Uncovering dynamic human brain phase coherence networks

Anders S. Olsen[1*], Anders Brammer[1], Patrick M. Fisher[2,3], Morten Mørup[1*]

[1]Department of Applied Mathematics and Computer Science, Technical University of Denmark, Kgs. Lyngby, Denmark

[2]Neurobiology Research Unit, Copenhagen University Hospital Rigshospitalet, Copenhagen, Denmark

[3]Department of Drug Design and Pharmacology, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark

[*]Corresponding authors: ansol@dtu.dk, mmor@dtu.dk

Keywords: Complex angular central Gaussian, Dynamic functional connectivity, Directional statistics, Probabilistic mixture models, Hidden Markov models, Phase coherence, Multivariate synchronization.

ABSTRACT

Complex behavioral and cognitive functions emerge from coordinated dynamic communication between often disparate brain regions, implying a systems-level arrangement of underlying neural signals. This study introduces multivariate mixture and hidden Markov modeling approaches to analyze phase coherence networks in fMRI data, capturing macroscale dynamic functional synchronization in the human brain. We show that 1) phase coherence is inherently a complex-valued phenomenon and should be analyzed as such, 2) statistical models for assessing complex-valued phase coherence, particularly the complex angular central gaussian (ACG) distribution, outperform amplitude and phase-amplitude methods on synthetic and task fMRI data as well as existing phase coherence modeling approaches (including LEiDA), 3) models of phase coherence should account for the inherent anisotropy of the brain's interconnections by including a covariance matrix in each mixture component. We provide a

Python-based toolbox ("Phase Coherence Mixture Modeling" (PCMM): github.com/anders-s-olsen/PCMM) to facilitate implementation of phase coherence models.

INTRODUCTION

Synchronization is a fundamental concept in science, observed across diverse fields from coherent light in lasers, anti-coherence in acoustic noise suppression, and the circadian rhythm aligning with the dark-light cycle of the environment. In neuroscience, synchronization is crucial across multiple scales: Neuron-to-neuron communication often occurs through synchronized bursts of firing (Coombes & Bressloff, 2005; Maeda et al., 1995) and synchronized mesoscale activity within cortical neuronal columns generates electrical signals that can be detected on the scalp (Kirschstein & Köhling, 2009). At the macroscale, synchronization underpins functional connectivity across brain areas, providing a framework for understanding cognitive functions and behaviors (Bola & Sabel, 2015; Uhlhaas et al., 2009; Womelsdorf et al., 2007). As such, accurate models of phase coherence, the quantification of synchronization, in multi-channel brain data are required to capture the complexity of brain-wide connectivity patterns, key to understanding the orchestrated dynamics within and across neural networks.

Traditionally, connectivity between regional neuroimaging signals such as those from blood oxygen level dependent functional magnetic resonance imaging (BOLD fMRI) has been measured using the Pearson correlation coefficient (Biswal et al., 1995), which captures the linear relationship between signals (Bastos & Schoffelen, 2016; Wang et al., 2014). Estimating time-varying correlation coefficients requires temporal windowing of input signals, decreasing stationarity and leading to challenges in determining, e.g., the optimal window length, function, and stride (Allen et al., 2014; Leonardi & Van De Ville, 2015; Nielsen et al., 2017; Preti et al., 2017), and has been suggested to be an unstable approach to identifying dynamic functional connectivity (Choe et al., 2017; Hindriks et al., 2016; Laumann et al., 2017; Liégeois et al., 2017), see (Laumann et al., 2024) for a recent review. Directly modeling the time-series using gaussian mixture models (GMM) or hidden Markov models (HMMs) offer window-free alternatives (Vidaurre et al., 2017), however, intensity-based time-series methods are vulnerable to amplitude-related noise such as head motion (Chen et al., 2017).

Instead, assessing phase coherence (a.k.a. phase coupling) between brain signals, focusing exclusively on co-oscillation and discarding (co-)amplitude information, offers a frame-wise connectivity proxy free from area-specific amplitude-related noise uninformative of connectivity (Glerean et al., 2012). Phase

estimation involves converting signals to the complex domain, typically via the Hilbert transform, which translates temporal evolution of a signal $x_j(t)$ into a revolution around the unit circle (Fig. 1). The phase $\theta_j(t)$ is then a circular quantity with a period of 2π, and co-oscillations between regions are reflected by differences in their phases (phase coherence), which may be assessed at each data sample. Most previous fMRI-studies involving phase coherence have focused exclusively on *cosine* phase coherence, i.e., $\cos\left(\theta_j(t) - \theta_{j'}(t)\right)$ (Cabral et al., 2017; Glerean et al., 2012; Honari et al., 2021; Kringelbach et al., 2020; Zarghami et al., 2020) however, this approach discards the imaginary dimension (the sine), which holds potentially valuable information about phase alignment. In contrast, volume conduction in electroencephalography (EEG) data causes most signals to be coherent, and consequently, the sine is favored (Nolte et al., 2004). Furthermore, some studies have focused exclusively on the leading eigenvector of the instantaneous cosine phase coherence matrix, particularly through Leading Eigenvector Dynamics Analysis (LEiDA) (Cabral et al., 2017), further explicitly discarding a portion of the data variance within the cosine projection of the phase coherence. In this paper, we show that interregional phase coherence is very easily modeled when considering the full rectangular form $e^{i(\theta_j(t)-\theta_{j'}(t))} = \cos(\theta_j(t) - \theta_{j'}(t)) + i\sin(\theta_j(t) - \theta_{j'}(t))$. Consequently, the complex dynamic phase coherence map for each volume t is rank-1 $e^{i\boldsymbol{\theta}_t}\left(e^{i\boldsymbol{\theta}_t}\right)^{\mathrm{H}}$, where $(\cdot)^{\mathrm{H}}$ is the complex conjugate transpose, rendering modeling a pool of instantaneous phase vectors $\left\{\mathbf{u}_t = e^{i\boldsymbol{\theta}_t}\right\}_{t=1}^{N}$ possible without the need for lossy dimensionality reduction.

Such instantaneous complex-valued phase vectors can be effectively modeled using the complex angular central Gaussian (ACG) distribution (Kent, 1997; Tyler, 1987). In this paper we contrast phase modeling with traditional amplitude modeling via the Gaussian distribution as well as phase-amplitude modeling via the complex Gaussian distribution (Fig. 1). We introduce probabilistic mixtures and HMMs enabling modeling dynamic functional connectivity via all three data representations (see Supplementary Figure S1 for an overview of the phase coherence clustering framework). All distributions include a covariance matrix, the rank of which we show is crucial for extracting dynamics in data. For completeness, we also apply mixture modeling tools for the cosine phase coherence using appropriate statistical distributions, as well as corresponding non-probabilistic K-means (see Supplementary Table S1 for an overview). All models have been implemented in our Python toolbox "Phase Coherence Mixture Modeling" (PCMM) (github.com/anders-s-olsen/PCMM).

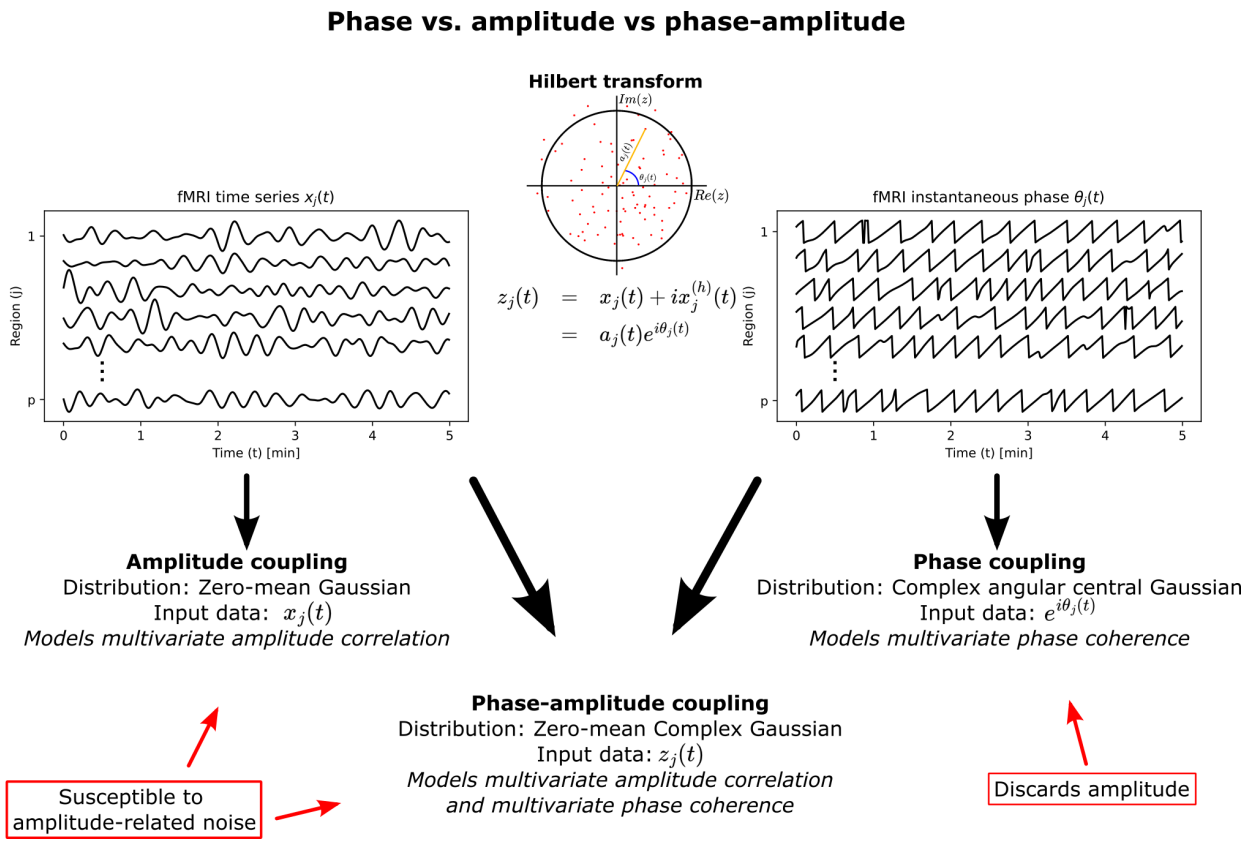

Figure 1: The Hilbert transform is used to create the analytic signal $z_j(t)$, which decomposes a time-series $x_j(t)$ into phase $\theta_j(t)$ and amplitude $a_j(t)$. While time-series data can be modeled using the Gaussian distribution, the analytic signal can be modeled using the complex Gaussian distribution. Discarding amplitude information results in unit-norm phase vectors, which can be modeled using the complex angular central Gaussian (ACG) distribution.

RESULTS

Using synthetic data, we first establish the advantage of modeling phase coherence in its entirety, with both the cosine and sine, in order to adequately recover data with systematic phase coherence fingerprints. We subsequently evaluate our approaches on both task-based and resting-state fMRI data from 255 unrelated subjects from the Human Connectome Project (Van Essen et al., 2013) downsampled to 100 cortical regions from the Schaefer atlas (Schaefer et al., 2018) augmented with 16 subcortical regions from the Tian atlas (Tian et al., 2020), showcasing the need for detailed modeling of networks with

covariance matrices, as well as how modeling phase-coherence networks improves the identification of states that are task-specific vis-à-vis conventional models.

The complex domain is required for modeling phase coherence in its entirety

Phase coherence methods convert signals to the complex domain via the Hilbert transform (Fig. 1), wherein the amplitude is discarded and only the instantaneous phase vectors $e^{i\boldsymbol{\theta}_t}$ retained, where $\boldsymbol{\theta}_t$ are the circular multivariate phases for time t and i is the imaginary unit. Phase coherence is then examined by fitting a model to a pool of phase vectors, explaining consistent phase differences between two or more signals. As illustrated in Fig. 2, modeling only the real part (cosine) of the complex phase vectors as suggested in, among others, (Cabral et al., 2017; Glerean et al., 2012), results in phase shifts of θ and $2\pi - \theta$ becoming indistinguishable, and consistent phase shifts $\pi/2$ or $3\pi/2$ cannot be differentiated from unrelated signals. Similarly, modeling the imaginary part of coherence (Nolte et al., 2004) results in signals that are in phase becoming indistinguishable from signals in antiphase. Consequently, the full complex representation is needed to account for all possible types of phase coherence and previous methods focusing on, e.g., only cosine or sine coherence are inherently limited in their representational value.

