaurre et al., 2018a; Quinn et al., 2018). Even then, a surprising amount of apparent homogeneity in the FC over time persists.

SWC and the HMM have in common that the mean activity and FC estimations are coupled; that is, changes in the mean activity level are by definition accompanied by changes in the FC, when both are modelled. For example, the most common version of the HMM for fMRI characterises each state as a Gaussian distribution with both mean and covariance. Since there is no biological reason for this coupling assumption, in this work we consider the consequences of relaxing that constraint, in fact showing that this change can offer a potential explanation for the homogeneity of TVFC seen in approaches that tie the dynamics of the mean and the FC together.

Our proposed method for uncoupling the dynamics of the mean activity and FC is called the Multi-dynamic Adversarial Generator-Encoder (MAGE). This proposes a multi-dynamic modelling approach that allows the mean activity and FC to fluctuate in time independently from each other. At the same time, MAGE captures long range temporal dynamics of FC and mean activity using recurrent neural networks (in contrast to the Markovian assumption of the HMM) (Gers et al., 2000; Graves and Schmidhuber, 2005). Finally, our proposed model allows multiple brain states to be simultaneously active (in contrast to the categorical “mutual exclusivity” assumption of the standard HMM). To estimate this model on fMRI data, we developed a novel adversarial generator-encoder network framework inspired by the adversarial regularization used in generative adversarial networks (GANs) (Goodfellow et al., 2014; Lowd and Meek, 2005; Makhzani et al., 2015).

In this paper, we demonstrate the validity of MAGE on several simulation studies and a cognitive task dataset. We also apply the method to resting-state fMRI studies and deduce several outcomes regarding TVFC temporal dynamics, revealing more distinct patterns of FC over time than are obtained with SWC and the HMM, providing an explanation as to why resting fMRI FC has appeared so temporally stable in previous work (Laumann et al., 2017; Hindriks et al., 2016; Leonardi and Van De Ville, 2015). Finally, we substantiate our findings by demonstrating improved prediction of behavioural and cognitive traits using MAGE-derived TVFC estimates.

## 2. Methods

### 2.1. multi-dynamic model

Our proposed approach assumes that fMRI data is generated by a multivariate Normal distribution process with a time-varying mean,  $m_t$ , and covariance,  $C_t$ :

$$Y_t = \mathcal{N}(m_t, C_t) \quad (1)$$

where  $Y_t$  is the value of standardised fMRI data,  $Y$  ( $N \times T$ ) at time point  $t$ , and  $N$  is the number of channels (e.g., brain regions).

In order to separately model the time-varying variances and the correlations, we partition the covariance into a  $N \times N$  diagonal matrix of standard deviations for each brain region,  $G_t$ , and an  $N \times N$  correlation matrix,  $F_t$ :

$$C_t = G_t F_t G_t \quad (2)$$

where the correlation matrix,  $F_t$ , captures instantaneous correlation between brain regions; which we refer to as instantaneous FC.

Clearly, it will not be possible to estimate the instantaneous correlation and means (and variances) at every timepoint without further constraints. As such, we model regularised versions of them by capturing their repeating spatio-temporal structure using an adaptive temporal prior.

Specifically, we assume that the means and correlations (and variances) are generated from their own linear mixture of a limited, underlying set of dynamic modes or states:

$$m_t = \sum_p \alpha_{tp} \mathcal{S}_p \quad (3)$$

$$F_t = \sum_q \beta_{tq} \mathcal{D}_q \quad (4)$$

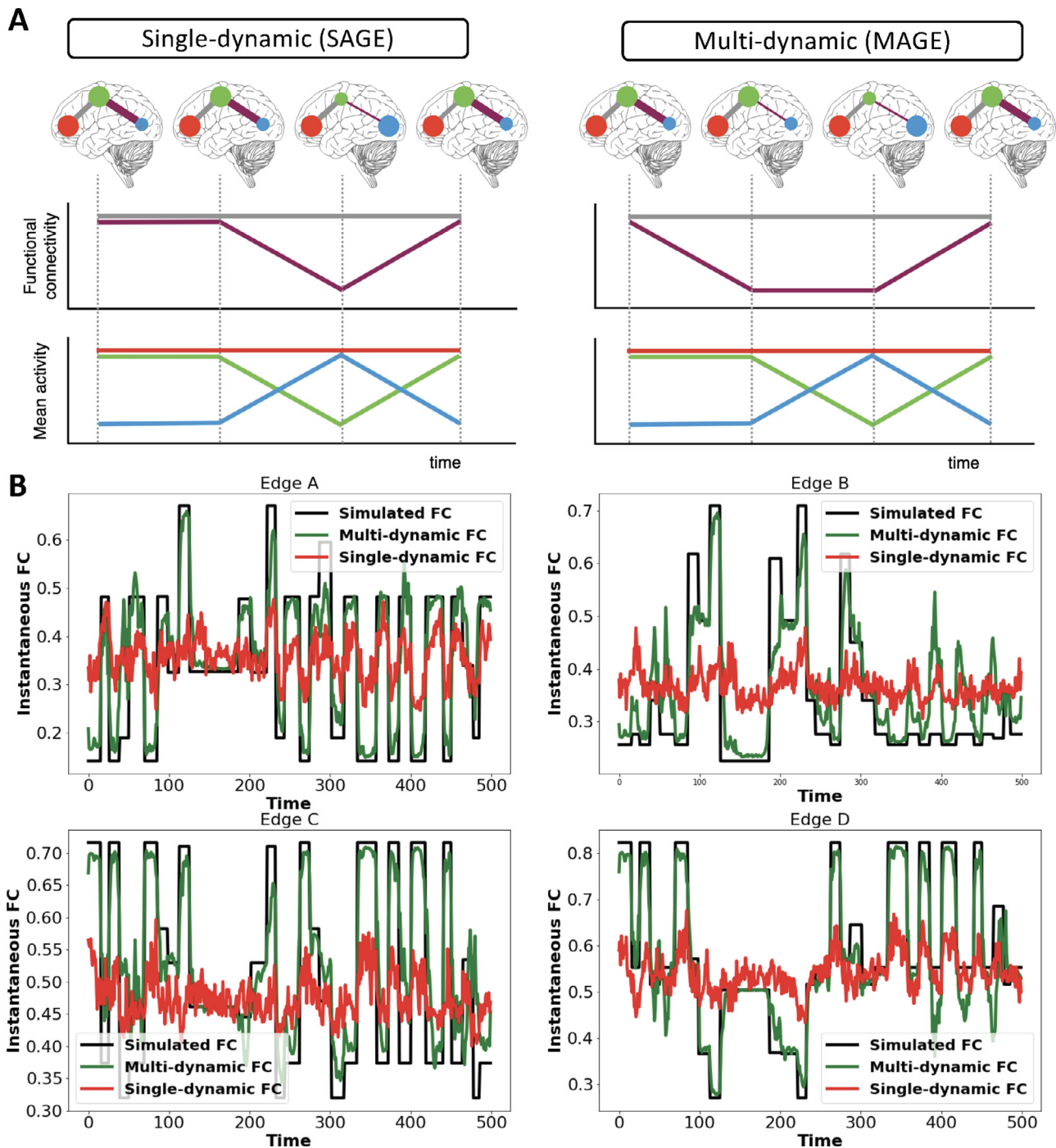
$$G_t = \sum_r \gamma_{tr} \text{diag}(\mathcal{E}_r) \quad (5)$$

where  $\alpha_{tp}$  is a scalar corresponding to the time course of the  $p$ th “mean activity” state ( $p=1\dots P$ ), and  $\mathcal{S}_p$  is its  $N \times 1$  spatial map of mean activity; and  $\beta_{tq}$  is a scalar corresponding to the time course of the  $q$ th “correlation” state ( $q=1\dots Q$ ), and  $\mathcal{D}_q$  is its  $N \times N$  correlation matrix; and  $\gamma_{tr}$  is a scalar corresponding to the time course of the  $r$ th “variance” state ( $r=1\dots R$ ), and  $\mathcal{E}_r$  is its  $N \times 1$  spatial map of the standard deviations. The  $\mathcal{D}_q$  correlation matrices capture the state-specific correlation between brain regions; which we refer to as the state-specific FC. As we shall describe in detail later, we assume that both the state time courses  $[\alpha, \beta]$  and the state-specific maps  $[\mathcal{S}_p, \mathcal{D}_q, \mathcal{E}_r]$  are unknown and need inferring from the data. Also,  $\alpha_{tp}$ ,  $\beta_{tq}$  and  $\gamma_{tr}$  are not coupled together and are completely free to fluctuate independently.

Note that in Eqs. (3), (4), (5) the correlations and means (and variances) are regularised using their own distinct priors, or dynamics. This is in contrast to most TVFC methods, which are *single-dynamic* approaches in that the dynamics of the means and the correlations (i.e., the FC) are tied together. However, there is no imperative to assume that the dynamics of the means and correlations are the same as each other. As such, the proposed method is designed to explore what happens when we do not tie all of the dynamics together. We refer to this approach as *multi-dynamic*, as illustrated in Fig. 1[A].

#### 2.1.1. Special case - Sliding window correlation

Sliding window correlation (SWC) measures of TVFC capture how the between-brain-region correlation fluctuates over time. For simplicity here (but without loss of generality), we assume non-overlapping sliding windows. This corresponds to Eqs. (3)–(5),



**Fig. 1.** Single-dynamic versus multi-dynamic brain networks. **[A]** Top plot shows example brain network dynamics for the two scenarios; line width indicates functional connectivity (FC) strength, and size of node indicates mean activity amplitude. Middle and bottom plots highlight how the single-dynamic case assumes that the means and the FC change together, whereas in the multi-dynamic case they can change at distinct times. **[B]** Simple simulation illustrating the potential benefits of a multi-dynamic approach. The ground truth is that there are different dynamics underlying the time-varying means and FC. The multi-dynamic approach is able to accurately adjust to fluctuating FC (illustrated for four edges of the FC matrix), whereas the single-dynamic approach is not able to accurately capture these fluctuations, producing time-varying, instantaneous FC that erroneously looks much more homogeneous, as is often observed in time-varying FC estimates in real fMRI data. Lastly, the multi-dynamic approach is also able to adjust to fluctuating mean as shown in Figure S6.

where  $\alpha_{tp}$ ,  $\beta_{tq}$  and  $\gamma_{tr}$  are set to be the same, and are fixed a priori to correspond to the  $P=Q=R$  different positions the sliding window can take as it slides through time (e.g., in the case of square windows,  $\alpha$  and  $\beta$  and  $\gamma$  have a value of one inside the sliding window, and a value of zero outside of it). Note that  $P=Q=R$  needs to typically be very large to span all the data over time; and so SWC approaches are normally followed by a clustering ap-

proach, such as k-means, to find repeating correlation patterns (Allen et al., 2014; Tagliazucchi et al., 2012). Typically, information about how the mean fluctuates over time is ignored. Since  $\alpha_{tp}$ ,  $\beta_{tq}$  and  $\gamma_{tr}$  are set to be the same, i.e., the means and correlations (and variances) are modelled using the same dynamics, then SWC corresponds to what we refer to as a *single-dynamics* approach (see Fig. 1[A]).

### 2.1.2. Special case - hidden Markov models

Hidden Markov Models (HMM), with multi-variate Normal observation models, capture how both the mean activity levels and the between-brain-region correlations fluctuate over time. This form of the HMM is the same as SWC in that it corresponds to Eqs. (3)–(5) where the time courses  $\alpha_{tp}$ ,  $\beta_{tq}$  and  $\gamma_{tr}$  are the same however,  $\alpha_{tp}$ ,  $\beta_{tq}$  and  $\gamma_{tr}$  are now inferred from the data at the same time as  $S_p$ ,  $D_q$  and  $\mathcal{E}_r$ . This means that the approach can adjust to how the transient network states are coming and going for different durations at different points in time. Note that in addition, the HMM assumes that  $\alpha_{tp} = \beta_{tq} = \gamma_{tr}$  are binary time courses, defining when states  $p$  switches on and off, and that no two states can be active at the same time (i.e., the states are mutually exclusive). As with SWC, since the means and correlations (and variances) are modelled using the same dynamics, then this version of the HMM corresponds to what we refer to as a *single-dynamics* approach (see Fig. 1[A]).

### 2.2. What do we mean by functional connectivity?

Typically, correlation-based functional connectivity in fMRI refers to the correlation between two fMRI time series. However, in the model we are proposing, there are multiple entities that could be described as FC, all of which capture some different aspect of correlation between brain regions. First, the model describes the correlation at each time point, i.e.,  $F_t$  (Eqs. (1),(2)). We refer to these as the **instantaneous FCs**. Second, a set of state-specific correlations, i.e.,  $D_q$ , are mixed together to regularise the instantaneous FC (see Eq. (4)). We will refer to these as **state-specific FCs**. Finally, we might also consider correlations in the mean activity time courses, which is a single correlation matrix that corresponds to the correlation of  $m_t$  over all time, i.e.,  $\text{Corr}(m_t)$ . We will refer to this as the **mean activity FC**; this is potentially quite independent of the state-specific FCs.

Our model provides a regularised estimate of the instantaneous FC, and so we can more parsimoniously represent the instantaneous FC using the state-specific FCs. Hence in this work, we focus on using the state-specific FCs as our main TVFC feature of interest, for example when predicting behavioural traits.

### 2.3. Multi-dynamic adversarial generator-encoder (MAGE)

Equations (1)–(5) describe the modelling framework for the multi-dynamic approach we are proposing. In this section, we will describe the more complete generative model, including the generative process assumed for the dynamics of  $\alpha_{tp}$ ,  $\beta_{tq}$  and  $\gamma_{tr}$ , and how we infer on the model. We refer to the overall approach as the *Multi-dynamic Adversarial Generator-Encoder (MAGE)*.

In practice, MAGE assumes that the dynamics of the means and variances are tied together, but crucially the correlation is assumed to fluctuate independently. This corresponds to Eqs. (3), (4), (5) with  $\alpha_{tp} = \gamma_{tr}$  ( $P=R$ ), but with  $\beta_{tq}$  free to be inferred separately.<sup>1</sup> Alongside the state time courses we also infer the state-specific features:  $S_p$ ,  $D_q$  and  $\mathcal{E}_r$ . Note that while the dynamics of both the means and the variances are always modelled (and are tied together with the same dynamics), for simplicity we will tend to refer to just the dynamics of the means throughout the paper.

In addition, unlike the HMM, MAGE does not assume that state time courses are binary and mutually exclusive. Instead, we assume that they describe time-varying linear mixtures with a sum-

to-one constraint ( $\alpha_{tp}$ ,  $\beta_{tq}$  and  $\gamma_{tr}$  separately sums to one at each time point), which we will refer to as “partial volume” state time-courses. This sum-to-one constraint is an identification condition that is imposed to ensure identifiability (Huang, 2014) on the inferred states and makes the post-model temporal analysis interpretable.

In summary, MAGE describes the data as being generated from a finite set of states. These states are linearly mixed together at each time-point to generate a time-varying description of the data’s mean and covariance.

