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, Complex Watson, Dynamic functional connectivity, Riemannian clustering, complex diametrical clustering, Grassmann clustering, Weighted Grassmann clustering, Matrix angular central Gaussian, Singular Wishart, directional statistics, Probabilistic mixture models, Hidden Markov models, Phase coherence, Multivariate synchronization.

ABSTRACT

Complex behavioral and cognitive processes emerge from coordinated communication between sometimes disparate brain regions, implying a systems-level dynamic synchronization of underlying neural signals. This study presents approaches to multivariate mixture modeling of functional brain imaging data for analyzing phase coherence networks for human brain macroscale dynamic functional connectivity. We show that 1) the complex domain is required to analyze phase coherence in its entirety, 2) statistical models for complex-valued phase coherence, particularly the complex angular central gaussian (ACG) distribution, greatly exceeds performance over models for time-series data, rank-2 cosine phase coherence maps, or the leading eigenvector of cosine phase coherence maps (i.e., LEiDA), 3) methods should account for the inherent anisotropy of the brain's interconnections. We emphasize the

need to utilize models that account for the manifold on which the data reside, and show that the proposed models provide valuable information about intrinsic functional connectivity networks and can easily distinguish task-related brain function. We provide an open-source, Python-based toolbox ("Phase Coherence Mixture Modeling" (PCMM): github.com/anders-s-olsen/PCMM) to facilitate and promote implementation of these models.

INTRODUCTION

Synchronization is a fundamental concept in science, observed across diverse fields from coherent light in lasers, to 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 indispensable for capturing the complexity of brain-wide connectivity patterns, key to understanding the orchestrated dynamics within and across neural networks.

Traditionally, macroscale brain connectivity between regional brain 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 both amplitude and oscillatory coactivation of brain signals (Bastos & Schoffelen, 2016; Wang et al., 2014). Brain connectivity is known to be time-varying (Chang & Glover, 2010; Lurie et al., 2018; Zalesky et al., 2014) reflecting the inherently dynamic nature of brain function and as a result, considerable research has focused on methods to characterize this time-varying nature (Calhoun et al., 2014; Hutchison et al., 2013; Lurie et al., 2018; Preti et al., 2017). Estimating time-varying correlation coefficients requires temporal windowing of input signals, 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 zero-mean 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 between brain signals, discarding amplitude information, offers a frame-wise connectivity proxy free from area-specific amplitude-related noise uninformative of connectivity (Glerean et al., 2012).

Assessing phase coherence involves converting signals to the complex domain, typically via the Hilbert transform, which translates temporal evolution into a revolution around the unit circle. The phase of a signal is a circular quantity with a period of 2π (Fig. 1A) from which a *complex dynamic phase coherence map* $\mathbf{A}_t$ for sample t with phase difference elements $A_{t,j,j'} = \mathrm{e}^{i\left(\theta_{t,j} - \theta_{t,j'}\right)}$ can be constructed. Most previous studies in fMRI phase coherence analysis have focused exclusively on *cosine* phase coherence, i.e., $\mathbf{A}_t^{(r)} = \mathrm{Re}\left(\mathbf{A}_t\right)$ (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, 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 imaginary dimension is favored (Nolte et al., 2004). We previously demonstrated that matrices with cosine difference entries are 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 $\mathbf{c}_t = \cos \boldsymbol{\theta}_t$ and $\mathbf{s}_t = \sin \boldsymbol{\theta}_t$ (Olsen et al., 2022). However, some studies have focused exclusively on the leading eigenvector of $\mathbf{A}_t^{(r)}$, particularly through Leading Eigenvector Dynamics Analysis (LEiDA) (Cabral et al., 2017), thereby explicitly discarding a portion of the data variance within the cosine projection of the phase coherence. Moreover, studies involving LEiDA typically model the leading eigenvectors using least-squares methods, thereby ignoring the underlying manifold structure of eigenvectors (i.e., unit norm and sign invariance), leading to inaccuracies (Olsen et al., 2022). A sign-flip imposed on eigenvectors, supposed to make these amenable to least-squares modeling, results in a non-affine restriction of the point space (see Fig. S1 in Olsen et al., 2022).

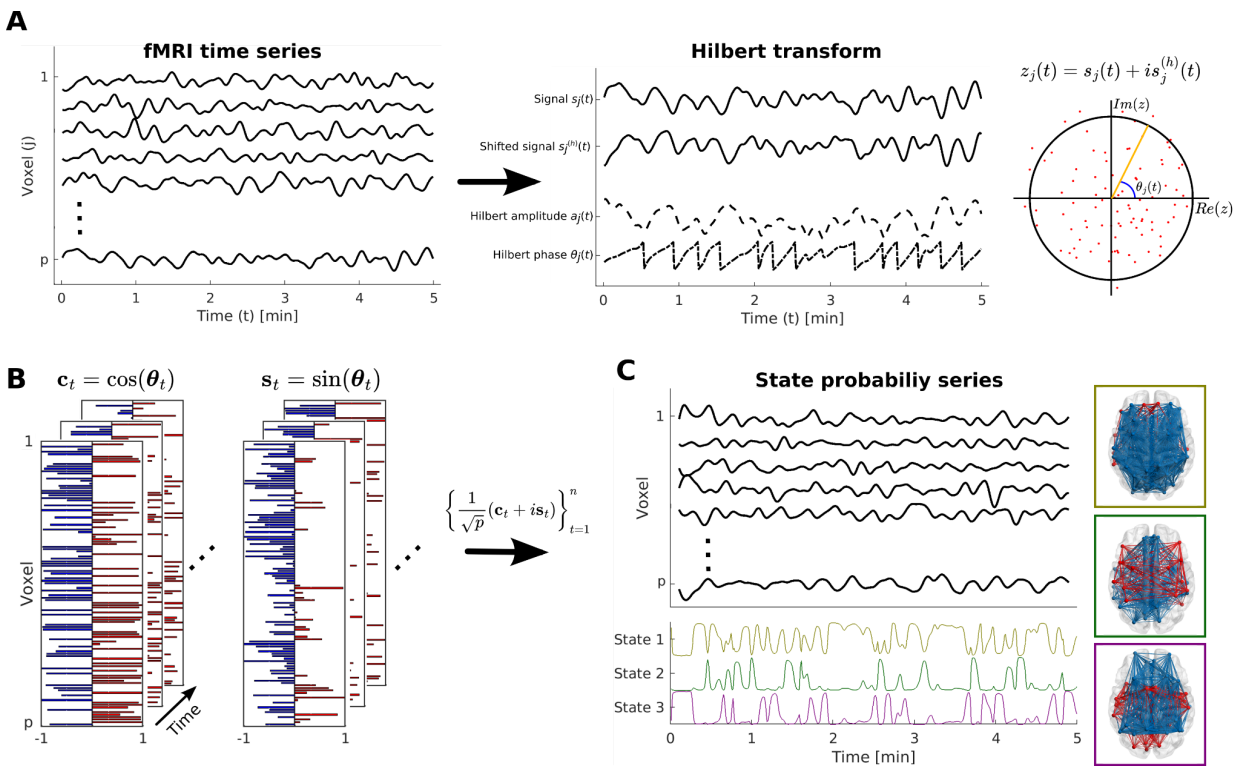

*Figure 1: Clustering complex dynamic phase coherence maps. **A**): The BOLD fMRI time-series $s_j(t)$ for each voxel or region j is decomposed into its phase $\theta_j(t)$ and amplitude $a_j(t)$ via the Hilbert transform. **B**): For each time point t, the cosine and sine phase vectors are combined into a normalized complex eigenvector. **C**): Clustering the pool of complex eigenvectors results in a set of K centroids, interpretable as phase coherence networks, and corresponding sample-to-state responsibilities.*