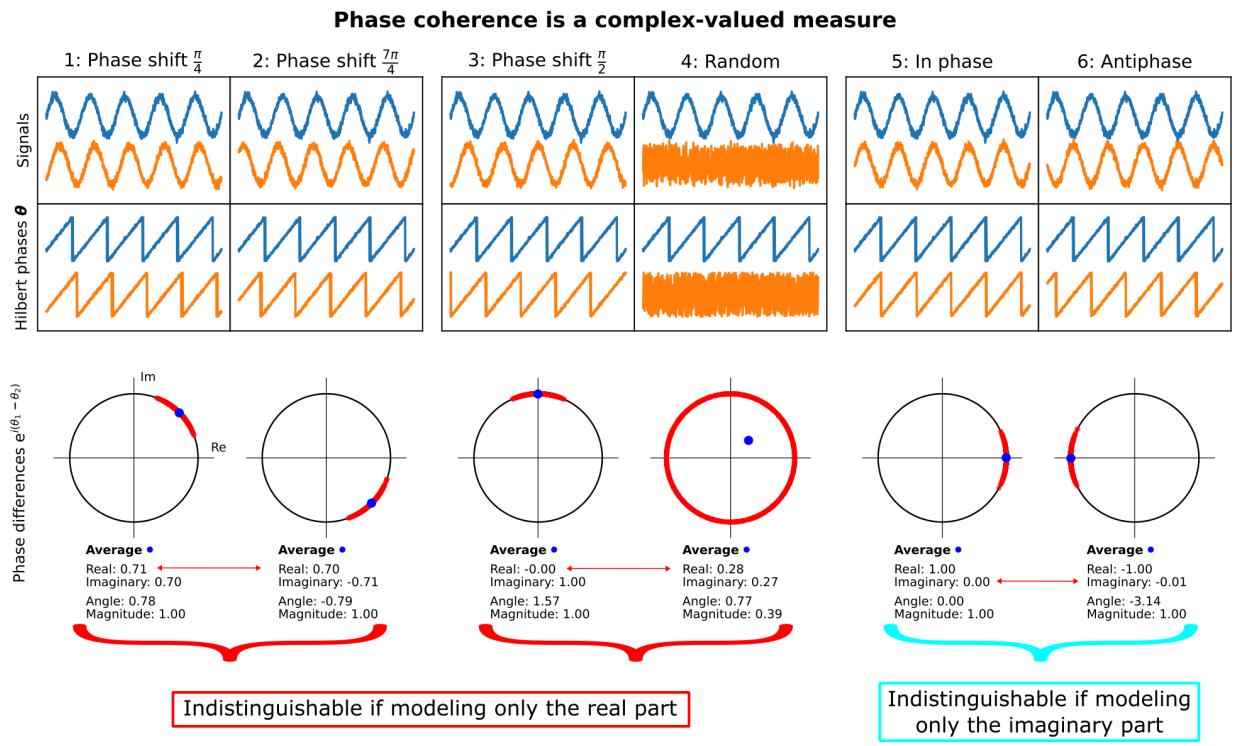

Figure 2: Modeling phase coherence in its entirety. Two oscillatory signals can be compared by examining their phases, which indicate the position of each sample within the oscillation cycle. Modeling only cosine phase coherence restricts modeling to the real part, making it impossible to distinguish phase shifts θ and $2\pi - \theta$ (cases 1 and 2) or differentiate unrelated signals from signals with a consistent phase shift of $\pi/2$ or $2\pi - \pi/2$ (cases 3 and 4). Similarly, modeling only the imaginary part fails to distinguish in-phase from antiphasic signals (cases 5 and 6). These issues are resolved when modeling both cosine and sine in a combined complex-valued representation.

On a synthetic mixture of $K = 2$ components with noisy, $p = 3$-dimensional phase vectors with consistent phase shifts, models for the complex-valued phase coherence, including the complex ACG model, outperformed methods that only evaluate cosine phase coherence (see Supplementary Fig. S2).

Rank-1 component covariances are insufficient for modeling anisotropic data

Multivariate brain data is anisotropic over time; i.e., two adjacent brain areas are likely to have similar neuroimaging time-series while a third, unrelated brain area may have a time-series unrelated to the first two (c.f. case 4 in Fig. 2). In this case, the distribution of such three-dimensional data tends to a

low-rank manifold, e.g., a hyperplane in the Euclidean vector space or a great circle for normalized data (Fig. 3A). Importantly, the rank of the subspace can fall anywhere between 1 and the data's dimensionality.

To account for anisotropy in data, a statistical distribution with a covariance matrix parameter $\mathbf{\Psi}$ may be reparameterized using a low-rank matrix plus the identity matrix, i.e., the so-called low-rank-plus-diagonal reparameterization: $\mathbf{\Psi} \approx \mathbf{Z} = \mathbf{M}\mathbf{M}^{\mathsf{H}} + \mathbf{I}$. Here, $\mathbf{M} \in \mathbb{C}^{p \times r}$ is a rank-r matrix, $\mathbf{I}$ is the identity matrix, and $(\cdot)^{\mathsf{H}}$ is the complex conjugate transpose, reducing to the transpose for real $\mathbf{M}$ (Olsen et al., 2023). This allows explicitly controlling the amount of information captured by each component in a mixture model. For some distributions, such as the Gaussian distribution, a weight-parameter on the identity matrix is also required (see Methods).

In the two-component synthetic data in Fig. 3A, an equally two-component mixture model with reparameterized ACG (real-valued for visualization purposes) distributions as mixture components captured the data to differing degrees: Rank-1 components are inherently isotropic, for which potentially infinitely many mixture components are needed to account for the anisotropy of the data. Instead, two rank-2 mixture components sufficiently explained the data with additional components or increased component rank not providing any gain in predictive likelihood. Those additional parameters may be viewed as redundant and should be removed in keeping with the law of parsimony. Of note, many K-means models and matrix decompositions also have rank-1 components and should be used with caution if the data is anisotropic.

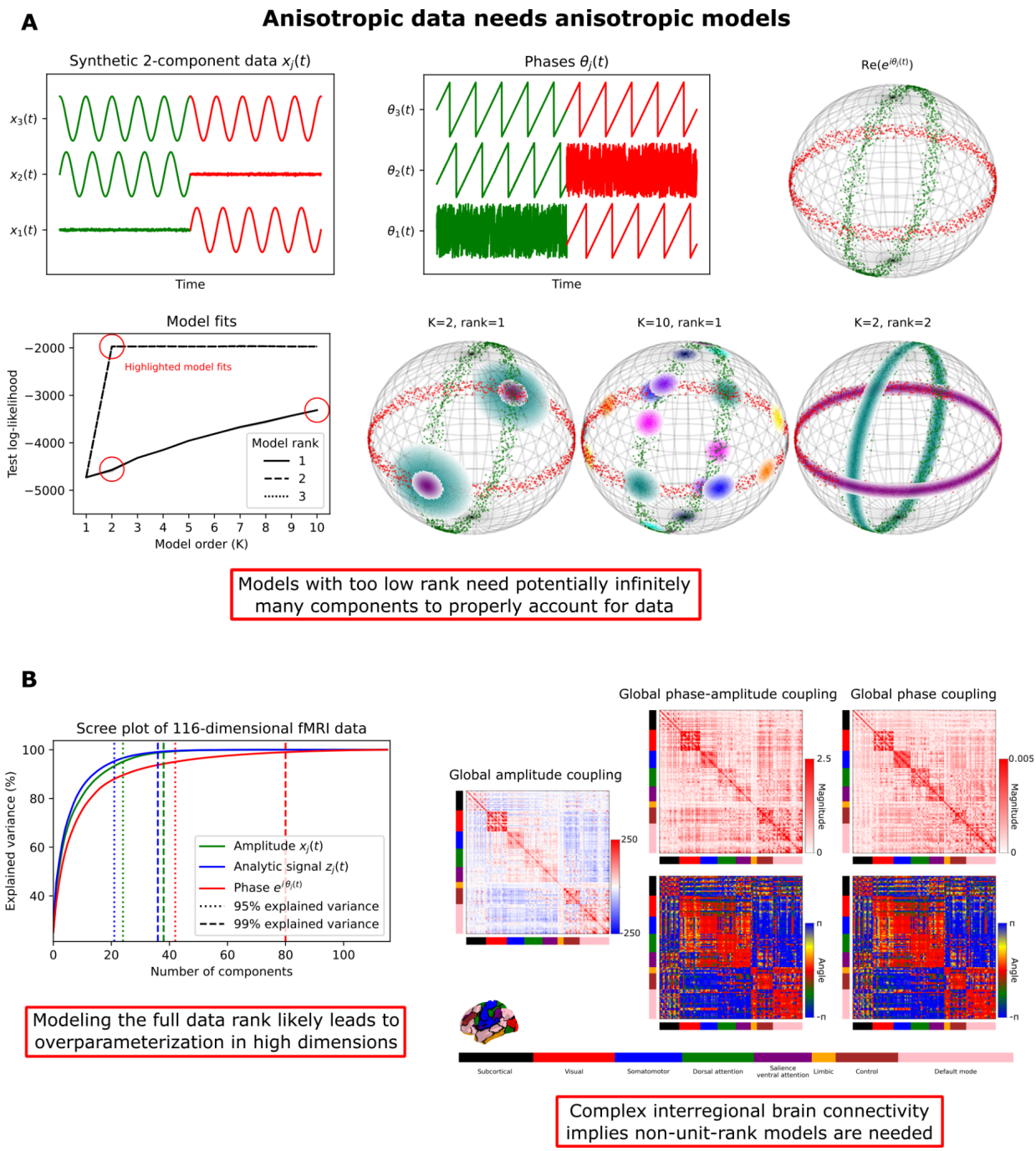

Figure 3: Non-unit rank statistical distributions are needed to model anisotropy present in functional neuroimaging data. ***A****): Real synthetic anisotropic phase vectors in two components. The (non-complex) ACG mixture model with a rank-1 covariance matrix for each component fails to account for the data anisotropy and instead fits isotropic approximations along the data manifold. With rank-2 or rank-3 covariance, the model adequately covers the spherical subspace along which the data is distributed. Shaded areas in test log-likelihood correspond to the standard error over 5 model initializations.* ***B****): Functional neuroimaging data is naturally grouped into networks, which are inadequately described by*

rank-1 approximations but likely overparameterized by full-rank models. Here, the solution includes modeling a certain predefined rank in between 1 *and* p*. Coupling matrices shown are an average across 255 subjects from the Human Connectome Project in resting state.*

Intercorrelations in fMRI data are seen by the naturally occurring network groupings of brain areas requiring non-unit rank covariances to model (Fig. 3B). Contrarily, a full covariance parameter $\mathbf{\Psi}$ can become singular leading to exploding normalization constants or model excessive numbers of parameters without explanatory gain.

On real phase-controlled resting-state fMRI data from one subject, where the phase in the Fourier-domain had been manipulated into $K = 5$ segments, models for complex phase coupling as well as for rank-2 cosine phase coupling outperformed all other models, showing their superiority in detecting phase changes in data unrelated to amplitude (Supplementary Fig. S3). Importantly, model performance was observed to stagnate around $r = 10$ with $p = 116$, thus indicating that unit-rank models underparameterize the data and full-rank models are likely overparameterized. Furthermore, we observed no performance benefits of using HMM over mixture models (Supplementary Fig S4) and therefore opted to only present mixture model results for the rest of the study.

Phase coherence mixture modeling provides meaningful partitions into task-related functional states

Next, we investigated whether a mixture model could extract meaningful patterns from task fMRI data. We concatenated all seven HCP tasks: emotion, gambling, language, motor, relational, social, and working memory. Each task also consisted of subtasks or alternating task/rest periods, which we did not take into account. We concatenated the task scans for 155 subjects in a train set; the remaining 99 subjects constituted an out-of-sample test set (one subject discarded due to incomplete data). We fitted seven-component mixture models to train data in search for potential one-to-one correspondences between mixture components and true task labels present also in the test set.