#### 2.3.1. Generative model

The generative model for the data,  $Y_t$ , corresponds to Eq. (1). Each instance of  $Y_t$  is described by a multivariate Normal distribution parameterised by the mean activity at time  $t$ ,  $m_t$ , (Eq. (3)) and the covariance matrix at time  $t$ ,  $C_t$  (Eq. (2)). We model the state time courses,  $\alpha_t$ ,  $\gamma_t$  and  $\beta_t$ , based on their underlying logits:

$$\alpha_t = \gamma_t = \zeta(\theta_t^m) \quad (6)$$

$$\beta_t = \zeta(\theta_t^c) \quad (7)$$

where  $\zeta$  is the softmax function, ensuring that  $\alpha$  and  $\beta$  sum to one (over states) at each time point.

The dynamics in  $\theta_t^m$  and  $\theta_t^c$  (and implicitly the dynamics of the state time courses) are captured using a unidirectional sequence-to-sequence long short-term memory (LSTM) model, which uses the history of  $\theta$ , to predict the value at the next time point. Specifically, it predicts the prior means of the latent, time-varying parameters  $[\theta_t^m, \theta_t^c]$ .

$$\hat{\mu}_t^{m,\theta} = f(LSTM_{uni}(\theta_{<t}^m, \omega_g^m), \lambda_g^m) \quad (8)$$

$$\hat{\mu}_t^{c,\theta} = f(LSTM_{uni}(\theta_{<t}^c, \omega_g^c), \lambda_g^c) \quad (9)$$

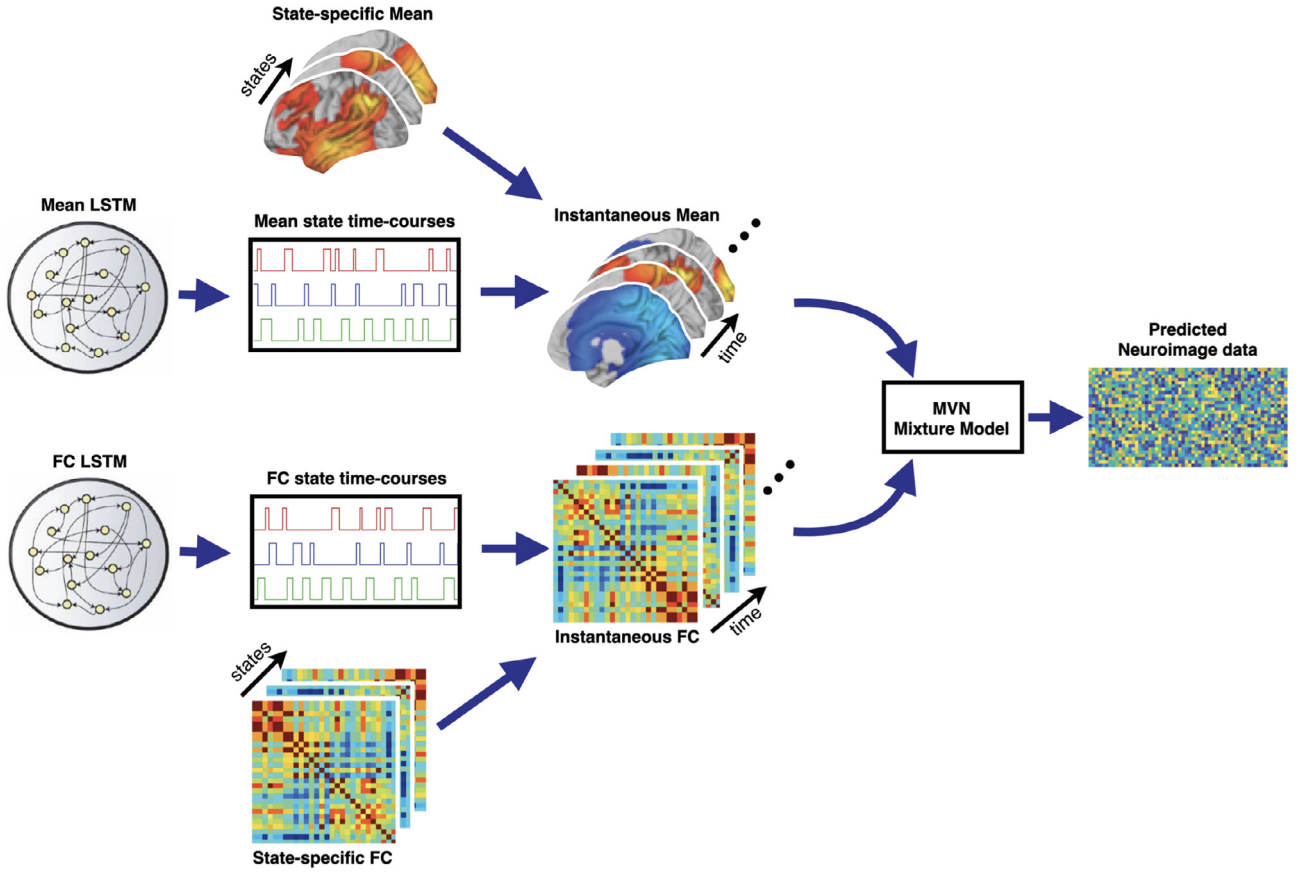
where  $\hat{\mu}_t^{m,\theta}$  and  $\hat{\mu}_t^{c,\theta}$  are the prior means of  $\theta_t^m$  and  $\theta_t^c$  respectively.  $f()$  is the affine transformation and  $\lambda_g^m$  and  $\lambda_g^c$  represents the weights of the affine transformation.  $\omega_g^m$  and  $\omega_g^c$  are the generative model LSTM weights. The proposed generative model for the data is illustrated in Fig. 2 (and the in-depth architecture is illustrated in Figure S2).

#### 2.3.2. Inference

Given fMRI data, we need to infer on the parameters  $[S_p, D_q, \mathcal{E}_r, \theta^m, \theta^c, \omega_g^m, \omega_g^c, \lambda_g^m, \lambda_g^c]$  in the generative model. We infer point estimates with uniform priors for the global parameters in the model, i.e.,  $[S_p, D_q, \mathcal{E}_r, \omega_g^m, \omega_g^c, \lambda_g^m, \lambda_g^c]$ ; but for the more challenging inference of the latent, time-varying parameters  $[\theta^m, \theta^c]$ , we want to regularise using priors (which here corresponds to the prior means of  $[\theta^m, \theta^c]$  as stated in Eqs. (8),(9)). This could be attempted using a number of different approaches. For example, Variational Bayes (VB) methods could be used based on the ideas from Variational autoencoders (VAEs) (Kingma and Welling, 2013), where the generative model acts as a decoder that maps from the latent, time-varying parameters  $[\theta^m, \theta^c]$  to the fMRI data. The loss function in VB inference is based on optimizing the variational free energy (Friston et al., 2007) and consists of two terms - the reconstruction likelihood, and the Kullback-Leibler (KL) divergence that regularises the solution by matching the learned posterior distribution to a prior distribution over the latent variables (i.e., Eqs. (8), (9)).

Here, we take an approach based on an adversarial autoencoder (AAE) (Makhzani et al., 2015). Again, the generative model acts as a decoder that maps from the parameters underlying the state time courses,  $[\theta^m, \theta^c]$ , to the fMRI data, and hence the reconstruction likelihood term of the AAE is the same as it would be in a VAE.

<sup>1</sup> The reason MAGE models the variances to be independent of the correlation (i.e.,  $\gamma_{tr}$  free to be not equal to  $\beta_{tq}$ ) is to allow maximum temporal independence to the correlation. We will expand onto this in Section 4 but allowing  $\gamma_{tr} = \beta_{tq}$  resulted in weaker state-specific FCs dynamics, and hence was not a preferable modelling choice.



**Fig. 2.** Generative model for MAGE. The proposed model generates data by first generating instantaneous means and the instantaneous correlation matrices (i.e., functional connectivity). MAGE also models instantaneous variances, whose dynamics are tied to be the same as the mean (not shown here). The instantaneous mean is modelled using an underlying set of states, for which the state time courses are generated using a long short-term memory (LSTM) model. The instantaneous correlation is also modelled using an underlying set of states, whose state time courses are generated using a completely different LSTM, making the approach multi-dynamic. The generative model for SAGE is illustrated in Figure S1.

However, the AAE replaces the regularising KL divergence term of the VAE with an adversarial training criterion (Makhzani et al., 2015; Mescheder et al., 2017). This takes the idea of adversarial loss from GANs to achieve the prior regularization - by forcing the posterior means underlying the state time courses,  $[\theta^m, \theta^c]$ , to be close to the specified prior means (i.e., Eqs. (8), (9)). To do this, we train a discriminator, separately from the training that corresponds to the generative model parameter inference, to be able to distinguish between the prior and posterior versions of the  $[\theta^m, \theta^c]$ . The trained discriminator is then used to enforce closeness between the posterior and prior (alongside the reconstruction likelihood) when performing the model parameter inference.

**Encoder Model** The encoder model maps from the fMRI data to the posterior estimates of the latent, time-varying parameters  $[\theta^m, \theta^c]$ . This can also be referred to as the inference network or inference model and corresponds to the idea of amortized inference in VAEs, whereby the encoder allows an efficient means to “look-up” the posterior estimates given the data (Kingma and Welling, 2013). The authors who proposed AAEs (Makhzani et al., 2015) explored the benefits of being fully Bayesian on the latent variables, by assuming that the posterior is a Gaussian distribution (with the mean and variance predicted by the encoder, in a similar manner to the VAE); but after an extensive hyper-parameter search, they did not find any additional advantages, and only reported results using a deterministic version of the posterior. As such, we learned a single nonlinear mapping that is used to map from the data,  $Y_t$ , to the posterior means at any time point:

$$\mu_t^{m,\theta} = f(LSTM_{bi}(Y, \omega_e^m), \lambda_e^m) \quad (10)$$

$$\mu_t^{c,\theta} = f(LSTM_{bi}(Y, \omega_e^c), \lambda_e^c) \quad (11)$$

where  $\mu_t^{m,\theta}$  and  $\mu_t^{c,\theta}$  are the posterior means of  $\theta_t^m$  and  $\theta_t^c$  respectively.  $LSTM_{bi}$  is a sequence-to-sequence, bidirectional LSTM model, and  $\omega_e^m$  and  $\omega_e^c$  are the encoder LSTM weights.  $f()$  is the affine transformation and  $\lambda_e^m$  and  $\lambda_e^c$  represents the weights of the affine transformation. Note that the approach is still stochastic since we use batching of the data.

**Discriminator Model** As mentioned above, a discriminator is needed to be able to distinguish between the prior and posterior mean estimates of the state time courses,  $[\theta^m, \theta^c]$ . The discriminator is also constructed using a sequence-to-sequence, bidirectional LSTM model and it classifies whether samples are from the posterior or prior estimates of  $[\theta^m, \theta^c]$ :

$$D_{\theta_t^m} = \sigma(f(LSTM_{bi}([\zeta(\hat{\mu}_t^{m,\theta}), \zeta(\mu_t^{m,\theta})], \omega_d^m), \lambda_d^m)) \quad (12)$$

$$D_{\theta_t^c} = \sigma(f(LSTM_{bi}([\zeta(\hat{\mu}_t^{c,\theta}), \zeta(\mu_t^{c,\theta})], \omega_d^c), \lambda_d^c)) \quad (13)$$

where  $\sigma$  is the sigmoid activation to ensure that discriminator output is either 0 or 1, and  $\omega_d^m$  and  $\omega_d^c$  are the discriminator LSTM weights.  $f()$  is the affine transformation and  $\lambda_d^m$  and  $\lambda_d^c$  represents the weights of the affine transformation. Note that the discriminator takes in as input the full history of both the prior and posterior mean state time courses into its LSTM. Taken together with the fact that we do not factorise the posteriors of  $[\theta^m, \theta^c]$  over time, this means that the trained discriminator will help drive the prior and posterior state time courses to have similar characteristics not just over short, but also over long, time scales.

### 2.3.3. How it works

We trained this model as a three-player game played between the generator (i.e., the generative model with parameters  $[S_p, D_q, \mathcal{E}_r, \theta^m, \theta^c, \omega_g^m, \omega_g^c, \lambda_g^m, \lambda_g^c]$ ), encoder (with parameters  $[\omega_e^m, \omega_e^c, \lambda_e^m, \lambda_e^c]$ ), and discriminator (with parameters  $[\omega_d^m, \omega_d^c, \lambda_d^m, \lambda_d^c]$ ). The discriminator model is trained to distinguish prior mean estimates of  $[\theta^m, \theta^c]$  from their posterior mean estimates. The discriminator is then used to enforce closeness between the posterior and prior of those parameters (while also honouring the reconstruction likelihood) when training the encoder and generator. In this game, when the discriminator successfully identifies the prior from posterior, it is rewarded; whereas both encoder and generator models are penalized. In contrast, when the combined efforts of the encoder and generator models fools the discriminator, the discriminator is penalized. Note, that at the same time the encoder and generator models must also look to maximise the reconstruction likelihood.

This three-player game is played with two main loss functions that are listed below:

1. **Reconstruction (Likelihood) loss:** The first loss corresponds to the likelihood - this tells us how well we are explaining our data according to the current estimate of the generative model, and is given by:

$$L = \sum_{t=1}^T \log p(Y_t | \theta_t^m = \mu_t^{m,\theta}, \theta_t^c = \mu_t^{c,\theta}) \quad (14)$$

2. **Regularization (Prior) Loss:** The second loss regularises the estimate of the latent, time-varying parameters  $[\theta^m, \theta^c]$  using an adaptive prior - this penalises when the posterior estimates of  $[\theta^m, \theta^c]$  deviate from the prior:

$$R = \sum_{t=1}^T [\text{CrossEntropy}(\mu_t^{m,\theta} || \hat{\mu}_t^{m,\theta}) + \text{CrossEntropy}(\mu_t^{c,\theta} || \hat{\mu}_t^{c,\theta})] \quad (15)$$

This corresponds to the cross-entropy between the prior and the posterior of the latent, time-varying parameters  $[\theta^m, \theta^c]$ .