Importantly, the complex dynamic phase coherence map $\mathbf{A}_t$ can be factored as $\mathbf{A}_t = \mathbf{u}_t\mathbf{u}_t^{\mathsf{H}}$, where $\mathbf{u}_t = \mathrm{e}^{i\boldsymbol{\theta}_t}$ and $(\cdot)^{\mathsf{H}}$ is the complex conjugate transpose. This means that instantaneous interregional phase coherence can be wholly represented by a complex-valued vector with the same dimensionality as the input data. Since the amplitude of the Hilbert transform is discarded, the scale of $\mathbf{u}_t$ is fixed, removing a degree of freedom (Fig. 1B). Moreover, the sign of $\mathbf{u}_t$ does not affect $\mathbf{A}$, meaning that $\mathbf{u}_t$ represents an axis, like an eigenvector, and the pool of axes over all samples can be modeled using statistical distributions defined on the complex projective hyperplane, the manifold of eigenvectors, (Mardia & Jupp, 1999), such as the complex Watson (Mardia & Dryden, 1999) and complex angular central Gaussian (ACG) distributions (Kent, 1997), where the latter allows modeling anisotropic data structures. We can capture phase synchronization in its entirety by modeling complex dynamic phase coherence maps probabilistically with such appropriate statistical distributions. This approach offers four distinct advantages: 1) it retains both the real and imaginary parts of the phase information, 2) it avoids discarding

eigenvectors with non-zero eigenvalue, 3) it accounts for the manifold structure of eigenvectors, and 4) probabilistic modeling allows quantifying the data likelihood.

Utilizing such complex distributions, time-varying phase coherence can be analyzed through clustering techniques such as mixture modeling, which describes "states" of brain activity as statistical distributions and assigns state assignment probabilities for each acquired sample (Fig. 1C). Furthermore, HMMs can be employed to account for temporal dependencies in the time-series, offering smoother state transition dynamics. State activation sequences can subsequently be analyzed in terms of fractional occurrence, dwell time, or transition dynamics and be linked to known biological phenomena, such as mental disorders (Farinha et al., 2022; Figueroa et al., 2019) or drug effects (Lord et al., 2019; Olsen et al., 2022). While classical expectation-maximization (EM) algorithms are commonly used for fitting probabilistic mixture models, they can be computationally expensive for HMMs. Consequently, stochastic numerical optimization methods in PyTorch provide a more efficient alternative (Olsen et al., 2023, 2024).

Here we introduce probabilistic mixture modeling including HMMs for clustering complex dynamic phase coherence maps using the complex Watson and complex ACG distributions enabling analysis of macroscale brain phase coherence networks in their entirety, without discarding relevant information by relying on existing cosine projections of phase differences or their associated leading eigenvector. We extensively contrast our approach to existing clustering procedures for phase coherence data. For completeness, we also provide clustering methods for the rank-2 cosine phase coherence maps and their leading eigenvectors by applying appropriate statistical distributions. We compare probabilistic mixture models to non-probabilistic K-means equivalents, including least-squares K-means, which is widely used in the literature, as well as conventional GMMs of the raw time-series data. An overview of the models considered is provided in Table 1 with further information in Supplementary Table S1. We compare mixture model estimation using both EM and stochastic optimization methods, and demonstrate our approach on synthetic, resting-state, and task-based functional magnetic resonance imaging (fMRI) data from 255 unrelated subjects from the Human Connectome Project (HCP) (Van Essen et al., 2013) filtered in the frequency range $[0.03; 0.07]$ Hz. We provide an implementation of these models in our Python toolbox "Phase Coherence Mixture Modeling" (PCMM) (github.com/anders-s-olsen/PCMM).

Table 1: Overview of clustering methods for phase coherence data. Models in blue indicate methods not previously applied to functional neuroimaging data. Data removal levels: 1) The imaginary component of the complex phase coherence map is discarded to give a rank-2 cosine phase coherence

map, i.e., with two eigenvalues and eigenvectors. 2) Eigenvalues of the cosine phase coherence map are further discarded, retaining only eigenvectors. 3) Only the leading eigenvector of the cosine phase coherence map is retained, with all other components removed. See Supplementary Table S1 for further information.

	Complex	Anisotropic	Probabilistic	Data removal level
Complex dynamic phase coherence maps				
Complex angular central Gaussian mixture	✓	✓	✓	-
Complex Watson mixture	✓	X	✓	-
Complex diametrical clustering	✓	X	X	-
Rank-2 cosine phase coherence maps				
Singular Wishart mixture	X	✓	✓	1
Matrix angular central Gaussian mixture	X	✓	✓	2
Weighted Grassmann clustering	X	X	X	1
Grassmann clustering	X	X	X	2
Leading eigenvector of cosine phase coherence maps				
Angular central Gaussian mixture	X	✓	✓	3
Watson mixture	X	X	✓	3
Diametrical clustering	X	X	X	3
Least squares K-means (sign-flipped eigenvectors)	X	X	X	3

RESULTS

We first establish, using synthetic data, the need for modeling phase coherence in its entirety in order to adequately recover data with systematic phase coherence fingerprints. We subsequently evaluate our approach on both resting-state and task-based fMRI data, showcasing the need for detailed modeling of phase-coherence networks beyond simple rank-1 representations, as well as how modeling phase-coherence networks in their entirety 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

Synchronization methods convert signals to the complex domain via the Hilbert transform (Fig. 1A), wherein the amplitude is discarded and only the instantaneous phase vectors $e^{i\boldsymbol{\theta}}$ retained, where $\boldsymbol{\theta}$ are the circular multivariate phases and i is the imaginary unit. As illustrated in Fig. 2A, 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.