The seven-component complex ACG model for complex phase coupling outperformed the Gaussian (amplitude coupling) and complex Gaussian (phase-amplitude coupling) models in terms of overlap between the true task-scan label and the test posterior probability, measured using normalized mutual information, a metric invariant to permutations of components and tasks (Fig. 4A). Furthermore, the complex Gaussian model data outperformed the real Gaussian model, suggesting that the decomposition

of amplitude data into oscillatory and magnitude constituents benefits the extraction of biologically meaningful components compared to the raw fMRI time series.

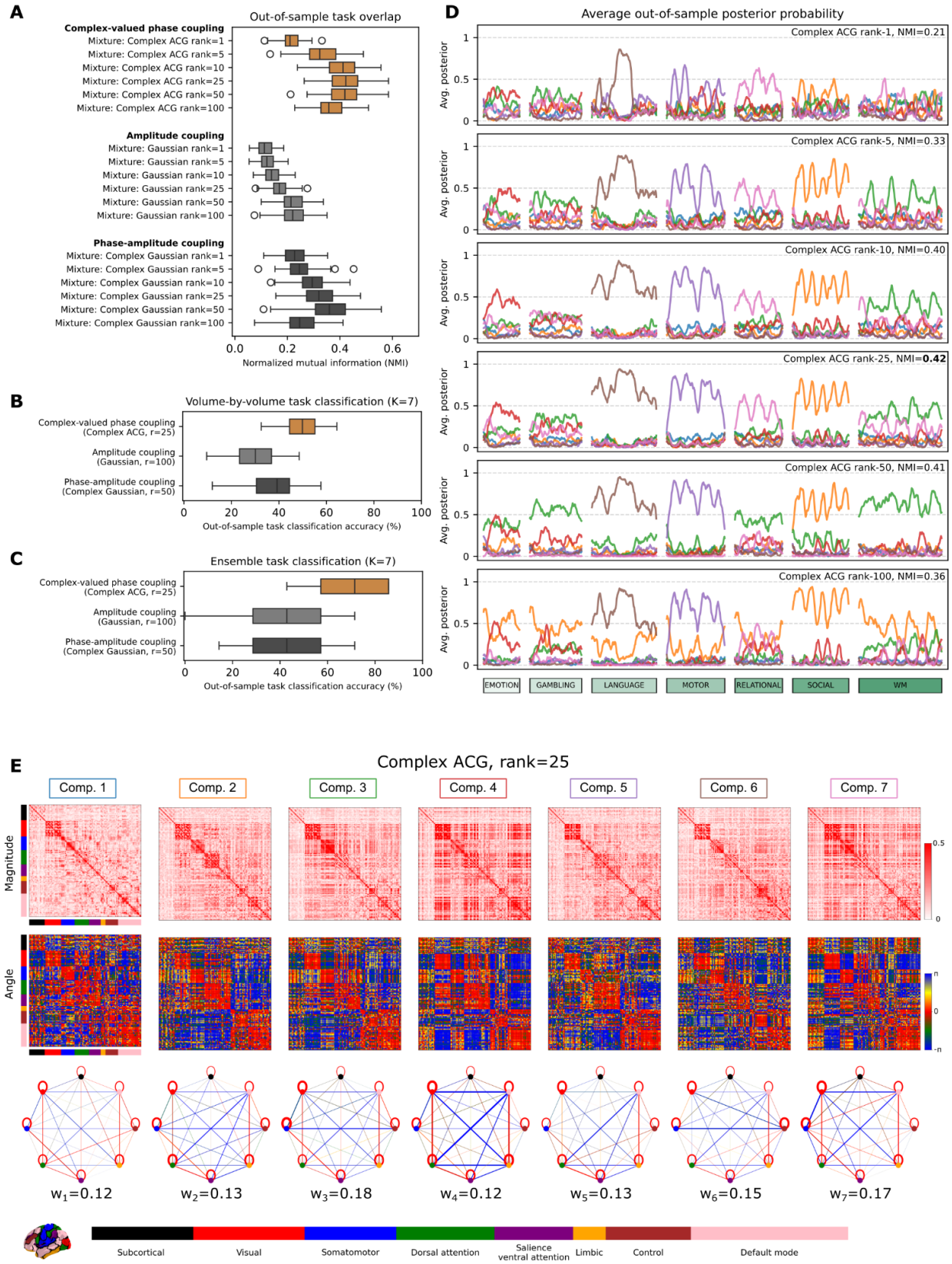
A
Out-of-sample task overlap
Complex-valued phase coupling
Mixture: Complex ACG rank=1
Mixture: Complex ACG rank=5
Mixture: Complex ACG rank=10
Mixture: Complex ACG rank=25
Mixture: Complex ACG rank=50
Mixture: Complex ACG rank=100
Amplitude coupling
Mixture: Gaussian rank=1
Mixture: Gaussian rank=5
Mixture: Gaussian rank=10
Mixture: Gaussian rank=25
Mixture: Gaussian rank=50
Mixture: Gaussian rank=100
Phase-amplitude coupling
Mixture: Complex Gaussian rank=1
Mixture: Complex Gaussian rank=5
Mixture: Complex Gaussian rank=10
Mixture: Complex Gaussian rank=25
Mixture: Complex Gaussian rank=50
Mixture: Complex Gaussian rank=100
0.0 0.2 0.4 0.6
Normalized mutual information (NMI)
B
Volume-by-volume task classification (K=7)
Complex-valued phase coupling (Complex ACG, r=25)
Amplitude coupling (Gaussian, r=100)
Phase-amplitude coupling (Complex Gaussian, r=50)
0 20 40 60 80 100
Out-of-sample task classification accuracy (%)
C
Ensemble task classification (K=7)
Complex-valued phase coupling (Complex ACG, r=25)
Amplitude coupling (Gaussian, r=100)
Phase-amplitude coupling (Complex Gaussian, r=50)
0 20 40 60 80 100
Out-of-sample task classification accuracy (%)
D
Average out-of-sample posterior probability
Complex ACG rank-1, NMI=0.21
Complex ACG rank-5, NMI=0.33
Complex ACG rank-10, NMI=0.40
Complex ACG rank-25, NMI=0.42
Complex ACG rank-50, NMI=0.41
Complex ACG rank-100, NMI=0.36
Avg. posterior
0 0.5 1
EMOTION
GAMBLING
LANGUAGE
MOTOR
RELATIONAL
SOCIAL
WM
E
Complex ACG, rank=25
Comp. 1
Comp. 2
Comp. 3
Comp. 4
Comp. 5
Comp. 6
Comp. 7
Magnitude
Angle
0.5
0
π
0
-π
w1=0.12
w2=0.13
w3=0.18
w4=0.12
w5=0.13
w6=0.15
w7=0.17
Subcortical
Visual
Somatomotor
Dorsal attention
Salience ventral attention
Limbic
Control
Default mode

Figure 4: Seven-component mixture models applied to the seven task fMRI scans from the human connectome project (HCP) concatenated. ***A****): Overlap between posterior probability and the true labels measured using normalized mutual information.* ***B****): Volume-by-volume task classification accuracy using the binarized posterior probabilities. Components were matched to tasks based on their average performance on the train set.* ***C****): Same as B), with only one predicted component per scan. In A) B), and C), variability is shown across the 99 test subjects.* ***D****) The average test posterior of each of the seven components (one color for each) of the complex ACG mixture model at different component ranks. The x-axis represents time with whitespace demarcating different task scans.* ***E****) The seven components learned using the best-performing model, i.e., the complex ACG for complex-valued phase coupling with rank-25 components. In the graph depiction, the color of edges corresponds to the average interregional "angle" and the line thickness to the average interregional magnitude.*

The complex ACG model reached maximum performance with a covariance rank of 25 with a test-NMI of 0.42, corresponding to 50% seven-class classification accuracy when evaluated at every acquired volume (Fig. 4B, see the performance for all models in Supplementary Fig S7). If instead labeling each scan only once, the average ensemble test-accuracy becomes 72%. Further increasing model rank resulted in lower task identification accuracy. If not performing global signal regression (GSR), the performance of phase coherence models decreases while that of the Gaussian and complex Gaussian models stays similar (Supplementary Fig. S8). The performance was similar between complex ACG and complex Normal without GSR, reaching an average volume-by-volume classification of 34% and 33%, respectively suggesting that GSR particularly benefits phase coupling methods.

The average test set posterior probability similarly indicated that the complex ACG model is increasingly able to discriminate tasks with higher rank, peaking at $r = 25$ (Fig. 4D). The language, motor, and social tasks were the easiest for the model to discern, while the emotion, relational, and working memory tasks were discernible to a lesser extent. The gambling task was more difficult to discern. One component was mostly in the background (blue), indicating the presence of a non-task-related background component with spurious activation. The average test posterior also indicated consistent within-task fluctuations, e.g., five peaks for the social task, which is composed of five task blocks, and three peaks for the relational task, which is composed of three blocks. The average posteriors for the Gaussian and complex Gaussian showed lower task discernibility but overall similar patterns (Supplementary Figs S5-6). The Gaussian model showed clear edge effects indicating that the time-series is particularly susceptible to amplitude-related artifacts, which both the complex Normal and

complex ACG models avoid by assigning more weight to the oscillatory rather than the amplitude constituent.

Differences between components were mostly confined to the strength of between-network (anti-)coherences, of which few were similar across all components (Fig. 4E). The non-task-related component 1 (blue) was characterized by low inter-Yeo-network coherence. Comparing to the background state, component 2 (yellow, social task) had increased DMN anti-coherences between two major groups: group 1 (visual, somatomotor, dorsal attention and salience/ventral attention) and group 2 (limbic, control, and DMN) as well as a generally higher within-network coherence in all of these networks. Component 3 (green, working memory) had higher within-network coherence in visual and somatomotor networks. Component 4 (red, emotion) was characterized by strong visual-DMN, dorsal attention-DMN, dorsal attention-limbic, and salience/ventral attention-DMN anticoherences as well as increased visual-dorsal attention, dorsal attention-salience/ventral attention, and Limbic-DMN coherences and increased within network coherence in the visual, dorsal attention, limbic, and DMN networks. Component 5 (purple, motor) was characterized by coherence within and between dorsal attention-salience/ventral attention as well as between the DMN and the two attention networks. Component 6 (brown, language) was primarily characterized by anti-coherence between visual-control and somatomotor-control networks and strong within-network coherence in the control network. Component 7 (pink, relational) was characterized by within-network coherence in visual and somatomotor and anti-coherence between visual-somatomotor, vis-limbic, visual-DMN, somatomotor-dorsal attention, and dorsal attention-DMN.

Resting-state phase coherence is better explained by higher component rank than more components

For the resting-state data we evaluated the number of mixture model components and component rank using predictive log-likelihood on the out-of-sample test set composed of 100 subjects while the models were trained on 155 subjects (Fig. 5A). Increasing the number of components altered likelihoods less than increasing component rank. When model rank tended towards the data's dimensionality ($p = 116$), fewer components were supported by the test-set resting state data. For the complex ACG mixture model, three components with rank-100 components achieved the highest predictive likelihood. Similarly, the Gaussian and complex Gaussian models peaked at $K = 3$ and $K = 4$, respectively, with rank-100 components. These results suggest that resting-state brain fMRI data is not easily temporally split into clusters that are representative across subjects and scans.