The total loss is the summation of these two terms:

$$F = \Lambda L + \frac{1 - \Lambda}{2} R \quad (16)$$

where we attribute a certain weight to each term using  $\Lambda$ . The expression in Eq. (16) can be thought of as an approximation of the variational free energy, but where the KL term is replaced by regularisation loss in Eq. (15). The complete derivation for the regularisation loss in Eq. (15) is given in Section S1.5. We trained the parameters (i.e.,  $S_p, D_q, \mathcal{E}_r, \theta^m, \theta^c, \omega_g^m, \omega_g^c, \omega_e^m, \omega_e^c, \omega_d^m, \omega_d^c, \lambda_g^m, \lambda_g^c, \lambda_e^m, \lambda_e^c, \lambda_d^m, \lambda_d^c$ ) using these losses in two phases - the reconstruction phase and the regularization phase - which are explained below:

- **Reconstruction Phase:** In the reconstruction phase (Eq. (14)), the generator model (with parameters  $[S_p, D_q, \mathcal{E}_r, \theta^m, \theta^c, \omega_g^m, \omega_g^c, \lambda_g^m, \lambda_g^c]$ ) and encoder model (with parameters  $[\omega_e^m, \omega_e^c, \lambda_e^m, \lambda_e^c]$ ) are updated to minimize the reconstruction loss. We kept the discriminator model (with parameters  $[\omega_d^m, \omega_d^c, \lambda_d^m, \lambda_d^c]$ ) constant during this phase, otherwise the generator and encoder would be trying to hit a moving target and might never converge.
- **Regularization Phase:** In the regularization phase (Eq. (15)), the discriminator model (with parameters  $[\omega_d^m, \omega_d^c, \lambda_d^m, \lambda_d^c]$ ) is first updated to distinguish prior and posterior mean estimates of  $[\theta^m, \theta^c]$ . Then, the parameters of the generator model and encoder model are updated to confuse the discriminator model.

### 2.3.4. Single-dynamic adversarial generator-Encoder (SAGE)

There is variant of the proposed model, SAGE, that assumes the time courses  $\alpha_{tp}$ ,  $\gamma_{tr}$  and  $\beta_{tq}$  are the same. In this regard it is the same as the HMM. However, SAGE maintains the long-range temporal modelling and the partial volume modelling of the network state dynamics that are not part of the standard HMM. Throughout this paper, we will compare MAGE with SAGE to highlight the benefits of the multi-dynamic modelling. SAGE is explained in Section S1.1 and is illustrated in Figure S1.

### 2.4. Dual-Estimation of the group-level MAGE

We performed a process of dual-estimation (analogous to dual-regression following group-ICA) (Beckmann et al., 2009) to infer the subject-level state versions of the group-level MAGE. This allows us to obtain a subject-specific description of time-varying co-relationship and time-varying mean. To achieve this, we used the trained parameters inferred from the group-level MAGE ( $\alpha_{tp}$ ,  $\beta_{tq}$ ,  $S_p$ ,  $D_q$  and  $\mathcal{E}_r$ ). Then we re-estimated the  $C_t$  and  $m_t$  for each subject by holding constant the  $\alpha$  and  $\beta$  parameters, and only retraining  $S_p$ ,  $D_q$  and  $\mathcal{E}_r$ .

### 2.5. Hyper-parameters tuning and cross-validation

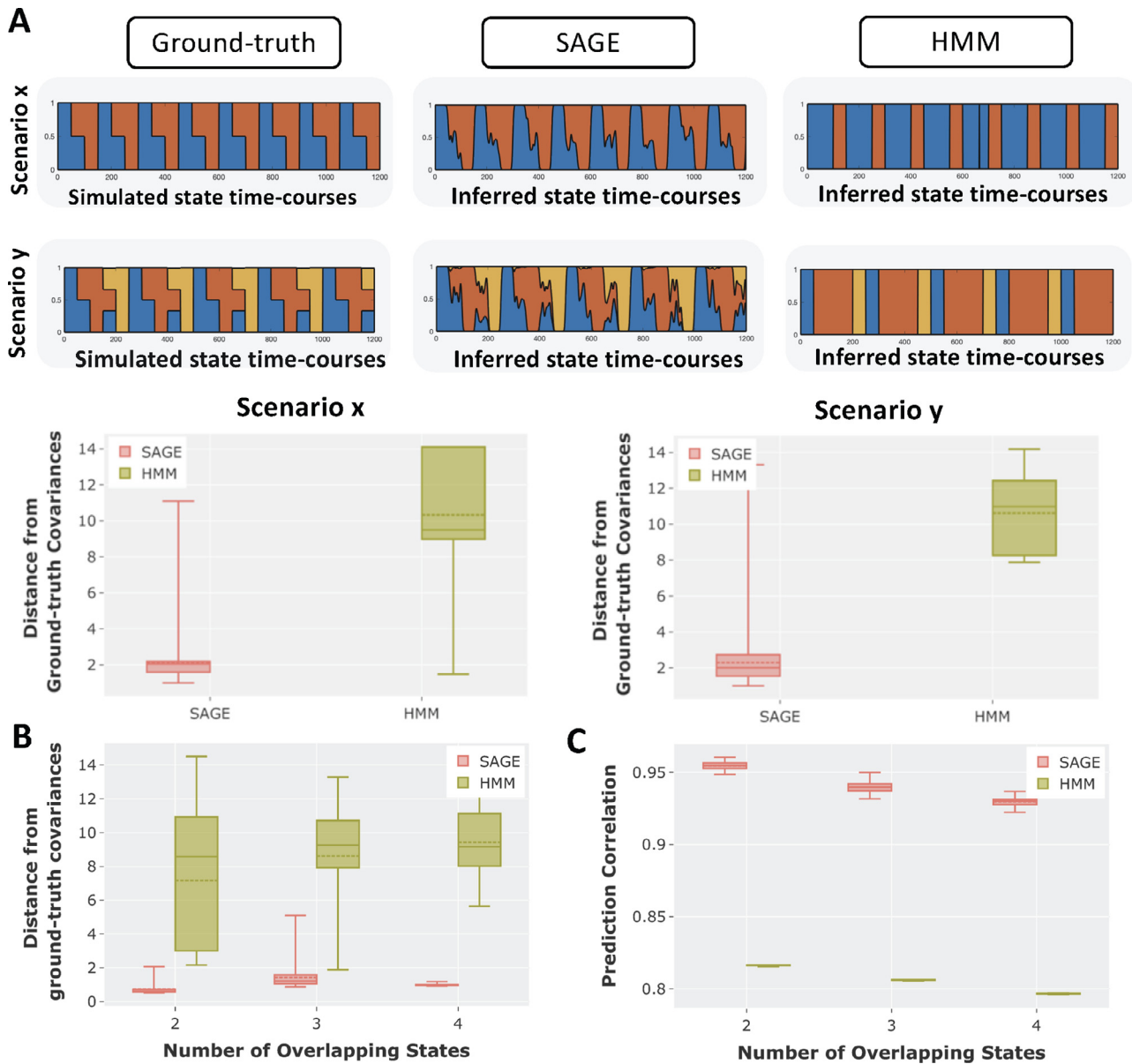
The hyper-parameters for MAGE were chosen using extensive trial and error by monitoring the validation loss. Some of the hyper-parameters are detailed as follows: 1) The sequence/window length,  $W$ , (number of data time-points) was set to 100 and batch-size,  $B$ , (number of grouped sequences) was set to 32 - this implies that each batch has a dimension of  $B \times W \times N$ . 2) a small learning rate of 0.0001 was used with large momentum value of 0.9. 3) Nesterov-accelerated Adaptive Moment Estimation (NAdam) optimizer was applied for training the data. 4) the number of units (the dimension of the inner cells in LSTM) for each LSTM layer was set between 8 to 64 (depending on the number of subjects). 5)  $\Lambda$  was optimised depending on the data-set (mostly constrained between 0.85 to 0.95 Higgins et al. (2016)). MAGE is an unsupervised algorithm, still, we were rigorous with the performance validation - we trained MAGE using the training/validation split with 80% data in the training subset and 20% data in the validation subset during any experiment.

To ensure that the reported MAGE results are reproducible, we performed cross-validations on different non-overlapping sets of subjects. We calculated the similarity of the MAGE estimated state-specific FC correlation and mean activity estimates across these different runs to highlight that the MAGE estimated states are replicable. For statistical testing conducted in this paper, we have performed paired  $t$ -test to determine whether there is statistical evidence that the mean difference between paired observations on a particular outcome is significantly different from zero. We have used symbols to report the test results; \* when  $P \leq 0.05$ , \*\* when  $P \leq 0.01$  and \*\*\* when  $P \leq 0.001$ .

## 3. Results

### 3.1. MAGE can infer dynamic linear mixtures of FC states

We start by considering the performance of our proposed method, MAGE, when we set aside the proposed multi-dynamic modelling. This version of MAGE models single-dynamics (where the state time courses for the mean activity and the correlations are the same, i.e.,  $\alpha_{tp} = \beta_{tq}$ ) and is referred to as SAGE. SAGE otherwise maintains all of the other features of MAGE, including the long-range temporal modelling and the partial volume modelling of the state dynamics.



**Fig. 3.** Comparison of the performance of the single-dynamic approach SAGE versus HMM, using simulations of dynamic linear mixtures of FC states (i.e., partial volume) and single-dynamics dynamics (i.e., where the state dynamics of the means and correlations are tied together). **[A]** Two illustrative partial volume examples, comparing SAGE performance with HMM. In scenario x, two states can be simultaneously active, whereas, in scenario y, up to three states can be simultaneously active. Note that these state-time courses were common to both the time-varying mean and correlations (i.e., instantaneous FC). The bottom row of **[A]** shows the Riemannian distance between the inferred and simulated covariances for both HMM and SAGE. **[B-C]** Comparison of the performance of SAGE with HMM on a number of partial volume simulations. Each box plot shows the distribution of prediction performance results over 50 different simulations. Riemannian distance between inferred and simulated covariances was used as a measure of success in **[B]**, and prediction correlation between the inferred and simulated state time-courses in **[C]**.

We first demonstrate SAGE’s ability to infer state time courses that are dynamic linear mixtures (i.e., partial volume mixtures) with more than one underlying state active at each time point,  $t$ . Note that this is in contrast to categorical simulations (i.e., HMM assumptions), where the ground-truth state time-courses,  $\alpha$  are either 1 or 0 (see Figure S3 for simulations demonstrating M/SAGE’s performance on categorical data). In the partial volume simulations, the ground-truth state time-courses,  $\alpha$  could be any continuous number from 0 to 1 for any state, but where the sum of  $\alpha$  for all states,  $\alpha_{jt}$  at any time-point,  $t$ , should be 1.

Figure 3 [A], shows illustrative examples where the number of simulated states is two and three respectively. Either one, two, or three can be active at any time point. The simulated ground-truth

in Fig. 3[A] is *single-dynamic* and the simulated state time courses (demonstrated in Fig. 3[A]) are common to both the time-varying mean and correlations. We compared SAGE with the HMM to see how they recover the underlying simulated state time courses and the observation model parameters. The HMM failed to infer these state time courses accurately, as HMM is built around the assumption that only one state is active at any time point. On the other hand, SAGE does a good job of inferring these partial volume state time courses. To quantitatively assess the performance of these models, we calculated the Riemannian distance between the simulated covariances and the inferred covariances for each time point. The lower the Riemannian distance between the inferred and simulated covariance, the better the performance of the model. We

can see that SAGE can accurately infer the underlying time-varying covariance, but this is not the case for the HMM.

Figure 3 [B,C] shows the performance of the approaches across many partial volume scenarios. We varied the number of overlapping states (either 2, 3, or 4), and then compared the performance of SAGE against the HMM. For each overlapping state configuration, we generated 50 different simulations in which we varied the state time courses (% of overlap), and the observation model itself. We illustrated the Riemannian distance between the inferred and simulated covariances. We also showed the prediction correlation (not accuracy, as state time courses, are no longer categorical) between the inferred and the simulated state time courses.

Lastly, we demonstrate the performance of MAGE for scenarios where the ground-truth is such that FC is static (i.e., no fluctuations in edges of the FC matrix over time). We show that MAGE does not infer time-varying correlations when we infer more than one state for such simulations (see Figure S4). This highlights that MAGE only infers time-varying FC when the underlying FC is time-varying.

### 3.2. Failing to model distinct mean activity dynamics can result in homogeneous estimation of FC over time

Commonly-used SWC and HMM approaches are what we refer to as single-dynamic models, as they tend to tie the dynamics of the correlations (i.e., the FC) to be the same as the dynamics of the means of the brain activity. However, if the underlying truth is that the dynamics of the means and correlations are actually distinct from each other, then using single-dynamic approaches to infer TVFC can have negative consequences. This is because forcing  $\alpha_{tp}$  to be the same as  $\beta_{tq}$ , when the ground truth is that they are not the same, causes the dynamics used for estimating the state-specific FC to be incorrect. This then leads to a smearing of the FC estimation over time, making it look more homogeneous than it really is.

We illustrate the mechanism of the multi-dynamic approach using a simple scenario in Fig. 1[A]. Furthermore, in Fig. 1[B], we demonstrate the potential benefits of the multi-dynamic approach on a simulation<sup>2</sup> in which the ground truth is such that the correlation dynamics and the mean activity dynamics are not tied together. Specifically, Fig. 1[B] shows the instantaneous FC estimated using SAGE or MAGE when we simulate using Eqs. (1)–(5) with  $\alpha_{tp}$  not equal to  $\beta_{tq}$ .

When we infer TVFC using SAGE, the single-dynamic version of MAGE (with  $\alpha_{tp}$  set to be the same as  $\beta_{tq}$ ), there are substantially reduced fluctuations in the inferred instantaneous FC over time (i.e., FC is mapped as being near-homogeneous/static) as is often observed in TVFC estimates in real fMRI data. In contrast, the multi-dynamic version of MAGE (with  $\alpha_{tp}$  free to be not equal to  $\beta_{tq}$ ) is better able to capture the true fluctuations in instantaneous FC over time.

### 3.3. MAGE can infer dynamics in the FC that are distinct from dynamics in the means of brain activity

We next sought to see if MAGE could identify multi-dynamics in a range of scenarios. For this, we simulated distinct dynamics (i.e., multi-dynamics) for the mean and correlations (i.e., the FC) by simulating two different state time courses,  $\alpha_{tp}$  and  $\beta_{tq}$ , using two different hidden semi-Markov model chains, whose state lifetime values were sampled from two different gamma distributions such

<sup>2</sup> For this particular simulation, the number of channels in the data is 20, the number of modelled states ( $P=Q=R$ ) is 12, and multi-dynamics state time courses for the means and the correlations are simulated using two different hidden semi-Markov model chains.

that  $\alpha_{tp}$  was not equal to  $\beta_{tq}$ . These state time courses were then used in Eqs. (1)–(5) to simulate the data.

We compared the performance of the multi-dynamic approach (MAGE) with the single-dynamic approach (SAGE) in Fig. 4[A,B]. As illustrated in Fig. 4[A], the lifetimes of both the mean and correlation states in the simulation were sampled from the gamma distribution. However, we fixed the distribution of lifetimes of the correlation state time courses by setting the gamma distribution parameter to  $k=30$ , whereas, for the mean state time courses, we varied  $k$  from 5 to 80, so that the mean state time courses spanned a variety of timescales in different scenarios. In Fig. 4[B], it is the other way around, where we fixed the lifetime distribution of the mean state time courses ( $k=30$ ), but varied the lifetime distribution for the correlation state time-courses (again from  $k=5$  to 80). In Fig. 4[A,B], we have plotted the prediction accuracy for the state time courses of correlation and mean. MAGE does a better job in predicting the underlying simulated labels for state time-courses as compared to SAGE. Note that we did not draw any comparison with the HMM here, as the HMM is also a single-dynamic approach, and the HMM is not able to infer the mean and correlation in such a multi-dynamic fashion (and hence the results would be expected to be closest to SAGE).