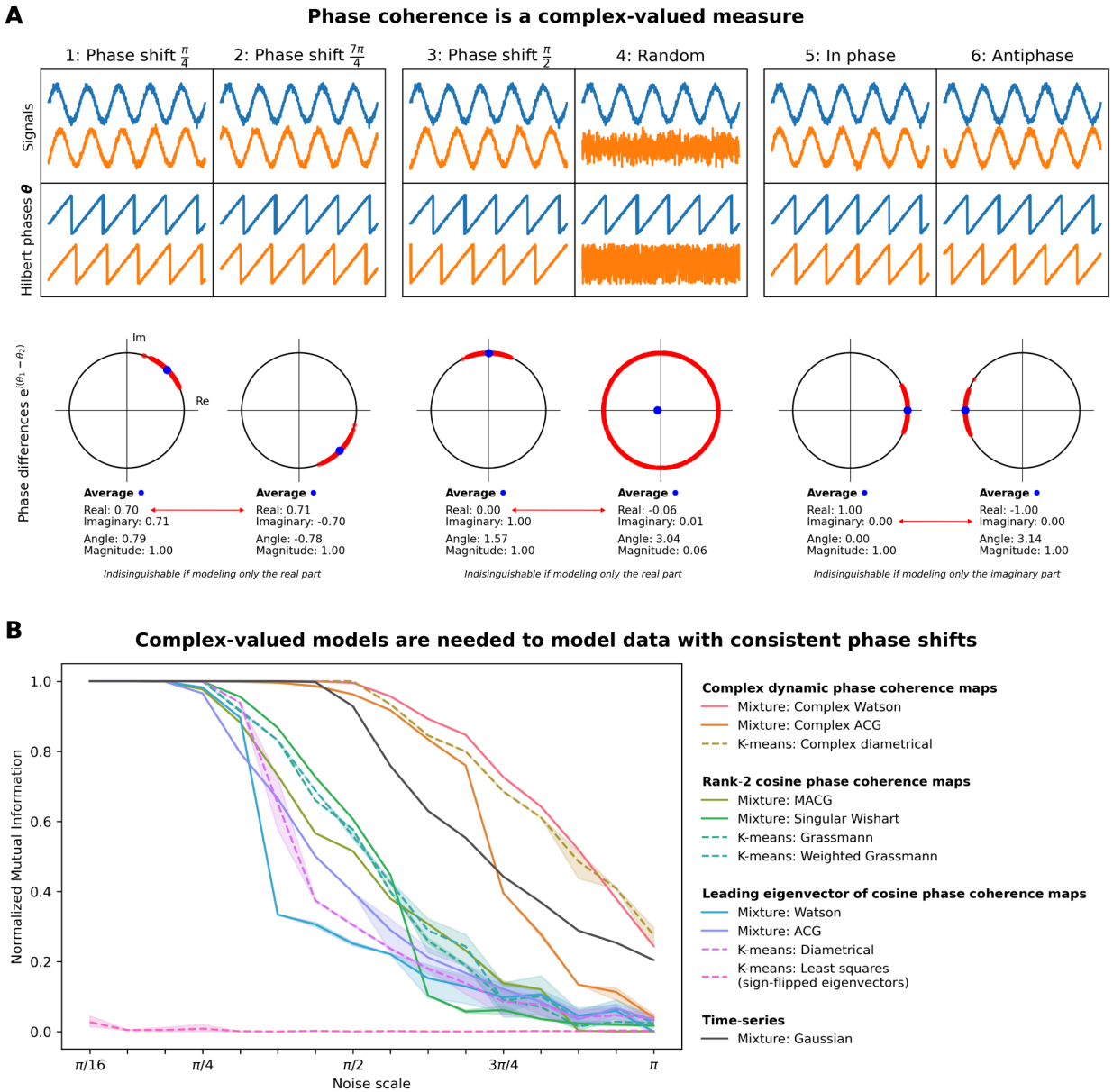

*Figure 2: Modeling phase coherence in its entirety. **A**): Two oscillatory signals can be compared by examining their phases, which indicate the position of each sample within the oscillation cycle. Modeling only cosine phase coherence restricts the centroid to its real part, making it impossible to distinguish phase shifts θ and $2\pi - \theta$ (cases 1 and 2) and 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 antiphase signals (cases 5 and 6). These issues are resolved when*

modeling both cosine and sine in a combined complex-valued representation. ***B****): Clustering of synthetic* $p = 3$*-dimensional phase coherence data with consistent phase shift (see Supplementary Fig. S1 for the synthetic generation; train and test sets were generated identically); clustering performance measured by normalized mutual information between ground-truth and predicted labels (K-means) or posterior probability (probabilistic mixtures) on the test set. Shaded areas represent the standard deviation across 10 initializations; probabilistic mixtures were initialized using K-means, which was seeded using K-means++.*

To showcase the difference between models when the data contains consistent phase shifts, we evaluated model performance on a synthetic mixture of $K = 2$ components with noisy, $p = 3$ -dimensional phase vectors (see Supplementary Fig. S1 for the synthetic data construction). The overlap between true labels and the predicted partition (K-means) or posterior (probabilistic mixtures) on a test set, measured via normalized mutual information (NMI) (Fig. 2B), showed that models for the complex dynamic phase coherence maps and time-series outperformed models for the cosine phase coherence maps, particularly in noisy environments. The complex Watson mixture and complex diametrical clustering, which both fit a rank-1 centroid, complying with the phase coherence structure in the synthetically generated data, performed the best. The complex ACG and Gaussian mixtures, which both fit a rank-3 centroid, performed slightly worse, which we attribute to the added model complexity that is not supported by the data generating mechanism. Models for the rank-2 cosine phase coherence maps only slightly outperformed models relying on the leading eigenvector, which we attribute to the clusters being easily distinguishable using only the leading eigenvectors for this data (Supplementary Fig. S1). The conventional least-squares K-means, which does not account for the manifold distribution of eigenvectors, failed to identify the correct cluster labels even for non-noisy data, likely due to the space restriction imposed by eigenvector sign-flipping (see Supplementary Fig. S1 in (Olsen et al., 2022) for an illustration).

Rank-1 components are insufficient for modeling anisotropic data

Multivariate brain data may be 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. 2A). In this case, the distribution of such phase vectors on the projective hyperplane tends to a great circle (Fig. 3). To account for such anisotropy, a model with a non-unitary component rank is needed, such as the (complex) ACG. The (complex) Watson and, by extension, other

rank-1 models such as least-squares K-means, are inherently isotropic, for which potentially infinitely many mixture components are needed to account for the anisotropy of the data.