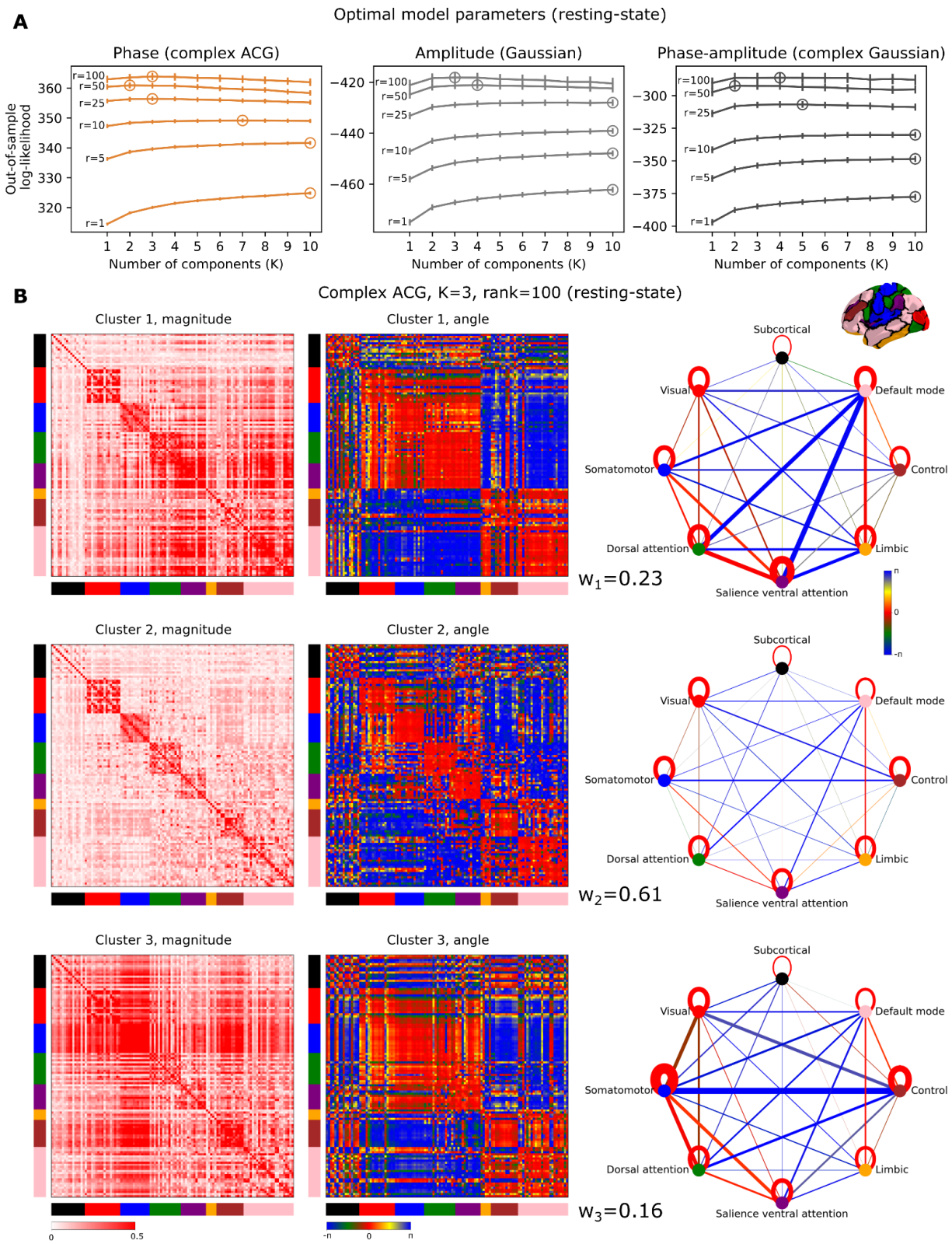

*Figure 5: Model parameter search and mixture estimate on resting-state fMRI. **A**): Component rank and model order via out-of-sample test log-likelihood (normalized by the number of samples); bars*

represent variability across subjects. ***B****): Complex ACG mixture estimate with* $K = 3$ *and component ranks* $r = 100$*. In the graph representation, average inter-network angle is represented by color, and magnitude by edge thickness.*

The second component of the $K = 3$, rank-100 complex ACG solution had the highest weight and was largely similar to the language-related sixth component in the task analysis, although with generally higher within-network coherence than between-network coherence. The first component strongly segregated the aforementioned two groups of sensorimotor/attention and limbic/control/DMN networks and also showed a higher within-network coherence in the attention and DMN networks. The third component, which was the least prevalent, was characterized by increased within-network coherence in the somatomotor and control networks, as well as increased coherence between sensorimotor/attention networks and anti-coherence between control and sensorimotor/attention networks. While the first components showed the strongest anti-coherences between DMN and other networks, the third component had the strongest anti-coherences between the control network and other networks.

Global signal regression (GSR) significantly altered the average resting-state complex global phase coherence map, revealing interregional anti-coherence in the angular representation (Supplementary Fig. S9). Without GSR, the vast majority of the brain was, on average in this data, coherent. Following GSR, the average angular global phase coherence was predominantly close to zero (in-phase) ir π (anti-phase) indicating that resting-state fMRI holds less important consistent phase shifts than the evaluated task fMRI data. Visually, a complex ACG model with rank 10 was able to fit the global average phase coherence map well, with a visually negligible effect of increasing the rank. In contrast, a unit-rank model provided a notably simplified estimate of the global coherence map, essentially dividing the cortical areas into two chunks, again suggesting that rank-1 models are too restrictive.

DISCUSSION

Here we explored and advanced multivariate probabilistic mixture modeling approaches for analyzing large-scale networks of phase coherence in their entirety, emphasizing the importance of modeling data while discarding as little potentially valuable information as possible. Based on our results, we argue that simplifying data—such as by discarding the imaginary component of phase coherence, relying on leading eigenvectors, or utilizing rank-1 components—can result in loss of cognitively relevant information. Our findings show that appropriate statistical distributions that focus on phase coherence rather than signal

amplitude better retrieve task information and provide a clean representation of brain dynamics. Importantly, our analyses were performed on unseen data from multiple subjects, indicating that the learned representations are well-represented across individuals.

Through synthetic data, we illustrated that focusing solely on cosine phase coherence may mask valuable information about phase shifts between signals, a drawback remedied by complex-valued components (Fig. 2). The estimated seven-component complex ACG model for task fMRI data also suggested valuable information in the imaginary component (green/yellow edges in Fig. 4E, "angle"). In the phase-controlled and task-fMRI experiments, appropriate models for the rank-2 cosine phase coherence maps performed slightly worse than the complex ACG for complex-valued phase coherence maps (Supplementary Fig. S7). Of note, modeling the complex rather than the rank-2 cosine phase coherence maps directly requires one less component rank. Taken together, constraining phase coherence models to the real domain risks discarding potential insights embedded in the imaginary component without obvious benefit.

Nevertheless, if the imaginary part is discarded, our analysis supports the use of the rank-2 cosine phase coherence over only its leading eigenvector (i.e., LEiDA) (Supplementary Figs. S2-S4, S7), as the latter omits data patterns associated with the trailing eigenvector. Although some studies have suggested that the trailing eigenvector captures "noise" (Ahrends & Vidaurre, 2023; Cabral et al., 2017), empirical evidence for this claim remains to be seen. In fact, recent work indicates that the trailing eigenvector may reveal localized connectivity patterns (de Alteriis et al., 2024). Our supplementary results reinforce that reducing dimensionality by modeling only the leading eigenvector of cosine phase coherence maps can be detrimental and even unnecessary, since both eigenvectors can be modeled using suitable statistical distributions such as the matrix ACG (MACG) (Chikuse, 1990) or singular Wishart (Khatri, 1970) that have the same number of parameters as those for the leading eigenvector ACG model. We also developed K-means approaches for the rank-2 cosine phase coherence maps to be used by researchers seeking a faster-to-implement model (see Supplementary Information). Thus, we hope to have shown that the otherwise popular LEiDA-approach can be vastly improved by making relatively small alterations in the modeling.

In multivariate statistical modeling, attention to the appropriate data manifold is crucial. Eigenvectors, which are by definition unit norm and sign-invariant, lie on the projective hyperplane, also known as the sign-symmetric unit hypersphere. Assuming the wrong manifold, such as using least-squares or cosine distance methods on eigenvectors, can introduce inaccuracies, even with eigenvector sign-flipping, which

imposes non-affine restrictions on the data space (Olsen et al., 2022). Some studies have suggested a K-means approach using either the l1-norm, cosine distance, or Euclidean distance between the upper-triangular part of the *full* cosine phase coherence maps and centroids of the same size (Castro et al., 2024; de Alteriis et al., 2024; Demertzi et al., 2019). Representing a triangular part of a symmetric matrix disregards the Riemmanian structure leading to systematic modeling errors. Instead, using the Frobenius norm results in a hard-assignment variant of the Wishart mixture model (Hidot & Saint-Jean, 2010; Nielsen et al., 2017), serving as a full-rank special case of the "weighted Grassmann clustering" introduced in our Supplementary Material. A Wishart K-means approach would likely lead to component over-parameterization in high dimensions, as brain connectivity often exhibits intercorrelations (Fig. 3B). Models that align with the correct manifold and allow for components of variable rank are thus more suitable.

We estimated our models using both analytically derived update rules via the expectation maximization (EM) as well as numerical optimization in PyTorch. We generally found the two methods to perform similarly with the latter being faster for large data sets. However, since the EM algorithm is more consistent (guarantees monotone convergence in many cases), we supply update rules for all included distributions, including low-rank-plus-diagonal approximation, in the Supplementary Material. Of these, to the best of our knowledge, EM update rules for the ACG, MACG, and Singular Wishart mixture models are new. While our analyses were limited to low-dimensional data, the low-rank-plus-diagonal approximation combined with low-rank initialization of high-rank models (see Methods) support the feasibility of future models in high, potentially native, resolution.

We showed that the global average phase coherence matrix without GSR is highly homogeneous (Supplementary Fig. S8). Clustering cosine phase coherence maps without GSR typically results in one dominant global component with a high weight (Lord et al., 2019; Olsen et al., 2022), while other clusters, previously referred to as ghost attractors (Vohryzek et al., 2020), carry less weight. We believe that applying GSR, which removes more trivial coherence structures, will directly lead to learned components having more contrast, characterizing more detailed and relevant connectivity information.

A disadvantage of synchronization methods employing the Hilbert transform is the requirement of data to be narrowband for phase estimation. We used a filter in the range $[0.03, 0.07]Hz$ based on previous studies (Glerean et al., 2012; Honari et al., 2021; Kringelbach et al., 2020; Pedersen et al., 2018), which is a relatively small restriction on the $[0.01, 0.1]Hz$ range that the BOLD signal is generally accepted to fall within. Future studies should further investigate the effect bandwidth on phase estimation

in BOLD fMRI data, e.g., focusing on detecting phase slips (Pikovsky et al., 2001). Previously, the empirical mode decomposition has been suggested as a data-driven method to extract narrowband signal "modes" prior to the HIlbert transform (Honari & Lindquist, 2022; Rehman & Mandic, 2010). However, it is difficult to reconcile frequency bands of interest across subjects and in practice, this method subdivides the frequency axis log-linearly with little indication of the relative relevance of each derived frequency band.

Besides the frequency range, our study has other limitations. The computational training speed increased markedly from K-means (time scale of seconds for 155 subjects) to mixture modeling (minutes) and hidden Markov Modeling (hours). However, we did not exploit GPU-processing, which is readily available in PyTorch and could be implemented in future studies. We compared methods for phase coherence with zero-mean Gaussian mixture models for "raw" time-series data, however, in the literature, it is more common to see hidden Markov models for fMRI data, perhaps augmented with autoregressive parameters (Ahrends et al., 2022; Liégeois et al., 2019; Stevner et al., 2019; Vidaurre et al., 2016, 2017). Here we observed little difference between mixture and hidden Markov modeling (Supplementary Fig. S4). With the PyTorch stochastic estimation framework, future studies could very easily add an autoregressive emission specification to the presented models. Finally, we only examined one dataset; the estimated components may very well differ in different data.

Conclusion

Modeling time-varying spatiotemporal coherence of functional brain imaging signals can provide meaningful insights about neural systems underlying cognition and illness. Here we demonstrated that dynamic macroscale functional brain phase coherence networks can be modeled in their entirety considering mixture modeling approaches with 1) complex-valued and 2) non-unit rank centroids. We systematically considered a variety of existing and new models for phase synchronization and demonstrated their efficacy on both synthetic and real resting-state and task fMRI. We suggest that future studies within phase coherence, whether in neuroscience or other fields, consider models such as the complex ACG distribution that account for data anisotropy and phase coherence in its entirety. Our Phase Coherence Mixture Modeling (PCMM) toolbox is readily available for general-purpose phase coherence modeling: github.com/anders-s-olsen/PCMM.

METHODS

Table 1: Signal preprocessing considerations regarding phase estimation via the Hilbert transformation. The processing steps are outlined for a univariate signal and should be repeated across ROIs/voxels $j = 1, \ldots, p$. Padding size in step one should be large enough to reasonably cover the filter impulse response and be applied to both the front and back end of the signal in the case of zero-phase (forward-reverse) filtering.