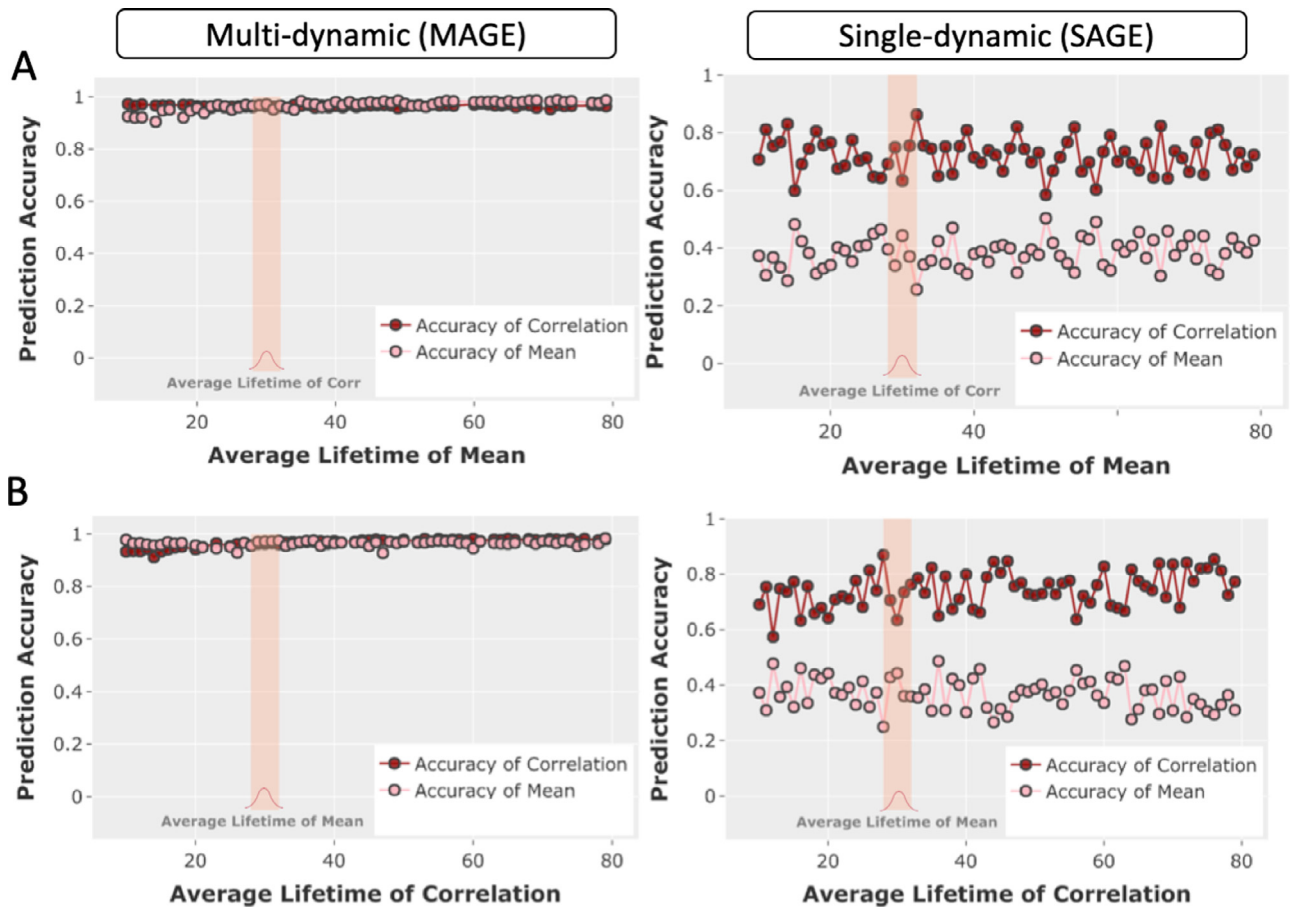
### 3.4. MAGE learns network state dynamics that show appropriate task dependencies

In order to help validate the different dynamics that MAGE is inferring, we applied it to task fMRI data. The idea is to train MAGE with no knowledge of the task timings (i.e., unsupervised), allowing us to see if the state dynamics being inferred by MAGE are meaningful in the sense that they exhibit dynamics that link to the task. Furthermore, we wanted to compare how the FC varies across the task conditions using the MAGE approach (and judge whether the inferred spatial maps are meaningful).

Task fMRI data used was collected as part of a previous study from 15 healthy volunteers (7 females, 8 males, age =  $27.3 \pm 4.4$ yr, all right-handed) without any previous neurological disorders (Kieliba et al., 2019; Duff et al., 2018). Participants performed four different tasks: rest, motor only, visual only, and simultaneous (but independent) visual and motor tasks, and each task was repeated four different times. The motor task involved continuous finger-tapping against the thumb using the right hand. The vision task consisted of videos of colourful abstract shapes in motion. During the combined motor-visual task, participants performed the finger-tapping task while simultaneously watching the videos. We applied a framework used for identifying large-scale probabilistic functional modes (PROFUMO) (Harrison et al., 2015; 2020), to identify parcels that are allowed to be correlated with each other in space and/or time, and which explicitly models between-subject variability in the parcels. We applied PROFUMO on this data with a dimensionality of 30, which outputs 30 parcel time-courses. These then correspond to the  $N=30$  channels of data that are fed into MAGE.

**Temporal Dynamics** In the main results shown in Fig. 5, we modelled four states using MAGE, so that the number of states is matched to the number of task conditions. In Fig. 5[A], we have plotted the task-evoked occupancy, which is the average of the MAGE inferred state time courses across the four repeated runs (as each experimental condition is repeated four times). The task-evoked occupancy shows clearly that the inferred states switch at the same time as the experimental task condition changes. It is worth noting from Fig. 5[A] that the estimated FC-driven state-time courses,  $\beta_{tq}$ , performed better than the mean activity/variance-driven time-courses,  $\alpha_{tp}$ . Moreover, in Figure S7, we have illustrated the results of task decoding when we modelled more than 4 states. In Figure S7[A], we have modelled 8 states and





**Fig. 4.** Comparison of the performance of the multi-dynamic approach (MAGE) and the single-dynamics approach (SAGE) on multi-dynamic simulations where multi-dynamics are the ground truth (i.e., where the state dynamics of the means and correlations are free to fluctuate independently). **[A,B]** The prediction performance of multi-dynamic and single-dynamic approaches are compared across a range of simulations. The x-axis shows the lifetime parameter (sampled from a gamma distribution) for the mean and correlation state time-courses, whereas the y-axis shows the prediction accuracy of the inferred state time-courses. Specifically, in **[A]**, we fixed the lifetime distribution of the correlation state time-courses ( $k = 30$ , as shown by the shaded bar) but varied the lifetime distribution for the mean state time-courses (from  $k = 5$  to 80), and it is the other way around in **[B]**.

in Figure S7[B], we have modelled 16 states, and we plotted the relative fractional occupancy (time spent in each state) distribution for each of the modelled states. It is evident that even with the increased number of states, we can still see dynamics that are task dependent and can identify task-specific states (as highlighted by the paired  $t$ -test values on the respective states).

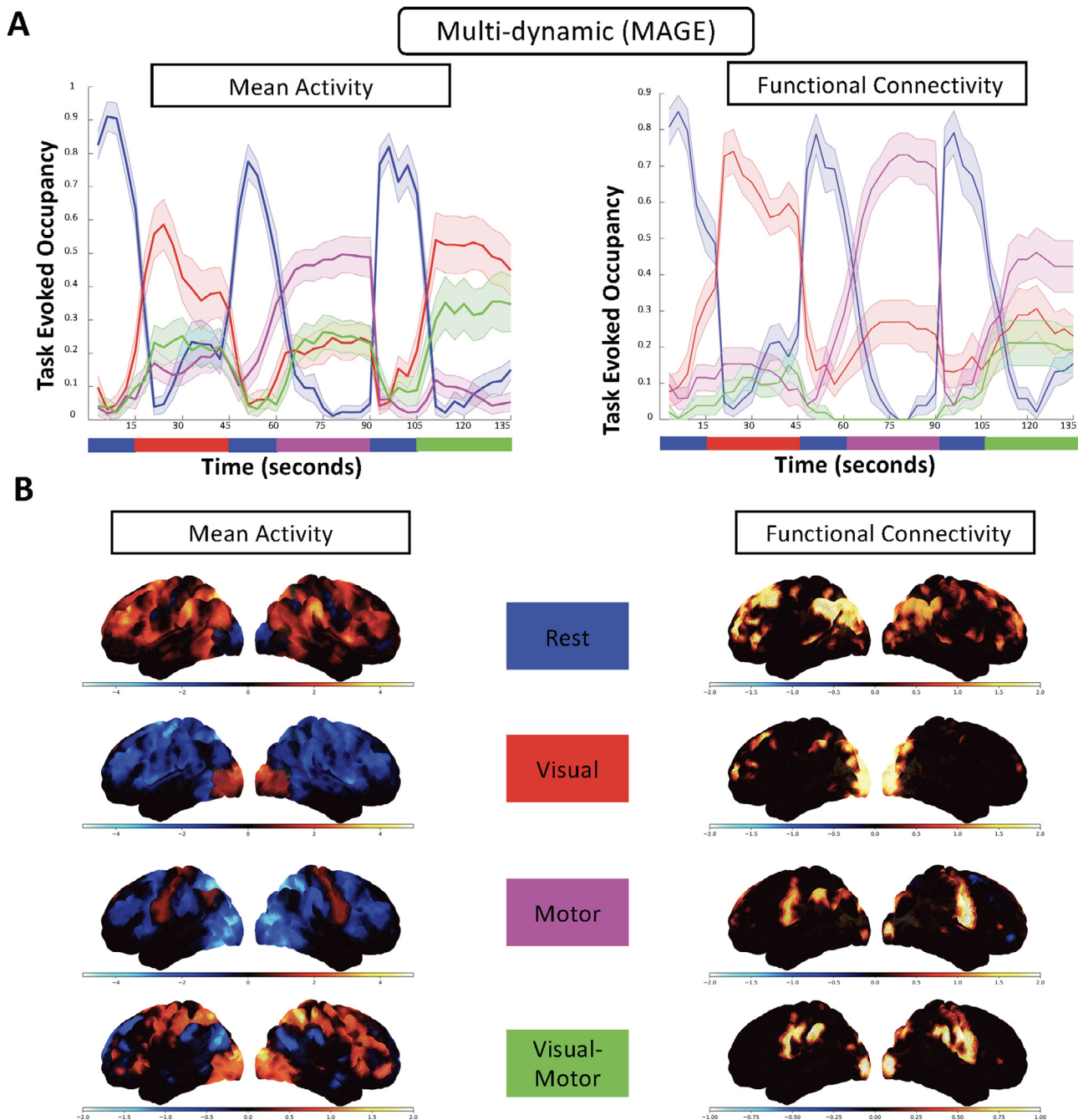
#### Spatial Dynamics

In Fig. 5[B], we show the state-specific FC spatial maps (corresponding to the first eigenvector of the state-specific correlation matrices) and mean activity spatial maps for each of the four inferred states. The variance spatial maps are not illustrated as there was no between-state variability in the inferred variance states (lack of time-varying content in variance). The maps are labelled according to the task conditions that their state time courses most correspond to. The state-specific FC maps for MAGE identifies appropriate active regions during each of the tasks. It can be seen during the motor task that the somatomotor area of the brain is active, and during the visual task that the visual cortex is active. During the rest task, the Default Mode Network (DMN) is active, and during the visual-motor task, both the motor and visual cortex are active. The mean activity level maps appear fairly similar and look appropriate with regards to the relevant task condition. During the visual task, the visual cortex of the brain is active, and during the motor task, the motor cortex is active for both approaches. During the rest task, DMN is active, and during the visual-motor task, only the visual cortex is active for the MAGE approach.

#### 3.5. Multi-dynamic approach reveals stronger changes in FC over time in resting fMRI

The simulations shown in Fig. 1 suggest that multi-dynamic approaches may be better able to identify TVFC than single-dynamic approaches, which may erroneously infer FC as being too homogeneous over time. We looked to test this idea by applying the single-dynamic (SAGE) and multi-dynamic (MAGE) approaches to resting-state fMRI data. In particular, we can analyse how the temporal dynamics of mean activity levels and FC differ (i.e.,  $\alpha_{tp}$ ,  $\beta_{tq}$ ) when modelled independently and see if this has any effect on the inferred state-specific FC maps. Lastly, we compared MAGE results with the HMM and SWC approaches, to further substantiate our findings.

We used rfMRI data from 13,301 subjects from the UK Biobank (UKB) (Miller et al., 2016) and 1003 subjects from the Human Connectome Project (HCP) (Van Essen et al., 2013). The pre-processing pipelines for UKB and HCP main steps can be summarized as (1) motion correction, (2) removal of structural artifacts with ICA (Independent Component Analysis and FMRIB's ICA-based X-noiseifier (FIX)). The length of the scanning session for UKB is 6 min, and for HCP is 1 h (4 sessions of 15 mins each). For UKB, we have 490 time-points and for HCP, we used data from all 4 sessions which gave us 4800 time points in total (1200 time points in each session). Lastly, we applied a popular data-driven parcellation scheme called spatial Group Independent Component Analysis (group-ICA)



**Fig. 5. [task fMRI data]** Validation of the performance of MAGE on a on a real task fMRI data-set. **[A]** There are four different tasks performed by the participants (rest, visual, motor, and visual-motor). Task-evoked occupancy plots of the state time courses from MAGE (learnt with no knowledge of the task timings) across the four tasks, where the x-axis represents the number of time points and the y-axis corresponds to the average activation value for that state. Specifically, the task-evoked plot curves show the mean and standard deviation across four repeated runs of each experimental condition for all 15 subjects. **[B]** The state-specific spatial maps for the functional connectivity (FC) and mean activity, which are labelled according to the task conditions that their state time courses most correspond to. The state-specific spatial maps for variance are not illustrated as there is no between-state variability in the inferred variance states. For the mean activity maps, the average amplitude of each node/channel is shown. The negative mean activity (e.g., deactivation of the visual network during the motor task) refers to decreases in signal change compared to the baseline. For the FC maps, rank-one decompositions (first Eigenvector) of the estimated state-specific FC correlation matrices are shown. The magnitude of the values of the state-specific FC spatial maps (colorbars) correspond to the degree of connectedness of each region with the rest of brain.

(Pervaiz et al., 2020)<sup>3</sup> to our data to identify the parcels/channels (N) that form the data to be input into MAGE<sup>4</sup>

<sup>3</sup> For HCP data, we use the minimally-preprocessed released data, in gray ordinates (cortical surface vertices and subcortical voxels). For UKB data, we used the preprocessed released data in standard volumetric space.