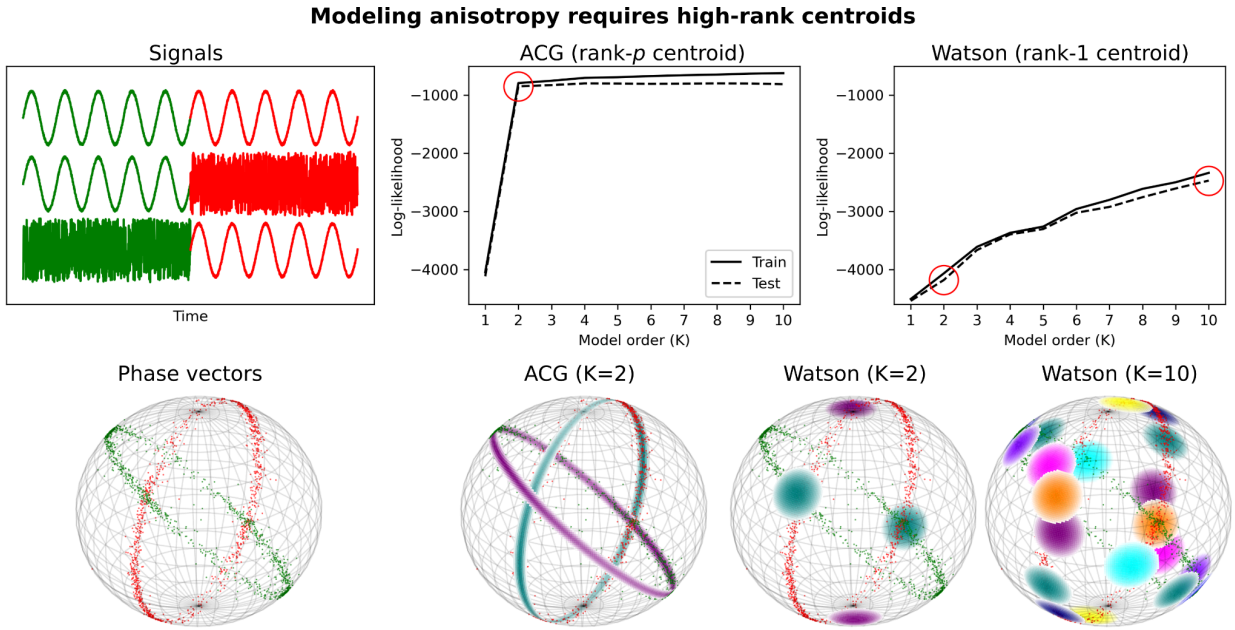

Figure 3: Synthetic phase vectors in two components that are both anisotropic. Data synthesization was repeated to create a train and test set. The Watson mixture model, which fits a rank-1 centroid for each component, fails to account for the data anisotropy and instead fits isotropic approximations along the data manifold.

Since brain areas are not fully uncorrelated, the optimal component rank is likely somewhere within $r \in \{1, \ldots, p\}$, with p being the number of brain areas. Estimation of models with a full covariance parameter $\mathbf{\Psi}$, e.g., the (complex) ACG, tends to be disrupted as more brain areas are included because estimates become singular. To address this, we introduced a low-rank-plus-diagonal reparametrization: $\mathbf{\Psi} \approx \mathbf{Z} = \mathbf{M}\mathbf{M}^{\mathsf{H}} + \mathbf{I}$, where $\mathbf{M} \in \mathbb{C}^{p \times r}$, $\mathbf{I}$ is the identity matrix, and $(\cdot)^{\mathsf{H}}$ is the complex conjugate transpose (reduced to the transpose for non-complex models) (Olsen et al., 2023).

To showcase models on real fMRI data with a ground truth label, we assessed model performance on phase-controlled resting-state fMRI data from one subject by manipulating the phase in the frequency domain into $K = 5$ segments. The data amplitude was untampered, such that only phase coherence methods should be able to untangle the true labels. Clustering partition or posterior probability overlap with true labels (Fig. 4) showed that the complex models best could extract the phase coherence clusters, albeit closely followed by models for the rank-2 cosine phase coherence, especially the matrix angular

central Gaussian (MACG) model. For the complex ACG, only a slight improvement was noted when increasing the centroid rank from 5 to 10, suggesting a potential component rank convergence point. Models for the leading eigenvector of cosine phase coherence maps and time-series performed inferiorly. Models were initialized gradually; high-rank models were initialized from lower-rank counterparts, mixture models were initialized from corresponding K-means models, and K-means initialized using K-means++, since this setup resulted in the best performance compared to random initialization (Supplementary Figs S2-S3). Little difference was observed between analytical maximization and stochastic optimization of mixture models, as well as HMMs (Supplementary Fig. S3), suggesting that, for these data, the added temporal smoothness in state-trajectories provided by HMMs does not result in improved performance. Consequently, the remainder of this report used analytical maximization of mixture models.

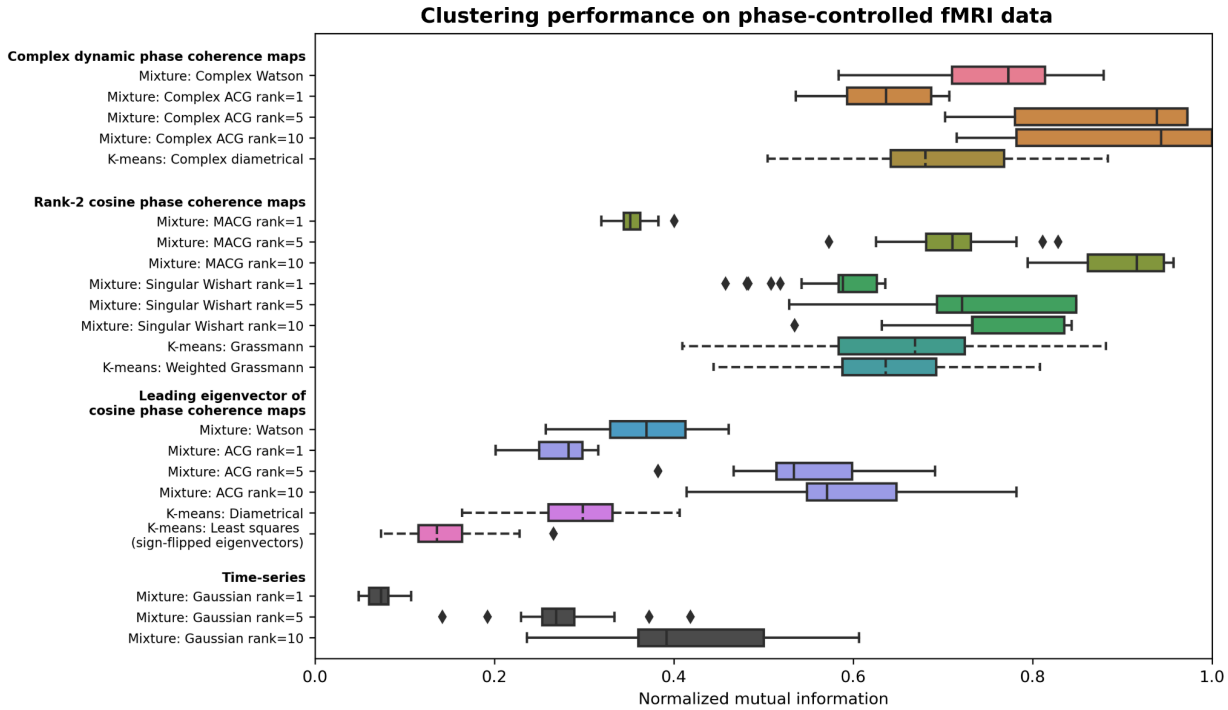

Figure 4: Clustering of phase-controlled 116-region fMRI data from a single subject. Model rank was increased gradually and rank-1 models initialized using K-means clustering seeded using K-means++. Normalized mutual information between true labels and the estimated train data partition or posterior probability from 25 initializations are shown.