Signal preprocessing step	**Reason**	**Example numpy/scipy code**
Pad signal $x_j(t)$ with mirrors of itself	To avoid edge effects from bandpass filtering and the Hilbert transform (the effect is smaller for phase than for amplitude data)	`pad = 100` `x_j_padded = numpy.concatenate([x_j[1:pad+1][::-1], x_j, x[-pad-1:-1][::-1]])`
Narrowband bandpass filter, e.g., $0.03 - 0.07Hz$ for BOLD fMRI	Phase estimation requires "narrowband" data. The exact signal bandwidth required is unclear	`b, a = scipy.signal.butter(N=2, Wn=[0.03,0.07], btype="bandpass", fs=fs)` `x_j_filtered = scipy.signal.filtfilt(b, a, x_j_padded)(Huang et al., 1998)`
Subtract the mean	Phase estimation breaks down for absolute signal mean higher than one, i.e., if $\frac{1}{N}\sum_t x(t) > 1$ (Huang et al., 1998)	`x_j_filtered_demeaned = x_j_filtered - numpy.mean(x_j_filtered)`

Decompose signal into phase $\theta_j(t)$ and magnitude $a_j(t)$ using the Hilbert transform	Separates the oscillatory and amplitude constituents of the signal.	`x_j_hilbert = scipy.signal.hilbert(x_j_filtered_demeaned)`
		`a_j = numpy.abs(x_j_hilbert)`
		`theta_j = numpy.angle(x_j_hilbert)`

Trim the obtained magnitude and phase signals to the original signal length	To avoid artificially increasing the signal length via padding	`a_j = a_j[pad:-pad]` `theta_j = theta_j[pad:-pad]`
Convert phase from polar to rectangular format and rescale by root dimensionality	Enables modeling of phase data on the complex projective hyperplane using, e.g., the complex ACG distribution	`u_j = 1/numpy.sqrt(p)*numpy.exp(1j*theta_j)`

Optional: Save both magnitude and phase for modeling or only the phase	Modeling only the phase $\{u_j = \frac{1}{\sqrt{p}} e^{i\theta_j}\}_{j=1}^{p}$ ensures an exclusive focus on *synchronization* and avoids amplitude-specific artifacts	`data[:,j] = u_j # complex phase vector`
		`data[:,j] = a_j * u_j # analytic signal`
		`data[:,j] = x_j_filtered_demeaned # filtered time-series`

Experimental data

Modeling phase in neuroimaging data requires some specific preprocessing considerations, see Table 1. Of particular note, phase estimation via the Hilbert transform requires narrowband data; however, no direct guidelines for the required signal bandwidth exist, only rules of thumb such as requiring the

bandwidth to be smaller than the mid-frequency (Feldman, 2011) or twice the mid-frequency (Sun & Small, 2009), both of which we uphold using $0.03 - 0.07Hz$ as our frequency range.

We used fMRI data from 255 unrelated healthy subjects from the openly available Human Connectome Project (Elam et al., 2021; Van Essen et al., 2013). Ethics statements and imaging protocols may be found elsewhere (Barch et al., 2013; Smith et al., 2013); we downloaded the MSM-All preprocessed version of the resting-state data and the minimally preprocessed version of the task data (Glasser et al., 2013). Additionally, we spatially downsampled the data to 100 cortical and 16 subcortical regions using the Schaefer and Tian atlases (Schaefer et al., 2018; Tian et al., 2020). Followingly, we bandpass-filtered the data in the range $[0.03, 0.07]Hz$ with a 2nd-order zero-phase butterworth filter. Due to the relatively short acquisition lengths and long impulse response of the filter giving rise to edge effects from both filtering and the Hilbert transform, we prepended and appended each signal with 100 flipped samples of the signal itself, which corresponds exactly to 1% of the peak of the filter impulse response for TR=0.72s. The added samples were discarded after the Hilbert transform. Of the 255 subjects, 155 randomly selected individuals were used for train and (for resting-state) within-subject validation sets, while the remaining 100 subjects were exclusively used as an out-of-sample test set. For the 155 subjects and resting-state data, one of two scans (REST1 or REST2, randomly selected) constituted the training set and the other a within-subject test set, which was not used due to a strong similarity between the out-of-sample and within-sample test analyses. One subject was discarded from the task analysis out-of-sample test set due to not having all task scans complete.

To evaluate performance upon increasing model complexity, we used phase-controlled 116-dimensional resting-state data from one subject and one scan, i.e., $n = 1200$ data points corresponding to 15 minutes of data acquisition. First we divided both n and p into $K = 5$ groups, such that cluster 1 corresponded to the first 240 time points and the first 23 regions, cluster 2 corresponded to time points 241 to 480 and regions 24-46 and so on. Then we computed, for all regions and temporal windows, the Fourier spectrum and if the region and window corresponded to one of the clusters, the phase information for the relevant frequencies was set to a pre-specified constant phase $\alpha_j \in [0; 2\pi]$. If not, the phase information within the frequency band was randomized.

Dynamic interregional phase coherence

Whereas the Pearson correlation coefficient between regional neuroimaging time-series over time implicitly assumes that the brain covariance matrix is static, time-varying phase coherence can be

estimated via the Hilbert transform. Evaluating dynamic brain coherence implicitly assumes that the signal in all brain areas is of the same scale and that coactivity between regional time courses can be wholly represented by the coherence between their phases.

The regional neuroimaging signal $x_j(t)$ for region $j \in \{1, \ldots, p\}$ is extended into the complex domain with the Hilbert transform $x_j^{(h)}(t) = x_j(t) * \frac{1}{\pi t}$, where $*$ is the convolution operator, to form the analytic signal

$$z_j(t) = x_j(t) + i x_j^{(h)}(t).$$

The Hilbert transform represents a phase shift by $\pi/2$ radians to every frequency in the signal. While the original signal $x_j(t)$ is oscillating along the time axis, the analytic signal $z_j(t)$ is revolving around the unit circle in the complex plane (Fig. 1 and Supplementary Fig. S1). In the complex plane every point may be described using polar coordinates, i.e., as a combination of an angle to the positive real axis

$$\theta_j(t) = \tan^{-1}\left(\frac{x_j^{(h)}(t)}{x_j(t)}\right),$$

also called instantaneous phase, and a distance from the origin $a_j(t) = \sqrt{x_j(t)^2 + \left(x_j^{(h)}(t)\right)^2}$, known as instantaneous magnitude. The phase captures solely the oscillatory information in the signal while signal strength, including potential artifactual spikes, forms part of the signal amplitude. Whereas the Hilbert transform itself does not impose frequency restrictions on the signal, reliable estimation of phase requires that the signal be sufficiently narrowband (Boashash, 1992). Furthermore, while the naming convention "instantaneous" implies that phases can be calculated using information from only a single time point, in reality the Hilbert transform suffers from the same estimation requirements and linearity/stationarity assumptions for a sufficient number of samples as the Fourier transform.

Each sample of the analytic signal is a complex number $z_j(t) = a_j(t)e^{i\theta_j(t)}$, where the radius is ignored when exclusively considering phase coherence, i.e., $a = 1$. The phase θ and difference between two phases $\theta_1 - \theta_2$ are circular variables with modus 2π. The polar representation can therefore be used to assess coherence via $e^{i(\theta_1 - \theta_2)}$ and more generally, all pairwise complex instantaneous phase coherences can be represented by the matrix $\mathbf{A}_t$ with elements

$$A_{t,j,j'} = e^{i(\theta_{t,j} - \theta_{t,j'})} = \cos(\theta_{t,j} - \theta_{t,j'}) + i\sin(\theta_{t,j} - \theta_{t,j'}),$$

such that every element for nodes $j, j' \in \{1, \ldots, p\}$ provides an estimate of the pairwise instantaneous synchrony between regions. Each element of $\mathbf{A}_t$ can be factored as $A_{t,j,j'} = e^{i(\theta_{t,j} - \theta_{t,j'})} = e^{i\theta_{t,j}} e^{-i\theta_{t,j'}}$,

and thus, the matrix can be written as $\mathbf{A}_t = \mathbf{u}_t \mathbf{u}_t^{\mathsf{H}}$, where $\mathbf{u}_t = e^{i\boldsymbol{\theta}_t}$ and $(\cdot)^{\mathsf{H}}$ represents the complex conjugate transpose. Thus, the rank of $\mathbf{A}_t$ is 1, and instantaneous interregional phase coherence can be represented using a vector, $\mathbf{u}_t$, instead of a matrix with $\frac{p(p-1)}{2}$ unique elements, greatly simplifying the modeling of all pairwise coherences. The sum of eigenvalues of a symmetric matrix is always equal to the trace, and since $\mathrm{diag}(\mathbf{A}_t) = \mathbf{1}$, i.e., a vector of all ones, the only non-zero eigenvalue of $\mathbf{A}_t$ is $\lambda = p$. Thus, $\mathbf{u}_t$ may readily be rescaled to unit norm, $\mathbf{u}_t = \frac{1}{\sqrt{p}} e^{i\boldsymbol{\theta}_t}$. Furthermore, flipping the sign of $\mathbf{u}_t$ does not change $\mathbf{A}_t$, meaning that $\mathbf{u}_t$ is a sign invariant eigenvector and models of interregional phase coherence should therefore also be sign-invariant. Consequently, all pairwise instantaneous phases can be compactly represented using a complex-valued sign-invariant vector residing on the unit sphere, also known as the complex projective hyperplane $\mathbb{CP}^{p-1} = \{\mathbf{u} | \mathbf{u} \in \mathbb{C}^p, \mathbf{u}^{\mathsf{H}}\mathbf{u} = 1, -\mathbf{u} = \mathbf{u}\}$.

It is also possible to model instantaneous phase coherence in the real domain by discarding the imaginary part of the complex phase coherence. Especially in fMRI research, the cosine phase coherence matrix has been widely studied, particularly in the context of the Leading Eigenvector Dynamics Analysis (LEiDA) (Cabral et al., 2017) considering the real symmetric matrix $\mathbf{A}_t^{(r)}$ with elements

$$A_{t,j,j'}^{(r)} = \cos(\theta_{t,j} - \theta_{t,j'}) = \mathrm{Re}(A_{t,j,j'}).$$

Notably, a matrix of cosine differences has rank-2 due to the angle difference identity

$$\mathbf{A}_t^{(r)} = \mathbf{c}_t \mathbf{c}_t^{\top} + \mathbf{s}_t \mathbf{s}_t^{\top},$$

where $\mathbf{c}_t = \cos \boldsymbol{\theta}_t = \mathrm{Re}(\mathbf{u}_t)$ and $\mathbf{s}_t = \sin \boldsymbol{\theta}_t = \mathrm{Im}(\mathbf{u}_t)$, which we showed previously (Olsen et al., 2022). Consequently, $\mathbf{A}_t^{(r)}$ can be expressed in terms of *two* eigenvectors with non-zero eigenvalue. The eigenvectors and eigenvalues may be computed easily by stacking the column vectors $\mathbf{c}_t$ and $\mathbf{s}_t$ horizontally and computing the singular value decomposition (SVD), i.e., $\mathbf{U}_t \boldsymbol{\Sigma}_t \mathbf{V}_t^{\top} = [\mathbf{c}_t, \mathbf{s}_t]$—the eigenvectors are given as the leading two left singular vectors $\mathbf{U}_t$ and the eigenvalues are the squared singular values $\boldsymbol{\Lambda}_t = \boldsymbol{\Sigma}_t^2$. Thus, the potentially very-high-dimensional $\frac{p(p-1)}{2} \times n$ phase coherence maps when defined across n time-points do not need to be explicitly calculated but implicitly analyzed in terms of a $P \times 2T$ matrix representation.

The leading eigenvector of the cosine phase coherence map automatically accounts for at least 50% of the variance and empirically captures anywhere between 50% and 100% (de Alteriis et al., 2024). Several previous studies have focused on only the leading eigenvector $\mathbf{u}_{t,1}$ of $\mathbf{A}_t^{(r)}$, i.e., the eigenvector

corresponding to the largest eigenvalue in the LEiDA-framework (Cabral et al., 2017). Furthermore, these eigenvectors have traditionally been sign-flipped, such that the majority of elements are of the same sign, and subsequently clustered using least-squares K-means (Cabral et al., 2017; Figueroa et al., 2019; Lord et al., 2019; Vohryzek et al., 2020). This approach is a heuristic, which we previously have shown may lead to inaccurate results (see Fig. S1 in (Olsen et al., 2022)).