<sup>4</sup> To apply group-ICA, we first generated the group-PCA (Principal Component Analysis) output using MIGP (MELODICs Incremental group-PCA) (Smith et al., 2014)

*Temporal and Spatial Dynamics* We first compare the single-dynamic (SAGE) and multi-dynamic (MAGE) approaches using the HCP data, where the only difference between the two methods is

using data from all subjects. This comprises the top few thousands (1000-5000) weighted spatial eigenvectors from a group-averaged PCA, which is a very close approximation to the original data (fully concatenating all subjects' time-series. Lastly, group-ICA is run on the output of group-PCA.

whether or not we model independent state dynamics for the state means and FC (i.e., correlations). In Fig. 6[A], we show the state-specific mean activity and FC correlation maps for SAGE and MAGE approaches for the HCP data. In total, we modelled  $P=Q=R=12$  states, but we have only displayed the observation model for four of the 12 states for ease of visualisation (the results for all 12 states are illustrated in Figure S9, Figure S10, Figure S11 and Figure S12). As described earlier, MAGE also models time-varying variances, whose dynamics are tied to be the same as the mean. However, we have not displayed the state-specific spatial maps for variance as MAGE infers that there was little between-state variability in the state-specific variances (i.e., the variances remained static across states).

Figure 6 [A] shows that the state-specific mean activity maps vary reasonably well across states using both SAGE and MAGE. However, with SAGE, it can be seen that the state-specific FC maps (the rank-one decomposition of the estimated state-specific correlation maps are shown to make these 2D FC maps easier to view) of all the states are predominantly similar, always resembling a somatomotor state. We also attempted another modelling option with SAGE in which we explicitly forced the mean activity to be zero<sup>5</sup> (i.e.,  $m_t = 0$  in Eq. (1)). This approach also resulted in homogeneous state-specific FCs as shown in Figure S13. In contrast, for the MAGE approach, we can see that the state-specific FC maps correspond to distinct brain networks, with the state-specific FCs looking more distinct from each other than in the SAGE results. To illustrate this quantitatively, we show the pairwise between-state Riemannian distance in Fig. 6[B]. Specifically, we calculated these pairwise distances over the actual 2D correlation matrices (i.e., FCs). It is clear that these between-state pairwise distances are higher for the state-specific FC maps when inferred using the multi-dynamic MAGE approach, revealing stronger changes in FC over time.

In Fig. 6[D], we compare summary measures of the state dynamics. Since MAGE models a linear mixture (i.e., a partial volume) at each time point, it is not straightforward to compute summary statistics normally used on binary state time courses. In order to allow us to do this we threshold the continuous partial volume state time courses ( $\alpha_{tp}$  and  $\beta_{tq}$ ) at 0.3. Firstly, we show the state switching rate for both models, which is a measure of the frequency of state fluctuation for each subject, and it can be seen that FC-driven state time-courses,  $\beta_{tq}$ , have the lowest switching rate. Then, we show the state lifetimes for both models, which correspond to the number of seconds per state visit. We can see that the FC-driven state time-courses have the highest values for state lifetimes (2 to 5 s) as compare to lifetimes of mean state time-courses (2 to 3 s). In Fig. 6[E], we show an example segment of the state time course (of a randomly chosen subject) showing the nature of the mixing of the inferred 12 states and the degree of overlap between the mean activity and the FC state time-courses. Furthermore, in Fig. 6[C], we quantitatively compare the extent of similarity between the mean activity and the FC state time courses. We plot the mean activity and the FC state time courses correlation matrices as histograms for the null and the HCP dataset. There is a significant non-zero correlation as compared to the equivalent null scenario, but there is clearly still a distinct difference between them (and hence the difference in the performance of SAGE vs MAGE). In short, the means and FCs have some shared dynamics but also some unique/independent dynamics.

It has been previously proposed that the observed estimates of TVFC must be compared against “static” null hypotheses to establish claims of non-stationary (Lurie et al., 2020). Considering that,

<sup>5</sup> For other such examples, see Vidaurre et al. (2021) for modelling only the covariance and Charquero-Ballester et al. (2020) for modelling only the mean activity.

we generated null data using autoregressive randomization (ARR) following (Liegeois et al., 2017). Specifically, a 10th order Gaussian autoregressive (AR) model was applied to generate null data. For each participant and each pair of brain regions, we compared the observed between-timepoint variances of the instantaneous FCs inferred by MAGE, which can be calculated from the FC state dynamics and state-specific FCs, from the real HCP data against the null data. This analysis resulted in one p-value for each edge of the FC matrix. On average, across HCP participants (compared against null data), 95% of edges were significant ( $p \leq 0.001$ ). Furthermore, in Figure S5, we show the histogram of null variance values compared with the real HCP data variance values (this highlights that between-brain region variability over time is significantly higher in the real HCP data as compared to null data).

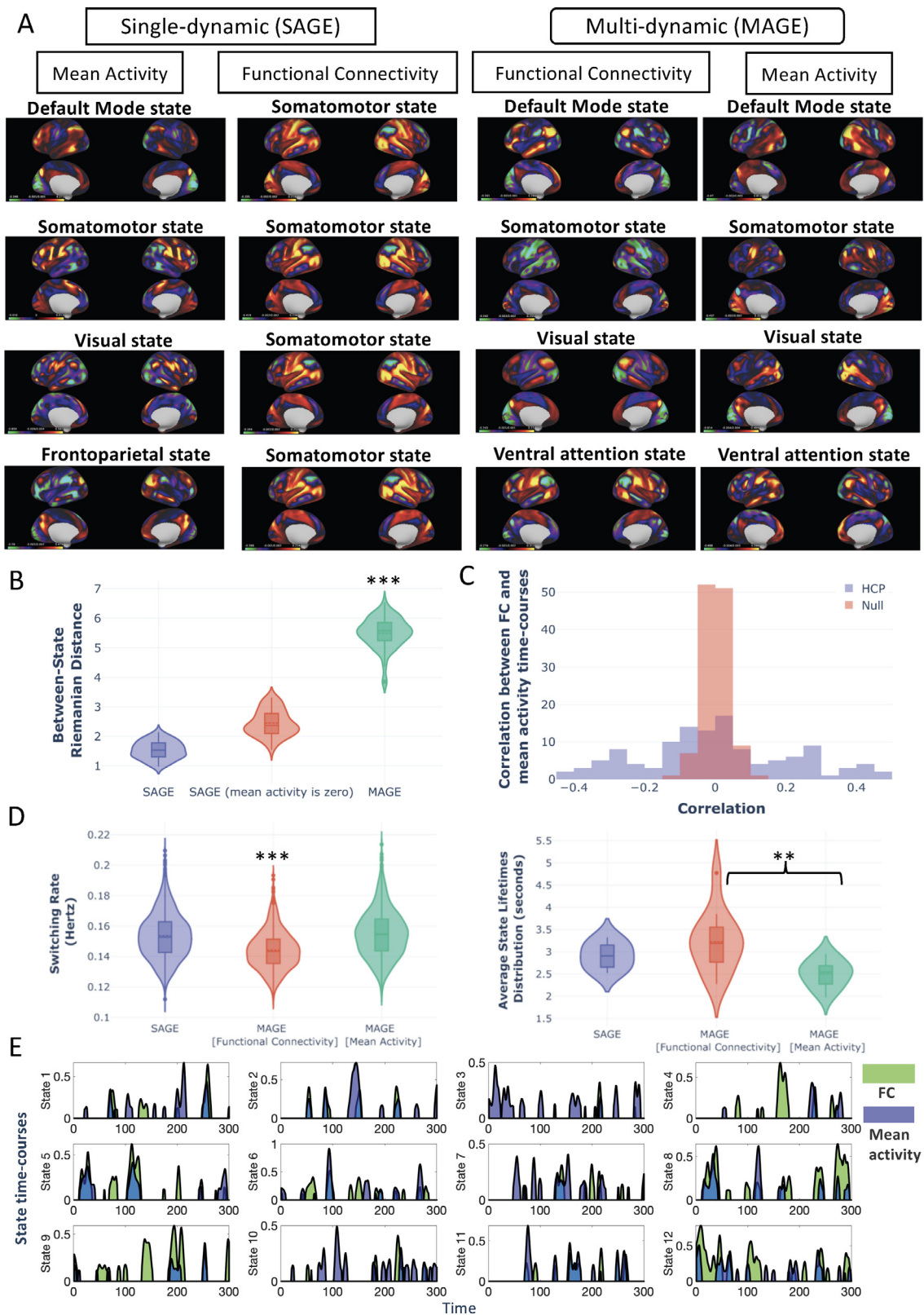
We find that we get similar results to the HCP dataset when we use MAGE on the resting fMRI data from the UKB dataset. Here, we also compare the performance of MAGE with other established methods (HMM and SWC) on the UKB data-set. Figure 7 shows the state-specific FC maps from MAGE, HMM and SWC, with the HMM and SWC approaches showing similar behaviour to that of SAGE in Fig. 6. It is also clear that there are smaller between-state differences in FC across time for both the HMM and SWC approaches compared to MAGE (as reflected by the between-state Riemannian distance plots in Fig. 7[D]), consistent again with the idea that MAGE reveals stronger changes in FC over time due to its ability to model multi-dynamics.

To ensure that the MAGE estimated FC and mean activity states are reproducible, we ran MAGE six times on different non-overlapping subsets of UKB data (2k subjects for each run). In Fig. 8[A,B], we compare the similarity between MAGE estimated state-specific spatial maps across the six runs. Specifically, in Fig. 8[A], we show the average correlation plot for the state-specific FC estimates across repeated runs, and in Fig. 8[B] for the state-specific mean activity estimates, both for a fixed number of states ( $P=Q=12$ ). This shows that the state-specific FCs are very reproducible and that the state-specific mean activity estimates are moderately reproducible but not to the same extent. We then replicated these reproducibility findings for a wide range of numbers of states (1 to 24) for the FC and mean activity, as shown in Fig. 8[C], also demonstrating that P does not have to be equal to Q. We show in Fig. 8[C] that the FC are mostly highly and mean activity are mostly moderately reproducible independent of the number of modelled states. Specifically, mean activity and functional connectivity states are more reproducible if the number of modelled states is more than approximately six.

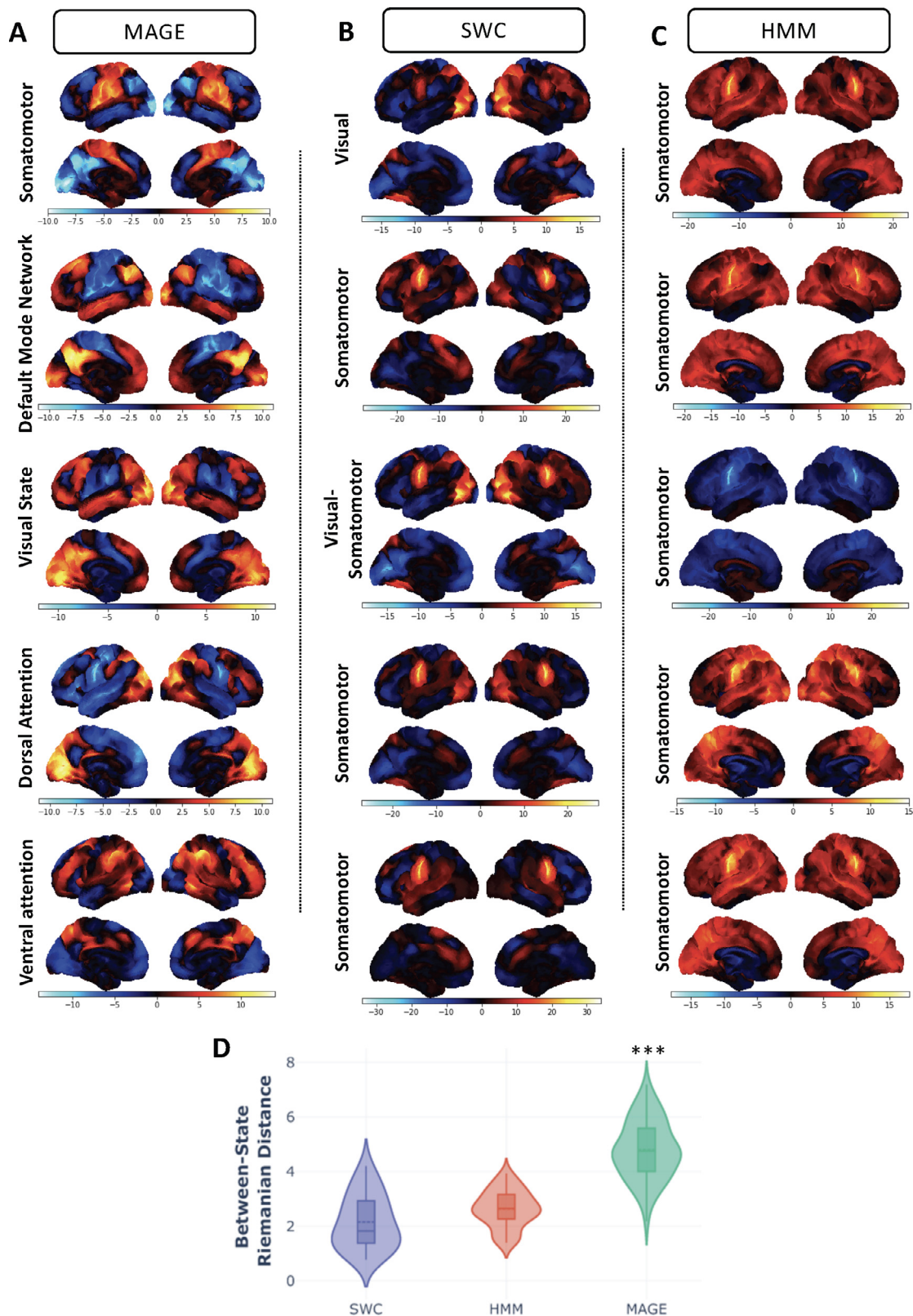
It has been previously reported in Vidaurre et al. (2017) that the brain network activity is organized into a hierarchy of two distinct meta-states, with one meta-state representing higher-order cognition, and the other representing the sensorimotor systems. We have taken a similar approach as Vidaurre et al. (2017), and computed  $[P \times d]$  or  $[Q \times d]$  matrices containing the fractional occupancy (time spent in each state) for each subject, where d is the number of subjects. We then use these to compute the  $[P \times P]$  or  $[Q \times Q]$  correlation matrices over subjects, which are shown in Fig. 8[D]. As in Vidaurre et al. (2017), Fig. 8[D] shows that a hierarchical meta-state structure is present. Here, it is apparent in both in the functional connectivity and the mean activity driven state time-courses, although it is more apparent in the mean activity state time-courses.