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

We investigated model performance on resting-state fMRI data from 255 subjects in $p = 116$ regions via the 100-region cortical Schaefer atlas (Schaefer et al., 2018) augmented with 16 subcortical regions (Tian et al., 2020). Data for 155 subjects formed a train and within-subject test set, whereas data for the other 100 subjects formed an out-of-sample test set. Notably, global signal regression (GSR) significantly altered the average resting-state complex dynamic phase coherence map, revealing interregional anti-coherence in the angular representation (Fig. 5A). Without GSR, the vast majority of the brain was, on average in this data, coherent. 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; Vohryzek et al., 2020), while other clusters, previously referred to as ghost attractors (Vohryzek et al., 2020), carry minimal weight. Therefore, to extract coherence networks that are not trivially induced by global signals, we estimated models on data following GSR. The average angular global phase coherence after GSR was predominantly close to zero (in-phase, red) or π (anti-phase, blue), indicating that, given the current resting-state fMRI data, signals are primarily strictly in-phase or anti-phase. Similarly, the real and imaginary representations (Supplementary Fig. S4A) suggest that on average, the imaginary part holds less information than the real part.

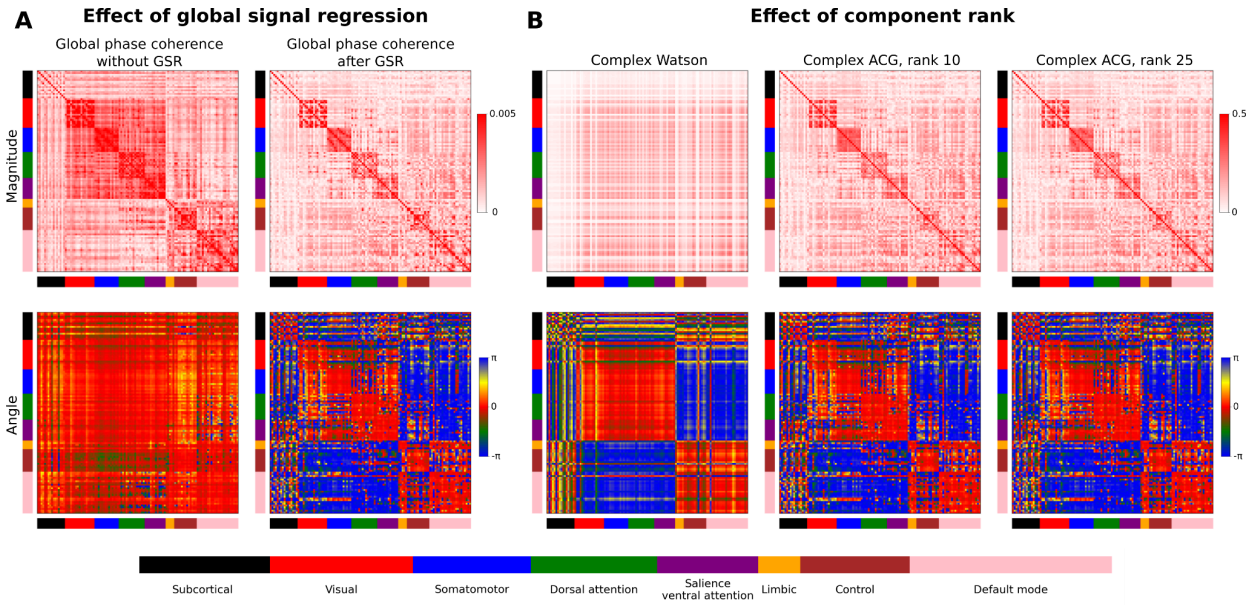

*Figure 5: Estimation of global resting-state phase coherence maps from 255 subjects in the 116-dimensional Schaefer-Tian atlas (Schaefer et al., 2018; Tian et al., 2020). **A**): The global average*

complex phase coherence matrix with and without global signal regression (GSR); the magnitude is also known as the phase-locking value. ***B****): Single-component complex Watson shown as* $\mathbf{Z} = \sqrt{\kappa}\boldsymbol{\mu}\boldsymbol{\mu}^{\mathsf{H}}$ *and complex ACG* $\mathbf{Z} = \mathbf{M}\mathbf{M}^{\mathsf{H}} + \mathbf{I}$ *estimates on data, where the global signal had been regressed out. The colored boxes indicate a partition of the 116 areas into 8 networks defined by the 7 Yeo networks and an added subcortical network.*

Considering that complex-valued models allow modeling phase coherence in its entirety and high-rank models better capture the anisotropy of brain data, we focused our attention on the complex ACG model. Train and test likelihoods for the single-component complex ACG model saturated around rank $r = 25$ (Fig. 6A). The high-rank complex ACG model was able to fit the global average phase coherence map (after GSR) well and displayed higher within-network than between-network coherence, referencing the seven Yeo networks (Yeo et al., 2011) with an added subcortical network (Fig. 5B). In contrast, the estimated single-component complex Watson 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. Visually, there was little difference between the rank-10 and rank-25 complex ACG models. The estimated single-component ACG, MACG, and singular Wishart centroids were very similar to the real part of the complex ACG estimate, suggesting that the latter supplements with the imaginary dimension without compromising information in the real dimension (Supplementary Fig. S4B-C).

Using complex ACG centroids with ranks $r = 10$ and $r = 25$ along, including the complex Watson model for comparison, we evaluated the model order (i.e., number of mixture components) based on the predictive log-likelihood (Fig. 6A). Generally, a lower component rank led to a higher model order being favorable. The best-performing complex ACG model had $K = 3$ components and a centroid rank of 25, showing marginally improved test log-likelihood over the single-component model. Generally, increasing the number of components altered likelihoods less than increasing component rank, suggesting that resting-state brain phase coherence data is not easily split into clusters that are representative across subjects and scans. Clustering stability was high for the three-component complex ACG model, and the complex Watson posterior probabilities differed drastically from those of complex ACG, again suggesting that increasing component rank significantly alters results (Supplementary Fig. S5).