Manifolds and statistical distributions for low-rank data

Below we present statistical distributions for analyzing multivariate brain data from amplitude, phase, and phase-amplitude perspectives; the presented methods and their characteristics are summarized in Supplementary Table S1. Further distributions for analyzing rank-2 cosine phase coherence matrices (matrix ACG and Singular Wishart) and a rank-1 model for phase coherence (Watson) may be found in the Supplementary Information.

Angular central Gaussian distribution

The angular central Gaussian (ACG) distribution for data on the projective hyperplane has density (Tyler, 1987)

$$f_{\mathrm{ACG}}(\mathbf{u}|\mathbf{\Psi}) = \frac{\Gamma\left(\frac{p}{2}\right)}{2\pi^{p/2}|\mathbf{\Psi}|^{\frac{1}{2}}}\left(\mathbf{u}^\top\mathbf{\Psi}^{-1}\mathbf{u}\right)^{-\frac{p}{2}},\ \mathbf{u}\in\mathbb{P}^{p-1}.$$

Here, $\mathbf{\Psi}\in\mathbb{R}^{p\times p}$ is a symmetric positive definite matrix identifiable up to a positive scale factor ($\mathbf{\Psi}$ may be scaled such that $\mathrm{tr}(\mathbf{\Psi})=p$ for convenience) and $|\cdot|$ is the matrix determinant. The ACG distribution is named based on its sampling procedure: samples from a centered multivariate normal distribution normalized to the sphere follows the ACG, i.e., $\mathbf{x}\sim N(\mathbf{0},\mathbf{\Psi}),\quad \frac{\mathbf{x}}{||\mathbf{x}||}\sim\mathrm{ACG}(\mathbf{\Psi})$ (Mardia & Jupp, 1999). The Bingham distribution, which serves a similar purpose, includes a more complicated normalization constant (the confluent hypergeometric function of matrix argument), which becomes intractable for non-trivial dimensionality, rendering the ACG distribution an attractive alternative.

Maximum likelihood estimation of $\mathbf{\Psi}$ involves solving the log-likelihood function

$$\mathcal{L}_{\mathrm{ACG}}\left(\mathbf{\Psi}|\mathbf{u}_1,\ldots,\mathbf{u}_n\right) = -\frac{n}{2}\ln|\mathbf{\Psi}| - \frac{p}{2}\sum_{t=1}^{n}\ln\left(\mathbf{u}_t^\top\mathbf{\Psi}^{-1}\mathbf{u}_t\right)+\delta,$$

where δ is a constant term unrelated to parameter estimation. For $n>p(p-1)$, the solution is given by the following fixed point update, which also satisfies $\mathrm{tr}(\hat{\mathbf{\Psi}})=p$ (Tyler, 1987) (see Supplementary Information for the derivation)

$$\hat{\mathbf{\Psi}}_{i+1} = p\left\{\sum_{t=1}^{n} \frac{\mathbf{u}_t\mathbf{u}_t^\top}{\mathbf{u}_t^\top\hat{\mathbf{\Psi}}_i\mathbf{u}_t}\right\} \Big/ \left\{\sum_{t=1}^{n} \frac{1}{u_t^\top\hat{\mathbf{\Psi}}_i\mathbf{u}_t}\right\}.$$

The fixed point update can be initialized using $\mathbf{\Psi}_0 = \mathbf{I}$.

Complex angular central Gaussian distribution

The complex version of the ACG distribution is characterized by the density (Kent, 1997)

$$f_{\mathrm{CACG}}(\mathbf{u}|\mathbf{\Psi}) = \frac{\Gamma(p)}{2\pi^p|\mathbf{\Psi}|}\left(\mathbf{u}^{\mathsf{H}}\mathbf{\Psi}^{-1}\mathbf{u}\right)^{-p},\ \mathbf{u} \in \mathbb{CP}^{p-1}.$$

Here, $\mathbf{\Psi} \in \mathbb{C}^{p\times p}$ is hermitian-symmetric positive definite and identifiable up to a positive (real) scale factor. The maximum likelihood estimator for the complex ACG has a similar structure to its real counterpart.

Gaussian distribution

The real-valued multivariate Gaussian distribution has the density

$$f_{\mathrm{G}}\left(\mathbf{x}|\mathbf{\Psi}\right) = \frac{1}{(2\pi)^{p/2}|\mathbf{\Psi}|^{1/2}}e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top\mathbf{\Psi}^{-1}(\mathbf{x}-\boldsymbol{\mu})}, \quad \mathbf{x} \in \mathbb{R}^p.$$

Similarly, the complex-valued multivariate Gaussian distribution has the density

$$f_{\mathrm{CG}}\left(\mathbf{x}|\mathbf{\Psi}\right) = \frac{1}{\pi^p|\mathbf{\Psi}}e^{-(\mathbf{x}-\boldsymbol{\mu})^\top\mathbf{\Psi}^{-1}(\mathbf{x}-\boldsymbol{\mu})}, \quad \mathbf{x} \in \mathbb{C}^p.$$

Probabilistic mixture modeling

Mixture modeling involves computing centroids for K components, each corresponding to a unique probability distribution mixed with component weights $\mathbf{w} = \{w_k\}_{k=1}^K$, $\sum_k w_k = 1$. The mixture density for each observation $\mathbf{y}_t$, which may correspond to, e.g., a (complex) eigenvector $\mathbf{y}_t = \mathbf{u}_t$ (phase coupling, complex ACG), a scaled phase vector $\mathbf{y}_t = \mathbf{a}_t \odot \mathbf{u}_t$ (phase-amplitude, complex Gaussian), the time-series $\mathbf{y}_t = \mathbf{x}_t$ (amplitude coupling, Gaussian), or the two eigenvectors of rank-2 cosine phase coherence maps $\mathbf{y}_t = \mathbf{U}_t$ (rank-2 cosine phase coupling, MACG), potentially scaled by the eigenvalues $\mathbf{y}_t = \mathbf{\Lambda}^{1/2}\mathbf{U}_t$ (rank-2 cosine phase coupling, Singular Wishart), is defined as

$$f_{\mathrm{mix}}(\mathbf{y}_t|\boldsymbol{\phi}, \mathbf{w}) = \sum_{k=1}^{K} w_k f(\mathbf{y}_t|\boldsymbol{\psi}_k),$$

where $\boldsymbol{\phi}_k$ are the parameters of the kth component, e.g., $\boldsymbol{\phi}_k = \mathbf{\Psi}_k$ for distributions with a covariance matrix.

We utilized two estimation schemes for mixture models involving these distributions: expectation-maximization (EM) and direct likelihood optimization. Both methods optimize the mixture model log-likelihood function evaluated over the entire data set $\mathcal{Y} = \{\mathbf{y}_t\}_{t=1}^{N}$

$$\mathcal{L}_{\mathrm{MM}}\left(\boldsymbol{\phi}, \mathbf{w}|\mathcal{Y}\right) = \sum_{t=1}^{N} \ln\left(\sum_{k=1}^{K} w_k f(\mathbf{y}_t|\boldsymbol{\phi}_k)\right).$$

EM computes assignment probabilities (i.e., state responsibilities) β_{kt} for sample t and component k in the expectation (E-step) according to

$$\beta_{kt} = \frac{w_k f(\mathbf{y}_t|\boldsymbol{\phi}_k)}{\sum_{k'} w_{k'} f(\mathbf{y}_t|\boldsymbol{\phi}_{k'})}.$$

These probabilities are used during the maximization (M-step) to guide analytical parameter updates. In contrast, the direct optimization procedure only evaluates the log-likelihood function and updates parameters directly using an optimizer and automatic differentiation. Although the E-step is not required for direct numerical optimization, it is still used to compute posterior probabilities after the parameter estimation.

The Supplementary Information includes the necessary analytical maximization formulas in the M step of the EM algorithm, including weighting by sample-to-component responsibilities β_{kt}, trace-normalization for (complex) ACG and MACG and some computational considerations including measuring the convergence of the fixed-point updates for (complex) ACG.

Hidden Markov Models

Mixture models do not necessarily capture the sequential dependencies present in ordered data, such as a multivariate time-series. Notably, Hidden Markov models (HMMs) explicitly model transition dynamics between components by imposing a transition probability matrix $\mathbf{T}$ with elements $T_{k,k'}$, which governs the transition probability from state k to state k'. Given the complexity of EM optimization for HMMs, especially for large data sets, we utilized the direct numerical optimization method of the log-likelihood function using the forward algorithm:

$$\alpha_{t,k}^{(s)} = \sum_{k'=1}^{K} \alpha_{t-1,k'}^{(s)} T_{k',k} f(\mathbf{y}_t^{(s)}|\boldsymbol{\phi}_k),$$

where $\alpha_{t,k}^{(s)}$ represents the probability of being in state k at time t for the sth sequence (e.g., a subject or separate recording), and $\alpha_{1,k}^{(s)} = w_k f(\mathbf{y}_1^{(s)}|\boldsymbol{\phi}_k)$. Of note, the $\mathbf{w}$-parameter for HMM only applies to the first sample in each sequence and thus has a different interpretation than for mixture models. The log-likelihood to be maximized numerically is

$$\mathcal{L}_{\text{HMM}}(\boldsymbol{\phi}, \mathbf{w}, \mathbf{T}|\mathcal{Y}) = \sum_{s=1}^{S} \ln \left(\sum_{k=1}^{K} \alpha_{n_s,k}^{(s)} \right),$$

where S is the number of observation sequences, and n_s is the length of sequence s. Following the parameter optimization, we used the Viterbi decoder to extract state assignments (Forney, 1973; Viterbi, 1967).

Low-rank-plus-diagonal approximation of $\boldsymbol{\Psi}$

Estimating covariance matrix parameters can be challenging due to the near-quadratic increase in parameters with dimensionality. Moreover, high-resolution brain connectivity networks exhibit rank deficiency due to neighboring and related areas having highly correlated time-series. To address this, a low-rank-plus-diagonal approximation $\boldsymbol{\Psi} \approx \gamma \left(\mathbf{M}\mathbf{M}^{\mathsf{H}} + \mathbf{I} \right)$, with $\mathbf{M} \in \mathbb{R}^{p \times r}$ (or $\mathbf{M} \in \mathbb{C}^{p \times r}$ for the complex ACG), allows capturing a pre-specified rank $r \leq p$ of the learned parameter matrix (Olsen et al., 2023). The constant $\gamma \in \mathbb{R}$ for the (complex) ACG distribution corresponds to a trace-normalization, i.e., $\gamma = \frac{p}{\text{tr}\left(\mathbf{M}\mathbf{M}^{\mathsf{H}} + \mathbf{I}\right)} = \frac{1}{o/p+1}$ where $o = ||\mathbf{M}||_F^2$, which does not affect mixture model results due to the scale-invariance of the (complex) ACG distribution. For the zero-mean Gaussian distribution, which is not scale-invariant, we used the reparameterization $\boldsymbol{\Psi} \approx \mathbf{M}\mathbf{M}^{\mathsf{H}} + \gamma \mathbf{I}$ such that the scale parameter affects the relationship between the added identity matrix and $\mathbf{M}$.

The use of low-rank-plus-diagonal approximation required a re-derivation of existing analytical maximization updates to be with respect to $\mathbf{M}$ (and γ for (complex) Gaussian) as detailed in the Supplementary Material. For scalable inference, we further explored the low rank plus diagonal structure using Woodbury's matrix inversion lemma and the matrix determinant lemma for efficient inversion and evaluation of determinants respectively.

Implementation details

We implemented mixture models with analytical maximization and numerical optimization in PyTorch, where the latter also included the possibility for computing HMMs. Recent advances in fast numerical gradient optimization frameworks such as automatic differentiation in PyTorch has enabled fast non-linear model inference while only requiring the specification of an objective function, in this case the log-likelihood function. An advantage of direct numerical optimization when compared to the more conventional EM inference procedure is that the (complex) ACG and Gaussian models with low-rank-plus-diagonal approximations all require fixed-point estimation for analytical maximization in the M-step. Secondly, while analytical updates ensure monotone convergence, the EM algorithm can be

prone to local optima (Shireman et al., 2016). Conversely, direct numerical optimization requires choosing a learning rate and learning algorithm; Here we used the ADAM optimizer (Kingma & Ba, 2014) with $\mathrm{LR} = 0.1$, in correspondence with adjacent studies (Olsen et al., 2023, 2024).