### 3.6. Multi-dynamic TVFC predicts behavioural variability better than the single-dynamic approaches

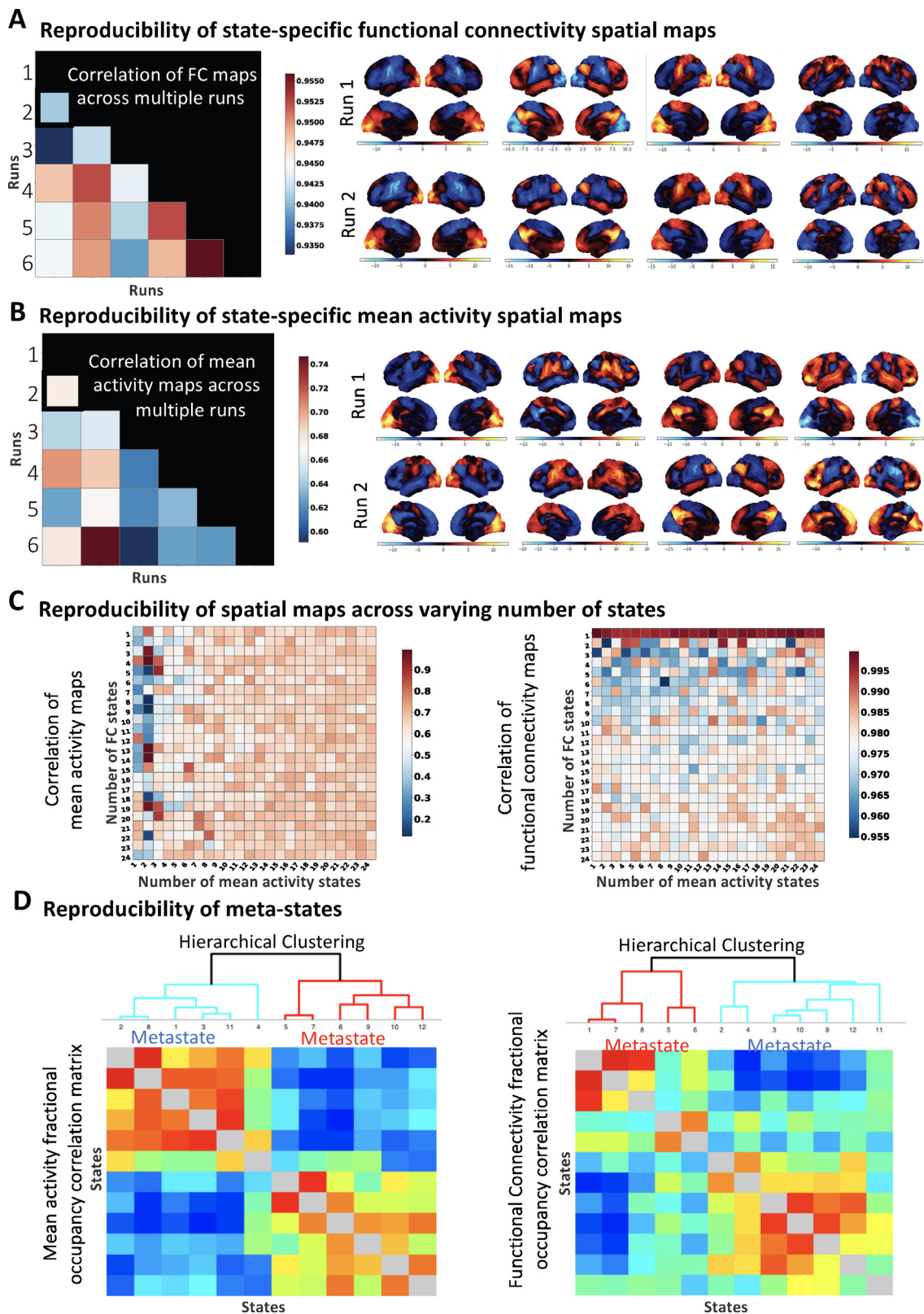
We next used the UKB data to see if the multi-dynamic MAGE approach provides spatial network features (i.e., the state-specific FCs) that can be used to predict individual behavioural traits, and



**Fig. 6. [HCP data] [A-B]** The multi-dynamic (MAGE) approach reveals stronger changes in state-specific FC maps compared to the single-dynamic (SAGE) approach. Using the HCP data, it is evident that qualitatively and quantitatively, MAGE-derived FC shows stronger changes over time. **[A]** State-specific mean activity and FC correlation maps for four of the modelled states for both approaches are shown. For the state-specific mean activity maps, the average amplitude of each node/channel is shown and for the state-specific FC maps, rank-one decomposition (first Eigenvector) of the estimated FC correlation maps are shown. **[B]** The pair-wise Riemannian distance of the state-specific FC correlation matrices between 12 FC states is shown for both MAGE and SAGE (with and without modelling the mean activity, see [Vidaurre et al. \(2021\)](#) as another example of modelling only the correlation, i.e.,  $m_t = 0$  in Eq. (1)). The stars referred to paired  $t$ -test values. **[C]** Correlations between the mean-derived state time-courses and the FC-derived state time-courses plotted as histogram over all pairs of states (and equivalent null scenario). The non-zero correlations indicate that the mean activity and FC have some shared dynamics but also have unique temporal dynamics. **[D]** Summary measures of the state dynamics for both approaches. It can be seen that FC state dynamics show the lowest switching rate (in Hertz) and the highest lifetimes values (in seconds) as compared to other approaches. **[E]** Example segment of a state time course of a randomly chosen subject is shown to qualitatively highlight the extent of similarity (and dissimilarity) between the mean activity and the FC state time courses. These also show how states in MAGE can overlap in time.



**Fig. 7. [UKB data]** Demonstration that the between-state spatial variability in state-specific FC maps is significantly higher when modelled using the MAGE approach on UKB data compared with the HMM or SWC approaches. **[A-C]** The pair-wise Riemannian distance of the state-specific FC correlation matrices between 12 FC states is shown for MAGE, HMM and SWC, alongside the FC correlation maps for 5 of the identified states. The state-specific FC maps are shown as rank-one decomposition (first Eigenvector) of the estimated state-specific FC correlation matrices. The channels with positive values (red regions) show they are positively correlated with other regions, whereas channels with negative values (blue regions) show they are negatively correlated with other regions. The state-specific FC maps for all 12 states using SWC, HMM and MAGE are illustrated in Figure S14, Figure S15, and Figure S16 respectively. **[D]** The between-state Riemannian distance for the modelled 12 states, shown for MAGE, SWC and HMM. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)



**Fig. 8. [UKB data]** MAGE estimated FC and mean activity spatial maps are reproducible across various runs on different non-overlapping sets of subjects. **[A]** shows the average correlation between the state-specific FCs across six repeated runs. Moreover, for runs 1 and 2, FC correlation maps are shown for 4 of the identified states (out of the 12 inferred states), demonstrating that the FC states are consistent across repeated runs. On the other hand, **[B]** shows the average correlation plot for the state-specific mean activity estimates across six repeated runs, alongside the mean activity spatial maps (for runs 1 and 2 only). The analysis in **[A]** and **[B]** is performed for a fixed number of states ( $P = Q = 12$ ), whereas, in **[C]**, we evaluated the reproducibility of state-specific FC and mean activity estimates by varying the total number of modelled states ( $P, Q$ ). Specifically, **[C]** shows the correlation plots for the state-specific FCs and state-specific mean activity estimates, where we varied the total number of inferred states ( $P$  and  $Q$ ) from 1 to 24. These results suggest that inferred states for the FC are highly and mean activity are moderately reproducible independent of the number of modelled states (Note the difference in the ranges of the colormaps). **[D]** We replicated the original findings of (Vidaurre et al., 2017) and illustrate that hierarchical meta-state structure exists both in the functional connectivity and the mean activity states, but is more well-defined in the mean activity states (matrices computed as  $\nu\nu'$ , where  $\nu$  is the [number of states by number of subjects] fractional occupancy matrix).

we compare its performance to the HMM and SWC approaches. Alongside the task fMRI results, this would provide further evidence that the state dynamics being inferred by MAGE are cognitively meaningful.

We applied the dual-estimation method to obtain subject-level estimates of TVFC. This means that for each subject, we can obtain a set of state-specific FC maps ( $Q \times N \times N$ ), where  $Q$  is the number of correlation (FC) states and  $N$  is the number of channels/brain regions. Then, we can use these subject-level, state-specific FC estimates to predict selected behavioural variables. The classifier/predictor we employed is Elastic Net (Zou and Hastie, 2005), which is a regularized regression method that combines the penalties of lasso ( $L_1$ ) and ridge ( $L_2$ ) methods (Vidaurre et al., 2013). The data matrix fed into the classifier is  $d \times [Q \times f]$ , where  $d$  is the number of subjects,  $Q$  is the number of states and  $f$  is the number of features. The state-specific FC matrices are symmetric and only values above the diagonal need to be retained and vectorized, and hence  $f = N(N-1)/2$ . Lastly, we concatenated the state-by-state subject-level, state-specific FC estimates, resulting in  $Q \times f$  features for each subject. We performed this prediction analysis for UKB data with  $d = 13301$ ,  $Q = 12$ , and  $N = 25$ .

We used four non-imaging variables to compare the performance of multi-dynamic (MAGE) approach with HMM and SWC. The details for the SWC and HMM approaches are explained in the Section S1.3 and Section S1.4 respectively. These variables are age, sex, fluid intelligence score, and neuroticism score. In Fig. 9[A], we display the prediction results (nested 10-fold cross-validated) from the subject-level FC estimates. We can see that the FC obtained using the MAGE approach is a more accurate predictor of behaviour as compared to the other approaches

### 3.7. Static FC is worse than multi-dynamic TVFC at predicting behaviour

Time-averaged FC (static) can be considered as an average of all state-specific FC states (and the time varying means) and is conventionally estimated simply by taking a full (Pearson) correlation between brain parcels over the whole scanning time. Unlike TVFC in the form of the state-specific FCs, time-averaged FC is a single FC estimate per subject and is already an established predictor of behavioural traits (Pervaiz et al., 2020). In Fig. 9[B], we compare the combined power of the state-specific FCs with the time-averaged FC. This shows that for the mental health variables, predictive performance is improved by using the state-specific FCs in comparison to time-averaged FC. We employed Elastic Net for this comparison and the reported results are nested 10-fold cross-validated.

## 4. Discussion

We have proposed a novel algorithm, MAGE, that reliably infers the time-point by time-point estimates of FC. Notably, MAGE frees up the dynamics of the FC to fluctuate independently from the dynamics of the means and variances of the brain activity. We found that this multi-dynamic approach reveals much stronger changes in FC over time than existing single-dynamic approaches (e.g., HMM, SWC), and is a better predictor of individual behavioural variability. This provides a potential explanation and solution as to why time-varying FC has appeared to look so stable in resting fMRI data in previous work (Hindriks et al., 2016; Liegeois et al., 2017).

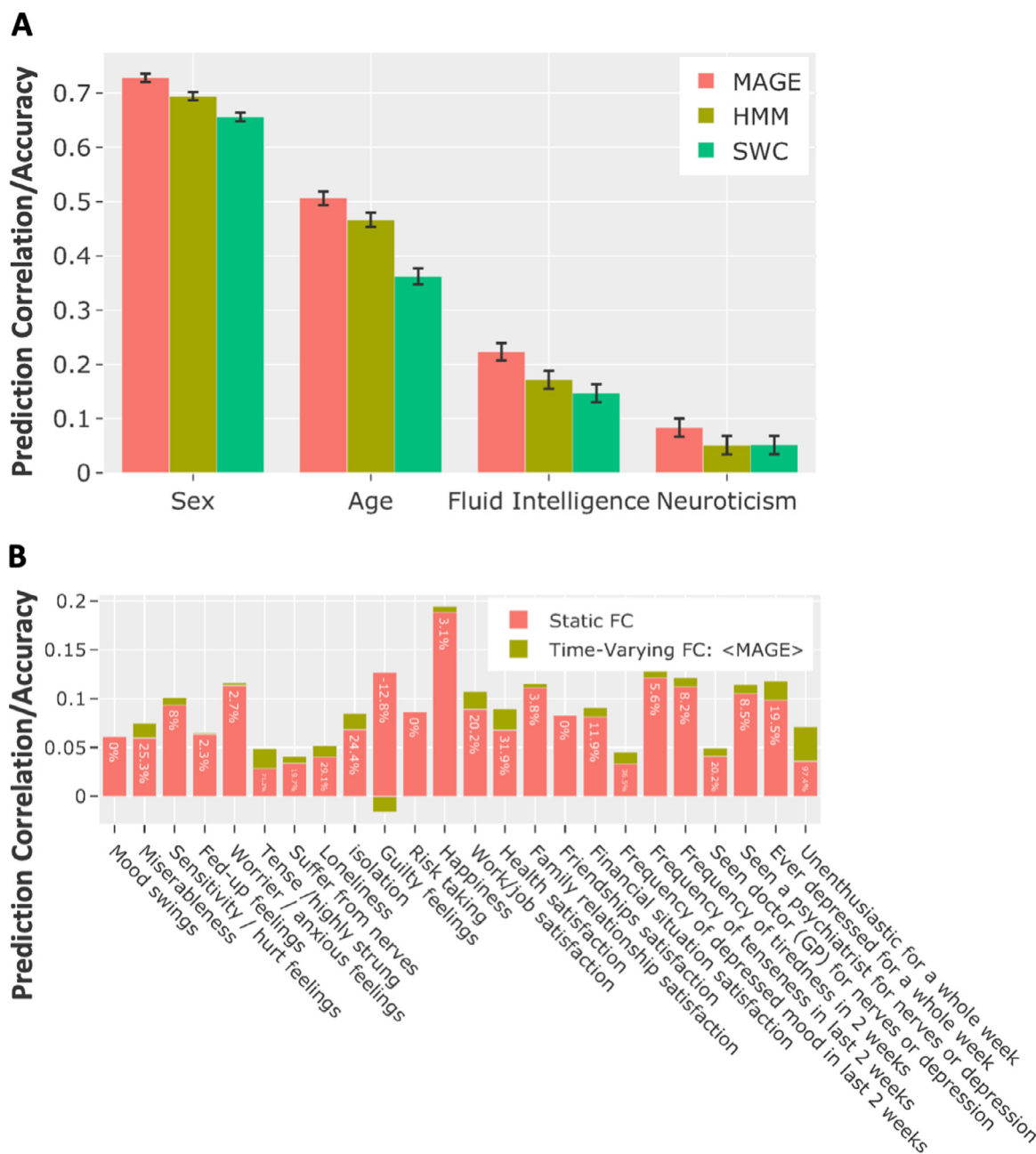
**MAGE inferred multi-dynamic linear mixtures of FC states.** We assessed the performance of MAGE on simulations where there can be more than one underlying state active at each time point (Fig. 3). We demonstrated that categorical models, such as the HMM, under-performed as expected when applied to such simulations. In contrast, MAGE remained accurate across a range of

these simulations. MAGE also shows improved performance in simulations designed to be suitable for categorical models (e.g., HMM), i.e., by having only one state active at each time point (mutual exclusivity) (Figure S3). This suggests that MAGE performance remained stable when the state time-courses have long-range temporal dependencies (i.e., violate the HMM Markovian assumption) (Figure S3[D]). Lastly, we assessed the performance on the multi-dynamic simulations and demonstrated that MAGE accurately identified unique dynamics of the correlations when the ground truth was that the correlations and mean activity dynamics were distinct from each other (Fig. 4).