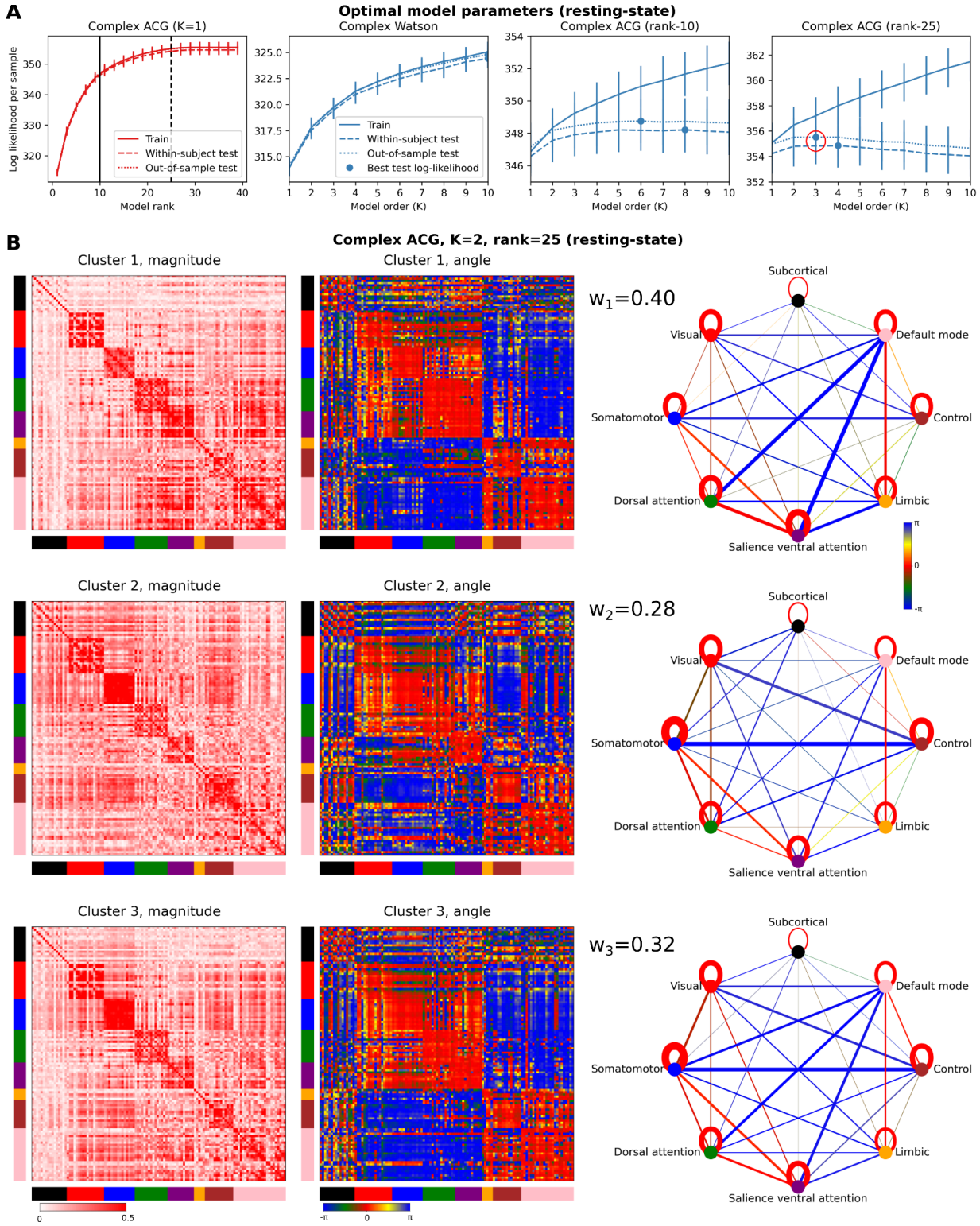

*Figure 6: Model parameter search and mixture estimate on resting-state fMRI. **A**): Component rank and model order measured using train and test log-likelihood (normalized by the number of samples) from 255 subjects from the human connectome project in $p = 116$ dimensions. The best of 10 estimates is*

shown, and bars represent variability across subjects. ***B****): Complex ACG mixture estimate with* $K = 3$ *and centroid ranks* $r = 25$*. In the graph representation, average inter-network angle is represented by color, and magnitude by edge thickness.*

Estimated resting-state phase coherence clusters primarily differ in subtle between-network coherence

The differences between the three clusters of the $K = 3$, rank-25 complex ACG solution were nuanced (Fig. 6B). All cluster centroids generally demonstrated high within-Yeo-network coherence (low interregional "angle") and, apart from the subcortical network, a high within-network magnitude. All centroids were characterized by coherence within and anti-coherence between two groups of areas: 1) subcortical, limbic, control, and default mode network (DMN), and 2) visual, somatomotor, dorsal attention, and salience. Differences between the clusters were primarily characterized by the magnitude of between-network anti-coherence: In cluster 1, the strength of the DMN anti-coherences were larger than in clusters 2 and 3, and the strength of the control network anti-coherences were strongest in cluster 2; cluster 3 served as a midpoint between clusters 1 and 2. The high similarity between cluster centroids suggests that the brain does not discretely "switch" from one subnetwork to another but rather subtly adjusts the nuances of, mostly, between-network coherence. Centroids were split evenly across subjects, suggesting a negligible effect of intersubject variability.

Mixture modeling can provide meaningful partitions into functional states

Given that a three-component complex ACG model only marginally outperformed the single-component equivalent on resting-state data and centroids were highly similar, we investigated whether a mixture model could extract meaningful components from task fMRI data. Specifically, we concatenated the two HCP tasks "motor", comprising visual cues and movement of the left/right fingers, toes, or tongue, and "social", comprising videos of moving shapes and subjective evaluation of whether the shapes interacted socially. The information overlap between scan type and posterior probabilities or predicted labels for models with $K = 2$ components showed that complex ACG, MACG, and singular Wishart models with rank-25 centroids performed well (Fig. 7), albeit with high intersubject variability. Clustering stability and similarity was also high between these models (Supplementary Fig. S6). In contrast, unit-rank models and models for LEiDA or the time-series performed inferiorly (Fig. 7).

Notably, when the global signal was not regressed out, model performance decreased for all models (Supplementary Fig. S7).

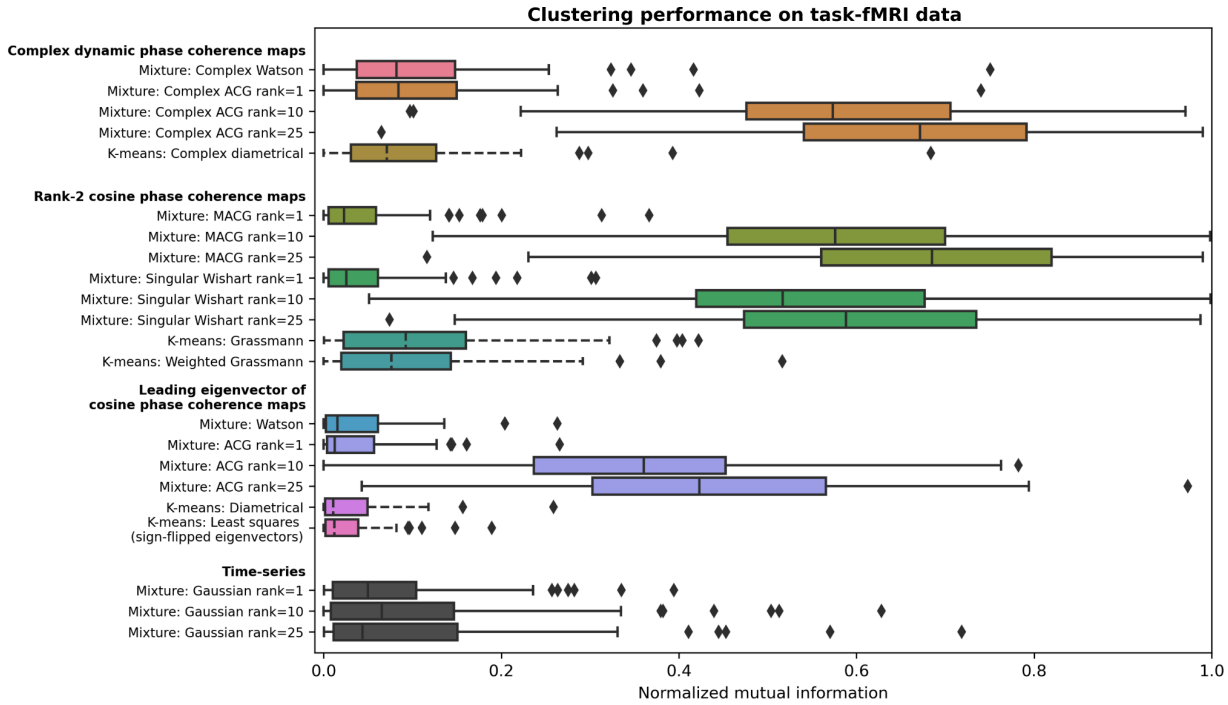

Figure 7: Clustering of 116-region task-fMRI data from 255 subjects after global signal regression, comprising "motor" and "social" tasks. Normalized mutual information between true scan type and estimated out-of-sample test data partition across 100 subjects is shown.