Parameter constraints for numerical optimization were implemented through reparametrization of unconstrained variables to comply with the imposed constraints: We used a softmax function for the mixture weights $w_k = \frac{e^{\tilde{w}_k}}{\sum_{k'} e^{\tilde{w}_{k'}}}$ and a softplus function for the γ-parameter for Gaussian distributions to ensure non-negativity. The transition probability matrix for HMM was treated as a row-wise probability mass function also implemented using the softmax function.

Analytical (EM) fixed-point updates for (complex) ACG and Gaussian distributions were evaluated until an absolute tolerance of 10^{-10}which we verified worked as expected (Supplementary Fig. S10). Mixture models, both using analytical and numerical optimization, were evaluated until a *relative* difference between the best and second-best iteration (divided by the best) in the latter 50 epochs fell below $\mathrm{tol} = 10^{-10}$. While the EM algorithm typically guarantees monotone convergence, this is not the case for the low-rank-plus-diagonal approximations, since these rely on iterative optimization procedures in the M-step.

K-means equivalents

To provide a comparison to the probabilistic mixture models and to function as initialization to these, we utilized existing hard assignment K-means models for the projective hyperplane and Grassmann manifold and introduced new K-means approaches for the space of symmetric positive semidefinite matrices (rank-2 cosine phase coherence matrices weighted by their eigenvalues), which we call *weighted Grassmann clustering*, as well as complex diametrical clustering for the complex projective hyperplane. We have outlined these approaches in the Supplementary Material.

Initialization schemes

Given the vast array of available models, gradually increasing model complexity may aid in ensuring model convergence to a suitable optimum. The K-means algorithms were initialized using their ++-versions, where an initial centroid is picked randomly and the remaining $K-1$ centroids iteratively picked based on a probability distribution proportional to the distance to existing centroids, ensuring that the starting points were spread far apart (Arthur & Vassilvitskii, 2007). This scheme provided slightly

better results than initializing centroids via the random uniform distribution (Supplementary Fig. S11). Probabilistic mixture models were initialized using centroids derived from K-means. For complex ACG and Gaussian distributions, the low-rank matrix $\mathbf{M}$ was initialized, given a partition of the data set $\mathbf{y}_t \in \mathcal{C}_k$, e.g., from K-means, as

$$\mathbf{M} = \mathbf{U}\left[\tfrac{1}{\epsilon}\left(\mathbf{\Sigma}_r - \epsilon\mathbf{I}\right)\right]^{1/2},$$

where $\mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top = \mathbf{Y}$ is the singular value decomposition and $\mathbf{Y} = [\mathbf{y}_t]_{\mathbf{y}\in\mathcal{C}_k}$ represents the samples assigned to cluster k stacked horizontally. $\mathbf{\Sigma}_r$ is the diagonal matrix of singular values $[\sigma_{ll}]_{l=1}^{r}$, where all other values than the leading r singular values were set to zero, and $\epsilon = \frac{\sum_{l=r+1}^{p}\sigma_l}{p-r}$, assuming the number of assigned samples to cluster k is $n_k > p$. Mixture component weights $\mathbf{w}$ and transition probability matrices $\mathbf{T}$ for HMMs were initialized based on the empirical transition probability matrix from the initialization partition. For the Normal distributions, γ was initialized to the same value as ϵ.

We also investigated a scenario where high-rank mixtures and HMMs were initialized by their lower-rank counterparts. For instance, given the parameter $\mathbf{M}_h$ of a rank $h < r$ model, the first h columns of $\mathbf{M}$ could be initialized as $\mathbf{M}_h$ and the remaining $r-h$ columns as the singular value decomposition on the residual matrix after projecting out the existing columns, i.e., $\mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top = \mathbf{Y} - \mathbf{M}_h\left(\mathbf{M}_h\mathbf{M}_h^\top\right)^{-1}\mathbf{M}_h^\top\mathbf{Y}$. Secondly, we investigated a scenario, where numerical optimization models were initialized from their EM-counterparts and HMM models from their direct numerically optimized mixture counterparts. The combination of these scenarios provided improved results (Supplementary Fig. S4).

Normalized mutual information

Clustering partitions may be easily compared using normalized mutual information (NMI), which is invariant to permutation in clustering labels. Given two binary or weighted responsibility matrices $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ of size $K_1 \times n$ and $K_2 \times n$, respectively, which sum to one for each column, normalized mutual information is defined as (Hinrich et al., 2016)

$$\mathrm{NMI}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) = \frac{2\mathrm{MI}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2)}{\mathrm{MI}(\boldsymbol{\beta}_1,\boldsymbol{\beta}_1)+\mathrm{MI}(\boldsymbol{\beta}_2,\boldsymbol{\beta}_2)},$$

$$\mathrm{MI}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) = \sum_{t,t'} p(t,t')\log\frac{p(t,t')}{p(t)p(t')},$$

$$p(t,t') = \frac{1}{K}\sum_k \beta_1^{kt}\beta_2^{kt'}.$$

We evaluated NMI between true labels and estimated posteriors or cluster labels for label-controlled experiments.

Code and data availability

All processing and analysis scripts used in the present study, implemented primarily in Python, are made available at github.com/anders-s-olsen/PCMM. Human connectome project data are readily available online.

AUTHOR CONTRIBUTIONS

ASO and MM conceived the study and developed the theoretical framework. ASO performed the experiments, wrote the toolbox, and drafted the manuscript. AB aided in model development and PMF aided in analysis of results. MM and PMF supervised the project. All authors contributed to interpretation of results and reviewed and approved the final version of the manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

REFERENCES

Ahrends, C., Stevner, A., Pervaiz, U., Kringelbach, M., Vuust, P., Woolrich, M., & Vidaurre, D. (2022). Data and model considerations for estimating time-varying functional connectivity in fMRI. *NeuroImage*, *252*, 119026. https://doi.org/10.1016/j.neuroimage.2022.119026

Ahrends, C., & Vidaurre, D. (2023). *Dynamic Functional Connectivity* (No. arXiv:2301.03408). arXiv. https://doi.org/10.48550/arXiv.2301.03408

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

Arthur, D., & Vassilvitskii, S. (2007). K-means++: The advantages of careful seeding. *Proceedings of the Annual ACM-SIAM Symposium on Discrete Algorithms*, *07-09-Janu*, 1027–1035.

Barch, D. M., Burgess, G. C., Harms, M. P., Petersen, S. E., Schlaggar, B. L., Corbetta, M., Glasser, M. F., Curtiss, S., Dixit, S., Feldt, C., Nolan, D., Bryant, E., Hartley, T., Footer, O., Bjork, J. M., Poldrack, R., Smith, S., Johansen-Berg, H., Snyder, A. Z., & Essen, D. C. V. (2013). Function in the Human Connectome: Task-fMRI and Individual Differences in Behavior. *NeuroImage*, *80*, 169. https://doi.org/10.1016/j.neuroimage.2013.05.033

Bastos, A. M., & Schoffelen, J. M. (2016). A tutorial review of functional connectivity analysis methods and their interpretational pitfalls. *Frontiers in Systems Neuroscience*,

9(JAN2016). https://doi.org/10.3389/fnsys.2015.00175

Biswal, B., Zerrin Yetkin, F., Haughton, V. M., & Hyde, J. S. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar mri. *Magnetic Resonance in Medicine*, *34*(4), 537–541. https://doi.org/10.1002/mrm.1910340409

Boashash, B. (1992). Estimating and interpreting the instantaneous frequency of a signal. I. Fundamentals. *Proceedings of the IEEE*, *80*(4), 520–538. Proceedings of the IEEE. https://doi.org/10.1109/5.135376

Bola, M., & Sabel, B. A. (2015). Dynamic reorganization of brain functional networks during cognition. *NeuroImage*, *114*, 398–413. https://doi.org/10.1016/j.neuroimage.2015.03.057

Cabral, J., Vidaurre, D., Marques, P., Magalhães, R., Silva Moreira, P., Miguel Soares, J., Deco, G., Sousa, N., & Kringelbach, M. L. (2017). Cognitive performance in healthy older adults relates to spontaneous switching between states of functional connectivity during rest. *Scientific Reports*, *7*(1), 1–13. https://doi.org/10.1038/s41598-017-05425-7

Castro, P., Luppi, A., Tagliazucchi, E., Perl, Y. S., Naci, L., Owen, A. M., Sitt, J. D., Destexhe, A., & Cofré, R. (2024). Dynamical structure-function correlations provide robust and generalizable signatures of consciousness in humans. *Communications Biology*, *7*(1), 1–12. https://doi.org/10.1038/s42003-024-06858-3

Chen, J. E., Rubinov, M., & Chang, C. (2017). Methods and Considerations for Dynamic Analysis of Functional MR Imaging Data. *Neuroimaging Clinics of North America*, *27*(4), 547–560. https://doi.org/10.1016/j.nic.2017.06.009

Chikuse, Y. (1990). The matrix angular central Gaussian distribution. *Journal of Multivariate Analysis*, *33*(2), 265–274. https://doi.org/10.1016/0047-259X(90)90050-R

Choe, A. S., Nebel, M. B., Barber, A. D., Cohen, J. R., Xu, Y., Pekar, J. J., Caffo, B., & Lindquist, M. A. (2017). Comparing test-retest reliability of dynamic functional connectivity methods. *NeuroImage*, *158*, 155–175. https://doi.org/10.1016/j.neuroimage.2017.07.005

Coombes, S., & Bressloff, P. C. (2005). *Bursting: The Genesis of Rhythm in the Nervous*

System. World Scientific.

de Alteriis, G., MacNicol, E., Hancock, F., Ciaramella, A., Cash, D., Expert, P., & Turkheimer, F. E. (2024). EiDA: A lossless approach for dynamic functional connectivity; application to fMRI data of a model of ageing. *Imaging Neuroscience*, *2*, 1–22. https://doi.org/10.1162/imag_a_00113

Demertzi, A., Tagliazucchi, E., Dehaene, S., Deco, G., Barttfeld, P., Raimondo, F., Martial, C., Fernández-Espejo, D., Rohaut, B., Voss, H. U., Schiff, N. D., Owen, A. M., Laureys, S., Naccache, L., & Sitt, J. D. (2019). Human consciousness is supported by dynamic complex patterns of brain signal coordination. *Science Advances*, *5*(2), eaat7603. https://doi.org/10.1126/sciadv.aat7603

Elam, J. S., Glasser, M. F., Harms, M. P., Sotiropoulos, S. N., Andersson, J. L. R., Burgess, G. C., Curtiss, S. W., Oostenveld, R., Larson-Prior, L. J., Schoffelen, J. M., Hodge, M. R., Cler, E. A., Marcus, D. M., Barch, D. M., Yacoub, E., Smith, S. M., Ugurbil, K., & Van Essen, D. C. (2021). The Human Connectome Project: A retrospective. *NeuroImage*, *244*(August), 118543. https://doi.org/10.1016/j.neuroimage.2021.118543

Feldman, M. (2011). Hilbert transform in vibration analysis. *Mechanical Systems and Signal Processing*, *25*(3), 735–802. https://doi.org/10.1016/j.ymssp.2010.07.018

Figueroa, C. A., Cabral, J., Mocking, R. J. T., Rapuano, K. M., van Hartevelt, T. J., Deco, G., Expert, P., Schene, A. H., Kringelbach, M. L., & Ruhé, H. G. (2019). Altered ability to access a clinically relevant control network in patients remitted from major depressive disorder. *Human Brain Mapping*, *40*(9), 2771–2786. https://doi.org/10.1002/hbm.24559

Forney, G. D. (1973). The viterbi algorithm. *Proceedings of the IEEE*, *61*(3), 268–278. Proceedings of the IEEE. https://doi.org/10.1109/PROC.1973.9030

Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J. L., Xu, J., Jbabdi, S., Webster, M., Polimeni, J. R., Van Essen, D. C., & Jenkinson, M. (2013). The minimal preprocessing pipelines for the Human Connectome Project. *NeuroImage*,