**MAGE learned network state dynamics that showed appropriate task dependencies.** Following rigorous performance evaluation of MAGE on more than a thousand simulations, we applied it to a task fMRI dataset. The rationale was that task fMRI gives an opportunity to assess the specific state dynamics that MAGE infers, by comparing them to the known task timings. In Fig. 5, we demonstrated that the FC-driven time courses performed well in terms of displaying task-dependencies as compared to the mean-driven state time courses when using MAGE. Moreover, we showed in Fig. 5[B] that plausible time-varying representations of FC and mean activity spatial maps were inferred using the MAGE approach. This suggests that the inferred TVFC is powerful enough to capture the dynamics in cognitive tasks, and is not static during these task experiments.

**Why do we model the mean of the brain activity separately from the functional connectivity?** Our assumption in Eq. (1) is that the mean and the FC (i.e., the correlation matrix) continuously vary over time. From that perspective, there is no imperative to assume that the dynamics of those two entities are the same as each other. As such, MAGE is designed to explore what happens when we do not tie the dynamics together. By allowing the state dynamics of the mean and FC to be independent of each other, MAGE also allows them to potentially operate on different time scales. However, in real fMRI data, we found that the dynamics of the mean states and the FC states ended up on fairly similar time-scales (between 2 to 5 s for FC states and 2 to 3 s for mean activity states, as shown in Fig. 6). It is also interesting to consider the extent to which the state dynamics of the mean and FC get inferred as being coupled together, despite being free to be inferred as completely independent. Figure 6[C] demonstrated that there was indeed some significant non-zero correlation between the two types of dynamics. However, this correspondence between the dynamics of the mean and the dynamics of the FC was not absolute. In other words, the fluctuations in the FC clearly have some unique temporal structure that is not shared with the fluctuations in the mean of the activity. Indeed, the benefit of capturing these unique fluctuations was demonstrated by the improved performance of MAGE (versus SAGE) in inferring stronger fluctuations in FC over time and in predicting behavioural traits.

**Why does MAGE reveal stronger functional connectivity dynamics in resting-state fMRI?** Our results demonstrated that MAGE revealed stronger changes in FC over time than single-dynamic approaches, where changes in FC over time correspond to between-state differences in the state-specific FCs (e.g., HMM, SWC, SAGE, see Figs. 6 and 7). The lack of change in FC over time in single-dynamic approaches has been observed in other, recent studies (Laumann et al., 2017; Hindriks et al., 2016; Leonardi and Van De Ville, 2015). Indeed, in some cases, it has been observed that the reported variability in the FC from one state to another state (or against null models) is statistically insignificant (Liegeois et al., 2017). So why does MAGE show comparatively stronger changes in FC over time? We can gain insight into the multi-dynamic model being used in MAGE, by considering a dynamic version of structural equation modelling (SEM), as explained in Appendix A. This shows that if the underlying truth is that there



**Fig. 9.** Multi-dynamic (MAGE) derived state-specific FCs are used to predict non-imaging variables in the UKB data. **[A]** The prediction performance of MAGE-derived state-specific FCs compared with the HMM and SWC approaches. The prediction accuracy/correlation is estimated using the combined features by concatenating the state-specific FC estimates from all states. Y-axis shows the prediction correlation (age, fluid intelligence, neuroticism) and prediction accuracy (sex), and the x-axis shows the predicted variable. To draw the confidence intervals on the prediction accuracies for sex variable (discrete output), we performed the Wilson test (suitable for the binomial distribution). Lastly, to generate confidence intervals on the prediction correlations (i.e., continuous outputs e.g., age, fluid intelligence and neuroticism), we computed the Fisher transformation (Pervaiz et al., 2020). **[B]** Comparison of the prediction performance of MAGE-derived state-specific FCs with the time-averaged FC. The bars in red colour show the prediction correlation/accuracy using time-averaged FC estimates, whereas bars in green show the increased/decreased prediction correlation/accuracy using state-specific FC estimates. The numbers in red bars represent the percentage in the change of performance using the state-specific FCs as compared to time-averaged FC. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

are distinct dynamics for the external driving inputs and connectivity, then distinct dynamics are needed for the mean activity and FC modelled by MAGE. If we were to ignore this, and use a single dynamic approach (e.g., HMM, SAGE, SWC) then we will infer common state dynamics that are a compromise. In turn, use of these incorrect state dynamics will tend to cause the inferred FC patterns to be temporally smeared versions of the true FC patterns, causing them to look more homogeneous over time. We demonstrated this phenomenon with a simple example (Fig. 1[B]), in which we simulated a ground truth where the FC dynamics and the mean

activity dynamics were not tied together. Using SAGE on this data substantially reduced the inferred change in instantaneous FC over time, whereas MAGE better captured the true variability in instantaneous FC across time. This work provides a potential explanation as to why FC in resting fMRI has previously appeared to look so homogeneous over time (Laumann et al., 2017; Hindriks et al., 2016; Liegeois et al., 2017).

**Reproducibility of MAGE states.** The presented results in Fig. 8 demonstrated that the MAGE estimated FC states are highly and mean activity states are moderately reproducible across re-



runs on non-overlapping subsets of the data (shown quantitatively and qualitatively in Fig. 8[A,B]). Also, as is typical with other unsupervised learning methods (e.g., ICA, HMM), the number of states to be modelled is an empirical parameter in our method and it needs to be chosen before running the model. In most of our presented work, we modelled anywhere from 4 to 16 states, though we had experimented with up-to 64 states in the resting-state and task data (and the results remained sensible). In Fig. 8[C], we found out the estimated states are more consistent across re-runs when the number of modelled states is more than six. Considering that, we suggest that the number of modelled states to be more than approximately six for such dynamic modelling (though we do not have a strong recommendation for an exact number of states).

**Multi-dynamic TVFC predicted behavioural variability better than existing methods.** We evaluated whether the inter-individual differences inferred by MAGE are meaningful. We demonstrated in Fig. 9[A] that the MAGE-estimated, state-specific FCs did a better job at predicting these non-imaging variables as compared to SWC and HMM approaches. For example, age was predicted with an average correlation of 0.50 using MAGE as compared to 0.46 using HMM, and 0.36 using SWC. This provides evidence that the TVFC estimated using the MAGE is a more accurate and meaningful representation of underlying functional networks and is indeed a powerful biomarker for cognitive traits.

**Static FC is worse than Multi-dynamic TVFC at predicting behaviour.** In Fig. 9[B], we compared the predictive power of MAGE-derived, state-specific FCs versus time-averaged FC, and demonstrated the prediction performance on mental-health related variables in UKB data. We have demonstrated that the prediction accuracy of these behavioural variables increased when TVFC was used, as compared to time-averaged FC e.g., for work/job satisfaction, prediction accuracy was improved by 20.2% and for depressed for a week, accuracy was improved by 19.5%, etc. This highlights that there are some aspects of behaviour that can be best explained using TVFC estimates. The better performance of time-varying FC in terms of behavioural variability might be because it should not contain as much potential confounding anatomical information as compared to time-averaged FC (Vidaurre et al., 2021); however, further detailed work is required to understand and address this issue.

**What do we mean by TVFC?** We can gain insight into the multi-dynamic model being proposed, by considering a dynamic version of structural equation modelling (SEM) as explained in Appendix A. The SEM perspective (Appendix A) suggests that the instantaneous FC is the closest representation that we have to the time-varying connectivity in a dynamic SEM (Eq. (19)). In MAGE, we have assumed that the instantaneous FC,  $C_t$ , is the TVFC measure of most interest. MAGE provides a regularised estimate of the instantaneous FC, and so in this paper we have represented the instantaneous FC parsimoniously using the state-specific FCs. Hence, in this work we have focussed on the state-specific FCs as our main measure of TVFC, e.g., for predicting behavioural traits. However, a more complete representation of TVFC includes the FC state time courses alongside the state-specific FCs (after all, we combine the state-specific FCs with the FC state time courses via Eq. (4) to get the instantaneous FC). As such, prediction may be improved further in the future by including features describing the FC state dynamics. For example, we can already show that inter-individual differences for the time spent in each FC state are linked to behavioural traits, as illustrated in Figure S8. MAGE also provides other potential features for use in prediction, for example, the mean-activity FC (i.e., correlations of the mean activity,  $m_t$ , over all time) or the state-specific mean activity and the mean-derived dynamics. In this paper, we have not looked into the details of mean-activity FC but mean-activity FC has been assessed in the literature. For example, Miller et al. (2015) highlights that

the spatiotemporal organization of brain activity correlates with the presence of diagnosed schizophrenia, as well as with gender and age.

**Variance of the brain activity.** We found through resting-state and task data-sets results (Figs. 6 and 7) that there was not much between-state variability in the variance of the activity, and the variance remained static over the scanning session. Nevertheless, we explored another configuration where dynamics of the variance of the activity was modelled with independent state dynamics from the mean activity and the FC, but that also did not result in any meaningful variability in the variance of the activity (though it is computationally more expensive to learn the variance state time courses separately). We also attempted to model state dynamics of the variance of the activity tied to the state dynamics of the FC fluctuations, but independently from the mean activity; however, this resulted in comparatively less between-state variability in the FC. In summary, we found out there was little between-state variability in the state-specific variances irrespective of the modelling choice.

In summary, we proposed a multi-dynamic approach (MAGE) that models temporal fluctuations in functional connectivity independently from fluctuations in the mean of the activity. Multi-dynamic modelling provided an explanation and a solution as to why resting fMRI FC has previously looked so stable and homogeneous. We found out that that separating fluctuations in the mean activity levels from those in the functional connectivity reveals much stronger changes in functional connectivity over time. Lastly, MAGE estimated time-varying FC is a better predictor of behavioural variability in the resting-state fMRI data than established methods.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## CRediT authorship contribution statement

**Usama Pervaiz:** Data curation, Conceptualization, Methodology, Writing – original draft, Validation, Visualization. **Diego Vidaurre:** Methodology, Writing – review & editing, Supervision. **Chetan Gohil:** Methodology, Writing – review & editing. **Stephen M. Smith:** Data curation, Conceptualization, Writing – review & editing, Validation, Supervision. **Mark W. Woolrich:** Conceptualization, Methodology, Writing – original draft, Writing – review & editing, Validation, Visualization, Supervision.

## Acknowledgements

Computation used the Oxford Biomedical Research Computing (BMRC) facility, a joint development between the Wellcome Centre for Human Genetics and the Big Data Institute supported by Health Data Research UK and the NIHR Oxford Biomedical Research Centre.

We are extremely grateful to all UK Biobank (Data Access Application 8107) and Human Connectome Project participants who generously donated their time to make this resource possible. UK Biobank (including the imaging enhancement) has been generously supported by the UK Medical Research Council and the Wellcome Trust. Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University.

We additionally thanks Eugene Duff for providing the data access to the cognitive task fMRI data and Samuel Harrison for pre-processing the task data. Furthermore, we are thankful to Ryan Timms and Evan Roberts for the helpful discussions.

MWW's research is supported by the NIHR Oxford Health Biomedical Research Centre, the Wellcome Trust (106183/Z/14/Z, 215573/Z/19/Z), the New Therapeutics in Alzheimer's Diseases (NTAD) study supported by UK MRC and the Dementia Platform UK (RG94383/RG89702) and the EU-project euSNN (MSCA-ITN H2020-860563). The Wellcome Centre for Integrative Neuroimaging is supported by core funding from the Wellcome Trust (203139/Z/16/Z).

UP's research is funded by an MRC Mental Health Data Pathfinder award (PI Clare Mackay) MC/PC/17215. SMS's research is supported by the XMI2: Wellcome Trust Collaborative Award 215573/Z/19/Z. DV's research is supported by Novo Nordisk Foundation Emerging Investigator Fellowship (NNF19OC-0054895) and ERC Starting Grant (ERC-StG-2019-850404).

## Appendix A. Dynamic structural equation modelling (SEM) perspective

We can gain insight into the multi-dynamic model being proposed, by considering a dynamic version of structural equation modelling (SEM) Penny et al. (2004):

$$Y_t = A_t Y_t + H \mu_t + e_t \quad (17)$$

where  $A_t$  is an  $N \times N$  dimensional matrix containing the connectivity information, where  $A_{ij}$  denotes a connection from region  $j$  to region  $i$  and  $A_{ii} = 0$  for all regions (in SEM,  $A_t$  is referred to as the path coefficient matrix,  $M$ ),  $\mu_t$  is an  $Q \times 1$  vector of  $Q$  structured, external inputs that are independent such that  $\mu_t \mu_t^T = I$ , and  $H$  is an  $N \times Q$  dimensional mixing matrix describing how much of each external input is contributing to each region. Finally,  $e_t$  is a random variable modelling the noise in  $Y_t$  using a Gaussian distribution with zero mean and an  $N \times N$  diagonal covariance matrix, i.e.,  $e_t \sim \mathcal{N}(0, \Sigma)$ . Note that the idea is that  $\mu_t$  would be identifiable through its non-Gaussian, spatio-temporal structure, which is what makes it distinct to  $e_t$ . Note that typically in SEM, only the noise,  $e_t$ , is modelled and an external input,  $\mu_t$ , is not (i.e.  $\mu_t = 0$ ). However, other variants of SEM have previously considered modelling structured, external inputs Shimizu et al. (2006).

It is straightforward<sup>6</sup> to then show that  $Y_t$  is Gaussian distributed, with a time-varying mean and covariance with the same form as in Eqs. (3), (4), where

$$m_t = (1 - A_t)^{-1} H \mu_t \quad (18)$$

$$F_t = (1 - A_t)^{-1} (\Sigma + H H^T) ((1 - A_t)^{-1})^T \quad (19)$$

In other words, if we assume that in the dynamic SEM the external inputs,  $\mu_t$ , and the dynamic connectivity,  $A_t$ , fluctuate distinctly from each other; then correspondingly  $m_t$  and  $F_t$  can also fluctuate distinctly from each other (see Eq. (18), (19)) also need to fluctuate distinctly from each other. This is what we are assuming in the *multi-dynamic* approach.

## Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.media.2022.102366.

<sup>6</sup> Such investigations to establish the relationship between SEM and FC has been previously conducted in detailed by Marrelec and Benali (2009).