For the complex ACG distribution, a centroid rank of 25 was again deemed sufficient and a dual-component model resulted in the best test log-likelihood (Fig. 8A). Importantly, on the out-of-sample test set, the posterior probability of cluster 1 for the complex ACG model largely overlapped with the "motor" task, and cluster 2 with the "social" task, and similarly for the MACG model. The Gaussian mixture model on the time-series data, and the widely used least-squares K-means on the sign-flipped leading eigenvector of cosine phase coherence maps did not result in data partitions overlapping well with task scans. These results suggest that modest-rank mixture modeling of phase coherence maps have the potential to provide meaningful data partitions. The two inferred clusters primarily differed in between-network visual connectivity (Fig. 8B). Interestingly, both clusters exhibited consistent phase shifts: cluster 1 (the motor task cluster) had a consistent phase shift of $\pi/2$ between the visual and dorsal attention networks, while cluster 2 (the social task cluster) had a consistent phase shift of $\pi/2$ between

the visual and control networks, connections that would not have been captured if only the cosine phase coherence had been modeled.

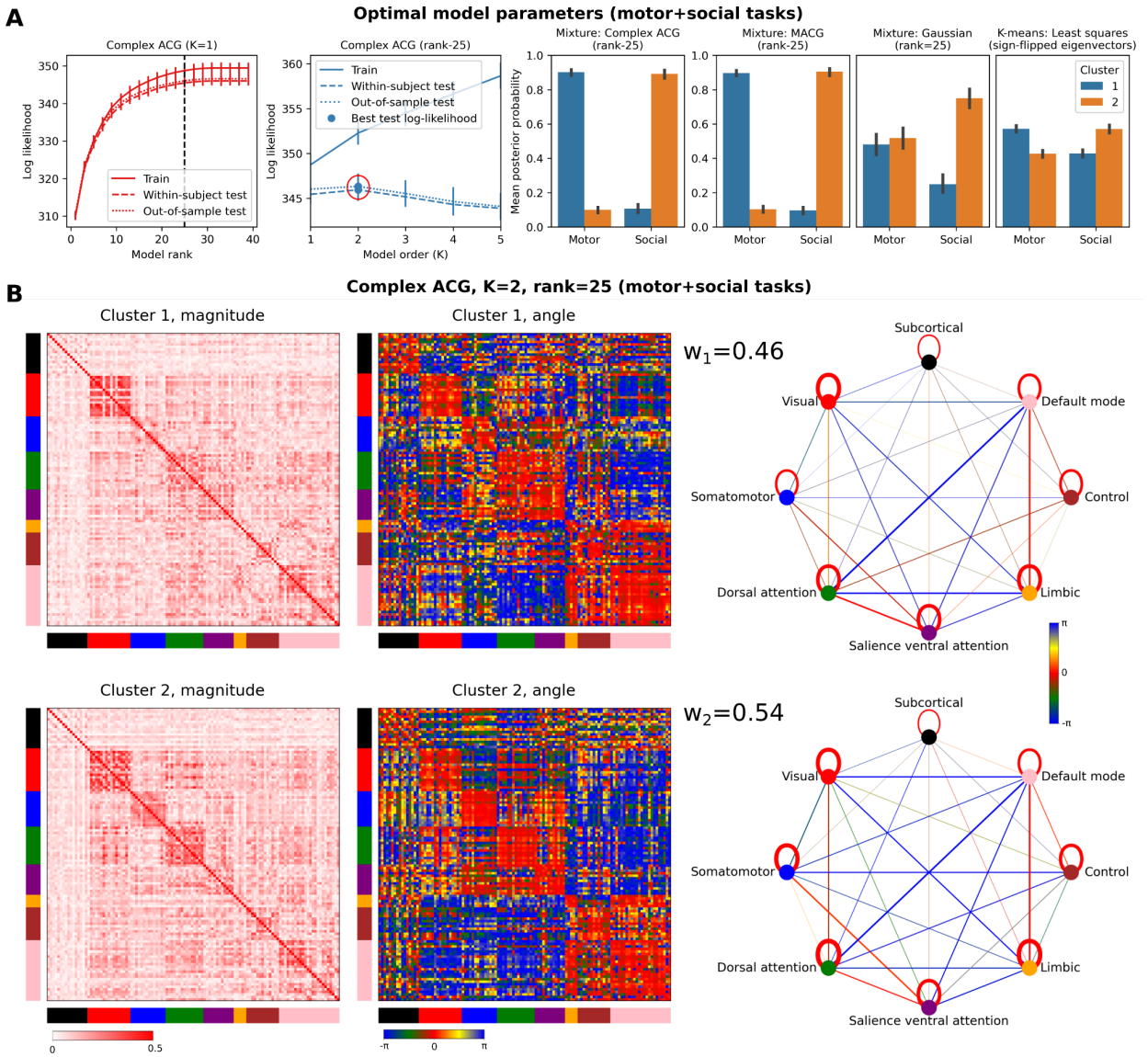

Figure 8: Model parameter search and mixture estimate on task data, where motor and social task-fMRI scans had been concatenated. ***A****): Centroid rank and model order measured using train and test log-likelihood across 255 subjects, as well as the mean posterior probability of each cluster during each scan for the best of 10 estimates in terms of train set likelihood, where the variation here indicates intersubject variability.* ***B****): Complex ACG mixture estimate with* $K = 2$ *and centroid ranks* $r = 25$*. In the graph representation, inter-network angle is represented by color, and magnitude by edge thickness.*

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 models—can result in loss of cognitively or behaviorally relevant information. Our findings show that statistical distributions that respect the underlying geometry of the manifold of phase coherence data more accurately capture components than traditional methods, including Gaussian mixture modeling of time-series or least-squares K-means modeling of leading eigenvectors.

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 centroids (Fig. 2). In resting-state fMRI data, our analysis suggests that the added complexity does not alter the real part of the complex centroids (Supplementary Fig. S4B-C). The estimated dual-component complex ACG model for task fMRI data also suggested valuable information in the imaginary component (Fig. 8B). Thus, constraining 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 representation of cosine phase coherence over only the leading eigenvector (i.e., LEiDA) (Figs. 4 and 7), 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 while the leading eigenvector represents broad connectivity, the trailing eigenvector may reveal localized connectivity patterns (de Alteriis et al., 2024). Our results reinforce that reducing dimensionality by modeling only the leading eigenvector can be detrimental and even unnecessary because both eigenvectors can be modeled using suitable statistical distributions such as the MACG or singular Wishart that have the same number of parameters as those for the leading eigenvector ACG model.