80, 105–124. https://doi.org/10.1016/j.neuroimage.2013.04.127

Glerean, E., Salmi, J., Lahnakoski, J. M., Jääskeläinen, I. P., & Sams, M. (2012). Functional Magnetic Resonance Imaging Phase Synchronization as a Measure of Dynamic Functional Connectivity. *Brain Connectivity*, *2*(2), 91–101. https://doi.org/10.1089/brain.2011.0068

Hidot, S., & Saint-Jean, C. (2010). An Expectation–Maximization algorithm for the Wishart mixture model: Application to movement clustering. *Pattern Recognition Letters*, *31*(14), 2318–2324. https://doi.org/10.1016/j.patrec.2010.07.002

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. https://doi.org/10.1016/j.neuroimage.2015.11.055

Hinrich, J. L., Bardenfleth, S. E., Roge, R. E., Churchill, N. W., Madsen, K. H., & Morup, M. (2016). Archetypal Analysis for Modeling Multisubject fMRI Data. *IEEE Journal on Selected Topics in Signal Processing*, *10*(7), 1160–1171. https://doi.org/10.1109/JSTSP.2016.2595103

Honari, H., Choe, A. S., & Lindquist, M. A. (2021). Evaluating phase synchronization methods in fMRI: A comparison study and new approaches. *NeuroImage*, *228*, 117704. https://doi.org/10.1016/j.neuroimage.2020.117704

Honari, H., & Lindquist, M. A. (2022). Mode decomposition-based time-varying phase synchronization for fMRI. *NeuroImage*, *261*, 119519. https://doi.org/10.1016/j.neuroimage.2022.119519

Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., Yen, N.-C., Tung, C. C., & Liu, H. H. (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, *454*(1971),

903–995. https://doi.org/10.1098/rspa.1998.0193

Kent, J. T. (1997). Data analysis for shapes and images. *Journal of Statistical Planning and Inference*, *57*(2), 181–193. https://doi.org/10.1016/S0378-3758(96)00043-2

Khatri, C. G. (1970). A Note on Mitra's Paper "A Density Free Approach to the Matrix Variate Beta Distribution." *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002)*, *32*(3), 311–318.

Kingma, D. P., & Ba, J. L. (2014). Adam: A Method for Stochastic Optimization. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*. https://doi.org/10.48550/arxiv.1412.6980

Kirschstein, T., & Köhling, R. (2009). What is the Source of the EEG? *Clinical EEG and Neuroscience*, *40*(3), 146–149. https://doi.org/10.1177/155005940904000305

Kringelbach, M. L., Cruzat, J., Cabral, J., Knudsen, G. M., Carhart-Harris, R., Whybrow, P. C., Logothetis, N. K., & Deco, G. (2020). Dynamic coupling of whole-brain neuronal and neurotransmitter systems. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(17), 9566–9576. https://doi.org/10.1073/pnas.1921475117

Laumann, T. O., Snyder, A. Z., & Gratton, C. (2024). Challenges in the measurement and interpretation of dynamic functional connectivity. *Imaging Neuroscience*. https://doi.org/10.1162/imag_a_00366

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., McCarthy, J. E., Tagliazucchi, E., Laufs, H., Schlaggar, B. L., Dosenbach, N. U. F., & Petersen, S. E. (2017). On the Stability of BOLD fMRI Correlations. *Cerebral Cortex*, *27*(10), 4719–4732. https://doi.org/10.1093/cercor/bhw265

Leonardi, N., & Van De Ville, D. (2015). On spurious and real fluctuations of dynamic functional connectivity during rest. *NeuroImage*, *104*, 430–436. https://doi.org/10.1016/j.neuroimage.2014.09.007

Liégeois, R., Laumann, T. O., Snyder, A. Z., Zhou, J., & Yeo, B. T. T. (2017). Interpreting temporal fluctuations in resting-state functional connectivity MRI. *NeuroImage*, *163*, 437–455. https://doi.org/10.1016/j.neuroimage.2017.09.012

Liégeois, R., Li, J., Kong, R., Orban, C., Van De Ville, D., Ge, T., Sabuncu, M. R., & Yeo, B. T. T. (2019). Resting brain dynamics at different timescales capture distinct aspects of human behavior. *Nature Communications*, *10*(1), 2317. https://doi.org/10.1038/s41467-019-10317-7

Lord, L. D., Expert, P., Atasoy, S., Roseman, L., Rapuano, K., Lambiotte, R., Nutt, D. J., Deco, G., Carhart-Harris, R. L., Kringelbach, M. L., & Cabral, J. (2019). Dynamical exploration of the repertoire of brain networks at rest is modulated by psilocybin. *NeuroImage*, *199*(April), 127–142. https://doi.org/10.1016/j.neuroimage.2019.05.060

Maeda, E., Robinson, H., & Kawana, A. (1995). The mechanisms of generation and propagation of synchronized bursting in developing networks of cortical neurons. *The Journal of Neuroscience*, *15*(10), 6834–6845. https://doi.org/10.1523/JNEUROSCI.15-10-06834.1995

Mardia, K. V., & Jupp, P. E. (1999). Directional Statistics. In *Directional Statistics*. John Wiley and Sons Ltd. https://doi.org/10.1002/9780470316979

Nielsen, S. F. V., Madsen, K. H., Schmidt, M. N., & Morup, M. (2017). Modeling dynamic functional connectivity using a wishart mixture model. *2017 International Workshop on Pattern Recognition in Neuroimaging, PRNI 2017*. https://doi.org/10.1109/PRNI.2017.7981505

Nolte, G., Bai, O., Wheaton, L., Mari, Z., Vorbach, S., & Hallett, M. (2004). Identifying true brain interaction from EEG data using the imaginary part of coherency. *Clinical Neurophysiology*, *115*(10), 2292–2307. https://doi.org/10.1016/j.clinph.2004.04.029

Olsen, A. S., Lykkebo-Valløe, A., Ozenne, B., Madsen, M. K., Stenbæk, D. S., Armand, S., Mørup, M., Ganz, M., Knudsen, G. M., & Fisher, P. M. (2022). Psilocybin modulation of

time-varying functional connectivity is associated with plasma psilocin and subjective effects. *NeuroImage*, 119716. https://doi.org/10.1016/J.NEUROIMAGE.2022.119716

Olsen, A. S., Nielsen, J. D., & Mørup, M. (2024). Coupled Generator Decomposition for Fusion of Electro- and Magnetoencephalography Data. *2024 32nd European Signal Processing Conference (EUSIPCO)*, 1357–1361. https://ieeexplore.ieee.org/document/10715032

Olsen, A. S., Ortvald, E., Madsen, K. H., Schmidt, M. N., & Mørup, M. (2023). Angular Central Gaussian and Watson Mixture Models for Assessing Dynamic Functional Brain Connectivity During a Motor Task. *2023 IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops (ICASSPW)*, 1–5. https://doi.org/10.1109/ICASSPW59220.2023.10193021

Pedersen, M., Omidvarnia, A., Zalesky, A., & Jackson, G. D. (2018). On the relationship between instantaneous phase synchrony and correlation-based sliding windows for time-resolved fMRI connectivity analysis. *NeuroImage*, *181*, 85–94. https://doi.org/10.1016/j.neuroimage.2018.06.020

Pikovsky, A., Rosenblum, M., & Kurths, J. (2001). *Synchronization: A Universal Concept in Nonlinear Sciences*. Cambridge University Press. https://doi.org/10.1017/CBO9780511755743

Preti, M. G., Bolton, T. A., & Van De Ville, D. (2017). The dynamic functional connectome: State-of-the-art and perspectives. *NeuroImage*, *160*, 41–54. https://doi.org/10.1016/j.neuroimage.2016.12.061

Rehman, N., & Mandic, D. P. (2010). Multivariate empirical mode decomposition. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *466*(2117), 1291–1302. https://doi.org/10.1098/rspa.2009.0502

Schaefer, A., Kong, R., Gordon, E. M., Laumann, T. O., Zuo, X.-N., Holmes, A. J., Eickhoff, S. B., & Yeo, B. T. T. (2018). Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI. *Cerebral Cortex*, *28*(9), 3095–3114.

https://doi.org/10.1093/cercor/bhx179

Shireman, E. M., Steinley, D., & Brusco, M. J. (2016). Local optima in mixture modeling. *Multivariate Behavioral Research*, *51*(4), 466–481. https://doi.org/10.1080/00273171.2016.1160359

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., Kelly, M., Laumann, T., Miller, K. L., Moeller, S., Petersen, S., Power, J., Salimi-Khorshidi, G., Snyder, A. Z., Vu, A. T., … Glasser, M. F. (2013). Resting-state fMRI in the Human Connectome Project. *NeuroImage*, *80*, 144–168. https://doi.org/10.1016/J.NEUROIMAGE.2013.05.039

Stevner, A. B. A., Vidaurre, D., Cabral, J., Rapuano, K., Nielsen, S. F. V., Tagliazucchi, E., Laufs, H., Vuust, P., Deco, G., Woolrich, M. W., Van Someren, E., & Kringelbach, M. L. (2019). Discovery of key whole-brain transitions and dynamics during human wakefulness and non-REM sleep. *Nature Communications*, *10*(1). https://doi.org/10.1038/s41467-019-08934-3

Sun, J., & Small, M. (2009). Unified framework for detecting phase synchronization in coupled time series. *Physical Review E*, *80*(4), 046219. https://doi.org/10.1103/PhysRevE.80.046219

Tian, Y., Margulies, D. S., Breakspear, M., & Zalesky, A. (2020). Topographic organization of the human subcortex unveiled with functional connectivity gradients. *Nature Neuroscience*, *23*(11), 1421–1432. https://doi.org/10.1038/s41593-020-00711-6

Tyler, D. E. (1987). Statistical Analysis for the Angular Central Gaussian Distribution on the Sphere. *Biometrika*, *74*(3), 579. https://doi.org/10.2307/2336697

Uhlhaas, P., Pipa, G., Lima, B., Melloni, L., Neuenschwander, S., Nikolić, D., & Singer, W. (2009). Neural synchrony in cortical networks: History, concept and current status. *Frontiers in Integrative Neuroscience*, *3*. https://doi.org/10.3389/neuro.07.017.2009

Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E. J., Yacoub, E., & Ugurbil, K.

(2013). The WU-Minn Human Connectome Project: An overview. *NeuroImage*, *80*, 62–79. https://doi.org/10.1016/J.NEUROIMAGE.2013.05.041

Vidaurre, D., Quinn, A. J., Baker, A. P., Dupret, D., Tejero-Cantero, A., & Woolrich, M. W. (2016). Spectrally resolved fast transient brain states in electrophysiological data. *NeuroImage*, *126*, 81–95. https://doi.org/10.1016/j.neuroimage.2015.11.047

Vidaurre, D., Smith, S. M., & Woolrich, M. W. (2017). Brain network dynamics are hierarchically organized in time. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(48), 12827–12832. https://doi.org/10.1073/pnas.1705120114

Viterbi, A. (1967). Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory*, *13*(2), 260–269. IEEE Transactions on Information Theory. https://doi.org/10.1109/TIT.1967.1054010

Vohryzek, J., Deco, G., Cessac, B., Kringelbach, M. L., & Cabral, J. (2020). Ghost Attractors in Spontaneous Brain Activity: Recurrent Excursions Into Functionally-Relevant BOLD Phase-Locking States. *Frontiers in Systems Neuroscience*, *14*, 20. https://doi.org/10.3389/fnsys.2020.00020

Wang, H. E., Bénar, C. G., Quilichini, P. P., Friston, K. J., Jirsa, V. K., & Bernard, C. (2014). A systematic framework for functional connectivity measures. *Frontiers in Neuroscience*, *8*. https://doi.org/10.3389/fnins.2014.00405

Womelsdorf, T., Schoffelen, J.-M., Oostenveld, R., Singer, W., Desimone, R., Engel, A. K., & Fries, P. (2007). Modulation of Neuronal Interactions Through Neuronal Synchronization. *Science*, *316*(5831), 1609–1612. https://doi.org/10.1126/science.1139597

Zarghami, T. S., Hossein-Zadeh, G.-A., & Bahrami, F. (2020). Deep Temporal Organization of fMRI Phase Synchrony Modes Promotes Large-Scale Disconnection in Schizophrenia. *Frontiers in Neuroscience*, *14*. https://doi.org/10.3389/fnins.2020.00214