## References

- Abrol, A., Rashid, B., Rachakonda, S., Damaraju, E., Calhoun, V.D., 2017. Schizophrenia shows disrupted links between brain volume and dynamic functional connectivity. *Front. Neurosci.* 11, 624.
- 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. *Cereb. Cortex* 24 (3), 663–676.
- Anand, A., Li, Y., Wang, Y., Wu, J., Gao, S., Bukhari, L., Mathews, V.P., Kainin, A., Lowe, M.J., 2005. Activity and connectivity of brain mood regulating circuit in depression: a functional magnetic resonance study. *Biol. Psychiatry* 57 (10), 1079–1088.
- Baker, A.P., Brookes, M.J., Rezek, I.A., Smith, S.M., Behrens, T., Smith, P.J.P., Woolrich, M., 2014. Fast transient networks in spontaneous human brain activity. *Elife* 3, e01867.
- 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.
- 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. *Magn. Reson. Med.* 34 (4), 537–541.
- Cai, J., Liu, A., Mi, T., Garg, S., Trappe, W., McKeown, M.J., Wang, Z.J., 2018. Dynamic graph theoretical analysis of functional connectivity in Parkinson's disease: the importance of fiedler value. *IEEE J. Biomed. Health Inform.* 23 (4), 1720–1729.
- Charquero-Ballester, M., Kleim, B., Vidaurre, D., Ruff, C., Stark, E., Tuulari, J.J., McManners, H., Bar-Haim, Y., Bouquillon, L., Moseley, A., et al., 2020. Effective psychological treatment for PTSD changes the dynamics of specific large-scale brain networks. *BioRxiv*.
- Chen, J., Sun, D., Shi, Y., Jin, W., Wang, Y., Xi, Q., Ren, C., 2018. Alterations of static functional connectivity and dynamic functional connectivity in motor execution regions after stroke. *Neurosci. Lett.* 686, 112–121.
- De Luca, M., Beckmann, C.F., De Stefano, N., Matthews, P.M., Smith, S.M., 2006. fMRI resting state networks define distinct modes of long-distance interactions in the human brain. *Neuroimage* 29 (4), 1359–1367.
- De Pasquale, F., Della Penna, S., Snyder, A.Z., Lewis, C., Mantini, D., Marzetti, L., Belardinelli, P., Ciancetta, L., Pizzella, V., Romani, G.L., et al., 2010. Temporal dynamics of spontaneous meg activity in brain networks. *Proc. Natl. Acad. Sci.* 107 (13), 6040–6045.
- De Pasquale, F., Della Penna, S., Snyder, A.Z., Marzetti, L., Pizzella, V., Romani, G.L., Corbetta, M., 2012. A cortical core for dynamic integration of functional networks in the resting human brain. *Neuron* 74 (4), 753–764.
- Duff, E.P., Makin, T., Cottaar, M., Smith, S.M., Woolrich, M.W., 2018. Disambiguating brain functional connectivity. *Neuroimage* 173, 540–550.
- d'Ambrosio, A., Valsasina, P., Gallo, A., De Stefano, N., Pareto, D., Barkhof, F., Ciccarelli, O., Enzinger, C., Tedeschi, G., Stromillo, M.L., et al., 2020. Reduced dynamics of functional connectivity and cognitive impairment in multiple sclerosis. *Mult. Scler. J.* 26 (4), 476–488.
- Fox, M.D., Raichle, M.E., 2007. Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nat. Rev. Neurosci.* 8 (9), 700–711.
- Fox, M.D., Snyder, A.Z., Vincent, J.L., Corbetta, M., Van Essen, D.C., Raichle, M.E., 2005. The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proc. Natl. Acad. Sci.* 102 (27), 9673–9678.
- Friston, K., Mattout, J., Trujillo-Barreto, N., Ashburner, J., Penny, W., 2007. Variational free energy and the laplace approximation. *Neuroimage* 34 (1), 220–234.
- Gers, F.A., Schmidhuber, J., Cummins, F., 2000. Learning to forget: continual prediction with LSTM. *Neural Comput.* 12 (10), 2451–2471.
- Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial networks. *arXiv preprint arXiv:1406.2661*.
- Graves, A., Schmidhuber, J., 2005. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Netw.* 18 (5–6), 602–610.
- Greicius, M.D., Srivastava, G., Reiss, A.L., Menon, V., 2004. Default-mode network activity distinguishes Alzheimer's disease from healthy aging: evidence from functional MRI. *Proc. Natl. Acad. Sci.* 101 (13), 4637–4642.
- Guo, X., Duan, X., Suckling, J., Chen, H., Liao, W., Cui, Q., Chen, H., 2019. Partially impaired functional connectivity states between right anterior insula and default mode network in autism spectrum disorder. *Hum. Brain Mapp.* 40 (4), 1264–1275.
- Hacker, C.D., Perlmuter, J.S., Criswell, S.R., Ances, B.M., Snyder, A.Z., 2012. Resting state functional connectivity of the striatum in Parkinson's disease. *Brain* 135 (12), 3699–3711.
- Hahn, A., Stein, P., Windischberger, C., Weissenbacher, A., Spindelegger, C., Moser, E., Kasper, S., Lanzenberger, R., 2011. Reduced resting-state functional connectivity between amygdala and orbitofrontal cortex in social anxiety disorder. *Neuroimage* 56 (3), 881–889.
- 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.
- 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.
- Higgins, C., Liu, Y., Vidaurre, D., Kurth-Nelson, Z., Dolan, R., Behrens, T., Woolrich, M., 2021. Replay bursts in humans coincide with activation of the default mode and parietal alpha networks. *Neuron* 109 (5), 882–893.

- Higgins, I., Matthey, L., Pal, A., Burgess, C., Glorot, X., Botvinick, M., Mohamed, S., Lerchner, A., 2016. beta-VAE: learning basic visual concepts with a constrained variational framework.
- Hindriks, R., Adhikari, M.H., Murayama, Y., Ganzetti, M., Mantini, D., Logothetis, N.K., Deco, G., 2016. Can sliding-window correlations reveal dynamic functional connectivity in resting-state fMRI? *Neuroimage* 127, 242–256.
- Huang, G.-H., 2014. Model Identifiability. *Wiley StatsRef: Statistics Reference Online*, pp. 1–4.
- Huang, M., Zhou, F., Wu, L., Wang, B., Guo, L., Zhao, Y., Wan, H., Li, F., Zeng, X., Gong, H., 2019. White matter lesion loads associated with dynamic functional connectivity within attention network in patients with relapsing-remitting multiple sclerosis. *J. Clin. Neurosci.* 65, 59–65.
- Hutchison, R.M., Womelsdorf, T., Allen, E.A., Bandettini, P.A., Calhoun, V.D., Corbetta, M., Della Penna, S., Duyn, J.H., Glover, G.H., Gonzalez-Castillo, J., et al., 2013. Dynamic functional connectivity: promise, issues, and interpretations. *Neuroimage* 80, 360–378.
- Kieliba, P., Madugula, S., Filippini, N., Duff, E.P., Makin, T.R., 2019. Large-scale intrinsic connectivity is consistent across varying task demands. *PLoS ONE* 14 (4), e0213861.
- Kingma, D. P., Welling, M., 2013. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*.
- Klugah-Brown, B., Luo, C., He, H., Jiang, S., Armah, G.K., Wu, Y., Li, J., Yin, W., Yao, D., 2019. Altered dynamic functional network connectivity in frontal lobe epilepsy. *Brain Topogr.* 32 (3), 394–404.
- Laumann, T.O., Snyder, A.Z., Mitra, A., Gordon, E.M., Gratton, C., Adeyemo, B., Gilmore, A.W., Nelson, S.M., Berg, J.J., Greene, D.J., et al., 2017. On the stability of bold fMRI correlations. *Cereb. Cortex* 27 (10), 4719–4732.
- Leonardi, N., Van De Ville, D., 2015. On spurious and real fluctuations of dynamic functional connectivity during rest. *Neuroimage* 104, 430–436.
- Liegeois, R., Laumann, T.O., Snyder, A.Z., Zhou, J., Yeo, B.T., 2017. Interpreting temporal fluctuations in resting-state functional connectivity MRI. *Neuroimage* 163, 437–455.
- Liégeois, R., Li, J., Kong, R., Orban, C., Van De Ville, D., Ge, T., Sabuncu, M.R., Yeo, B.T., 2019. Resting brain dynamics at different timescales capture distinct aspects of human behavior. *Nat. Commun.* 10 (1), 1–9.
- Liégeois, R., Ziegler, E., Phillips, C., Geurts, P., Gómez, F., Bahri, M.A., Yeo, B.T., Soddu, A., Vanhaudenhuyse, A., Laureys, S., et al., 2016. Cerebral functional connectivity periodically (de) synchronizes with anatomical constraints. *Brain Struct. Funct.* 221 (6), 2985–2997.
- Lindquist, M.A., Xu, Y., Nebel, M.B., Caffo, B.S., 2014. Evaluating dynamic bivariate correlations in resting-state fMRI: a comparison study and a new approach. *Neuroimage* 101, 531–546.
- Lowd, D., Meek, C., 2005. Adversarial learning. In: *Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining*, pp. 641–647.
- Lurie, D.J., Kessler, D., Bassett, D.S., Betzel, R.F., Breakspear, M., Kheilholz, S., Kucyi, A., Liégeois, R., Lindquist, M.A., McIntosh, A.R., et al., 2020. Questions and controversies in the study of time-varying functional connectivity in resting fMRI. *Netw. Neurosci.* 4 (1), 30–69.
- Lynall, M.-E., Bassett, D.S., Kerwin, R., McKenna, P.J., Kitzbichler, M., Muller, U., Bullmore, E., 2010. Functional connectivity and brain networks in schizophrenia. *J. Neurosci.* 30 (28), 9477–9487.
- Makhzani, A., Shlens, J., Jaitly, N., Goodfellow, I., Frey, B., 2015. Adversarial autoencoders. *arXiv preprint arXiv:1511.05644*.
- Marrelec, G., Benali, H., 2009. A theoretical investigation of the relationship between structural equation modeling and partial correlation in functional MRI effective connectivity. *Comput. Intell. Neurosci.* 2009.
- Mescheder, L., Nowozin, S., Geiger, A., 2017. Adversarial variational bayes: unifying variational autoencoders and generative adversarial networks. In: *International Conference on Machine Learning*, PMLR, pp. 2391–2400.
- Miller, K.L., Alfaro-Almagro, F., Bangarter, N.K., Thomas, D.L., Yacoub, E., Xu, J., Bartsch, A.J., Jbabdi, S., Sotiropoulos, S.N., Andersson, J.L., et al., 2016. Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nat. Neurosci.* 19 (11), 1523–1536.
- Miller, R.L., Erhardt, E.B., Agcaoglu, O., Allen, E.A., Michael, A.M., Turner, J.A., Bustillo, J., Ford, J.M., Mathalon, D.H., Van Erp, T.G., et al., 2015. Multidimensional frequency domain analysis of full-volume fMRI reveals significant effects of age, gender, and mental illness on the spatiotemporal organization of resting-state brain activity. *Front. Neurosci.* 9, 203.
- O'Reilly, J.X., Woolrich, M.W., Behrens, T.E., Smith, S.M., Johansen-Berg, H., 2012. Tools of the trade: psychophysiological interactions and functional connectivity. *Soc. Cogn. Affect. Neurosci.* 7 (5), 604–609.
- Parr, T., Rees, G., Friston, K.J., 2018. Computational neuropsychology and Bayesian inference. *Front. Hum. Neurosci.* 12, 61.
- Penny, W.D., Stephan, K.E., Mechelli, A., Friston, K.J., 2004. Modelling functional integration: a comparison of structural equation and dynamic causal models. *Neuroimage* 23, S264–S274.
- Pervaiz, U., Vidaurre, D., Woolrich, M.W., Smith, S.M., 2020. Optimising network modelling methods for fMRI. *Neuroimage* 211, 116604.
- 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. *Front. Neurosci.* 12, 603.
- Rogers, B.P., Morgan, V.L., Newton, A.T., Gore, J.C., 2007. Assessing functional connectivity in the human brain by fMRI. *Magn. Reson. Imaging* 25 (10), 1347–1357.
- Sakoğlu, Ü., Pearlson, G.D., Kiehl, K.A., Wang, Y.M., Michael, A.M., Calhoun, V.D., 2010. A method for evaluating dynamic functional network connectivity and task-modulation: application to schizophrenia. *Magn. Reson. Mater. Phys. Biol. Med.* 23 (5–6), 351–366.
- Shimizu, S., Hoyer, P.O., Hyvärinen, A., Kerminen, A., Jordan, M., 2006. A linear non-gaussian acyclic model for causal discovery. *J. Mach. Learn. Res.* 7 (10).
- Smith, S.M., 2012. The future of fMRI connectivity. *Neuroimage* 62 (2), 1257–1266.
- Smith, S.M., Beckmann, C.F., Andersson, J., Auerbach, E.J., Bijsterbosch, J., Douaud, G., Duff, E., Feinberg, D.A., Griffanti, L., Harms, M.P., et al., 2013. Resting-state fMRI in the human connectome project. *Neuroimage* 80, 144–168.
- Smith, S.M., Hyvärinen, A., Varoquaux, G., Miller, K.L., Beckmann, C.F., 2014. Group-PCA for very large fMRI datasets. *Neuroimage* 101, 738–749.
- Tagliazucchi, E., Von Wegner, F., Morzelewski, A., Brodbeck, V., Laufs, H., 2012. Dynamic bold functional connectivity in humans and its electrophysiological correlates. *Front. Hum. Neurosci.* 6, 339.
- Tavor, I., Jones, O.P., Mars, R.B., Smith, S., Behrens, T., Jbabdi, S., 2016. Task-free MRI predicts individual differences in brain activity during task performance. *Science* 352 (6282), 216–220.
- Van Essen, D.C., Smith, S.M., Barch, D.M., Behrens, T.E., Yacoub, E., Ugurbil, K., Consortium, W.-M.H., et al., 2013. The WU-Minn human connectome project: an overview. *Neuroimage* 80, 62–79.
- Van Schependom, J., Vidaurre, D., Costers, L., Sjøgård, M., D'hooghe, M.B., D'haeseleer, M., Wens, V., De Tiège, X., Goldman, S., Woolrich, M., et al., 2019. Altered transient brain dynamics in multiple sclerosis: treatment or pathology? *Hum. Brain Mapp.* 40 (16), 4789–4800.
- Vidaurre, D., Abeysuriya, 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.
- Vidaurre, D., Bielza, C., Larranaga, P., 2013. A survey of l1 regression. *Int. Stat. Rev.* 81 (3), 361–387.
- Vidaurre, D., Hunt, L.T., Quinn, A.J., Hunt, B.A., Brookes, M.J., Nobre, A.C., Woolrich, M.W., 2018. Spontaneous cortical activity transiently organises into frequency specific phase-coupling networks. *Nat. Commun.* 9 (1), 1–13.
- Vidaurre, D., Llera, A., Smith, S.M., Woolrich, M.W., 2021. Behavioural relevance of spontaneous, transient brain network interactions in fMRI. *Neuroimage* 229, 117713.
- Vidaurre, D., Smith, S.M., Woolrich, M.W., 2017. Brain network dynamics are hierarchically organized in time. *Proc. Natl. Acad. Sci.* 114 (48), 12827–12832.
- Zhi, D., Calhoun, V.D., Lv, L., Ma, X., Ke, Q., Fu, Z., Du, Y., Yang, Y., Yang, X., Pan, M., et al., 2018. Aberrant dynamic functional network connectivity and graph properties in major depressive disorder. *Front. Psychiatry* 9, 339.
- Zou, H., Hastie, T., 2005. Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B* 67 (2), 301–320.