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, whereas matrices with orthonormal columns (e.g., the rank-2 cosine

phase coherence maps) reside on the Grassmann manifold (Mardia & Jupp, 1999). 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). The choice of distance measure affects interpretability; 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 here. A Wishart K-means approach would likely lead to component over-parameterization in high dimensions, as brain connectivity centroids often exhibit intercorrelations, a point illustrated by rank-25 representations being sufficient in 116-dimensional fMRI data (Fig. 6A and 8A). Models that align with the correct manifold and allow for centroids of specified rank are thus more suitable. In contrast, models that do not account for data covariance are insufficient to model even simple cases (Fig. 3).

Our methodological contributions are numerous. We introduced, for the first time, EM update rules for the ACG, MACG and singular Wishart mixture models including low-rank-plus-diagonal approximations as well as weighted Grassmann and complex diametrical clustering for K-means on the symmetric positive semidefinite matrix manifold and complex projective hyperplane, respectively, all models applicable also outside the realm of functional neuroimaging. While EM procedures for the complex Watson (Tran Vu & Haeb-Umbach, 2010) and complex ACG distributions (Ito et al., 2016) existed already, these have not, to our knowledge, been used to model functional neuroimaging data. While our analyses were limited to low-dimensional data, the low-rank-plus-diagonal approximation combined with low-rank initialization of high-rank models 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 (Fig. 5A) and models failed to distinguish between the two task fMRI scan types when the global signal had not been regressed out (Supplementary Fig. S7). Previous results using K-means approaches showed that a “global” component was by far the most prevalent one, with the more informational states being less pronounced (Olsen et al., 2022; Vohryzek et al., 2020). We believe that applying GSR, which removes more trivial coherence structures, will directly lead to centroids having more contrast, characterizing more detailed and relevant connectivity information.

Our study is not without its 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. 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 small restriction on the $[0.01, 0.1]Hz$ range that the BOLD signal is generally accepted to fall within. Future studies could further investigate the effect of empirical mode decomposition to extract frequency-specific signal "modes" prior to the Hilbert transformation (Honari & Lindquist, 2022; Rehman & Mandic, 2010). 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). 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.

Conclusion

Modeling time-varying spatiotemporal coherence of functional brain imaging signals can provide meaningful insights about neural systems underlying cognition and illness. Appropriate models are necessary for the analysis of phase coherence between multivariate signals. 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 as much of the data variance as possible. Our Phase Coherence Mixture Modeling (PCMM) toolbox is readily available for general-purpose phase coherence modeling: github.com/anders-s-olsen/PCMM.

METHODS

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

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

The Hilbert transform represents a phase shift by $\pi/2$ radians to every frequency in the signal. While the original signal $s_j(t)$ is oscillating along the time axis, the analytic signal $z_j(t)$ is revolving around the unit circle in the complex plane (Fig. 1A). 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

$$\theta_{t,j} = \tan^{-1}\left(\frac{s_j^{(h)}(t)}{s_j(t)}\right),$$

also called instantaneous phase, and a distance from the origin, known as instantaneous amplitude. The phase captures solely the oscillatory information in the signal while signal strength, including potential artifactual spikes, is disregarded as 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. Furthermore, the naming convention “instantaneous” implies that phases can be calculated using information from only a single time point. However, in reality the Hilbert transform suffers from the same estimation requirements for a sufficient number of samples as the Fourier transform to reliably estimate the phase spectrum.

Each sample of the analytic signal is a complex number $ae^{i\theta}$, where the radius is typically ignored when 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 (complex) 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, $\mathbf{u}_t = \frac{1}{\sqrt{p}}\mathbf{u}_t$. 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 coherences can be compactly represented using a sign-invariant vector residing on the unit sphere.

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 $\mathbf{s}_t = \sin\boldsymbol{\theta}_t = \mathrm{Im}(\mathbf{u}_t)$ (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

Consider a real symmetric matrix $\mathbf{A}^{(r)} \in \mathbb{R}^{p\times p}$ with corresponding eigendecomposition $\mathbf{A}^{(r)} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^\top$ with eigenvectors $\mathbf{U}$ and positive eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p$ in the diagonal of $\boldsymbol{\Lambda}$. If $\mathbf{A}^{(r)}$ is of rank q, only q eigenvalues will be non-zero, and if $q < p$, $\mathbf{A}^{(r)}$ is said to be positive semidefinite. The reduced subspace $\mathbf{U}^{(q)}$, or more generally, any orthonormal matrix, belongs to the Grassmann manifold $\mathrm{Gr}\,(p,q) = \{\mathbf{U}|\mathbf{U}^\top\mathbf{U} = \mathbf{I}_q\}$ (Mardia & Jupp, 1999). Individual eigenvectors $\{\mathbf{u}\}_1^q$ represent axes, meaning they are sign-invariant and normalized to unit norm, and thus belong to the projective hyperplane $\mathbb{P}^{p-1} = \{\mathbf{u}|\mathbf{u} \in \mathbb{R}^p, \mathbf{u}^\top\mathbf{u} = 1, -\mathbf{u} = \mathbf{u}\}$, also known as the antipodally symmetric unit hypersphere (Mardia & Jupp, 1999). When considering a complex hermitian-symmetric matrix $\mathbf{A} \in \mathbb{C}^{p\times p}$, the eigenvectors are also complex. Here we only deal with rank-1 complex-valued matrices $\mathbf{A}$, and thus we only consider the complex projective hyperplane $\mathbb{CP}^{p-1} = \{\mathbf{u}|\mathbf{u} \in \mathbb{C}^p, \mathbf{u}^\mathsf{H}\mathbf{u} = 1, -\mathbf{u} = \mathbf{u}\}$, bypassing the need for the complex Grassmann or complex positive semidefinite matrix manifolds. Below we present statistical distributions on the above manifolds; the presented methods and their characteristics are summarized in Table 1 and Supplementary Table S1.

Watson distribution

The multivariate Watson distribution (Watson, 1965) is parameterized by a mean direction $\boldsymbol{\mu} \in \mathbb{P}^{p-1}$ and a scalar precision parameter $\kappa \in \mathbb{R}$, and its probability density function is given by

$$f_{\mathrm{W}}\left(\mathbf{u}|\boldsymbol{\mu},\kappa\right) = c_p(\kappa)e^{\kappa\left(\boldsymbol{\mu}^\top\mathbf{u}\right)^2},\ \mathbf{u} \in \mathbb{P}^{p-1}$$

with normalization constant

$$c_p(\kappa) = \frac{\Gamma(p/2)}{2\pi^{p/2}M\left(\frac{1}{2},\frac{p}{2},\kappa\right)}.$$

Here, $\Gamma(\cdot)$ is the Gamma function and $M(a,c,\kappa)$ is Kummer's confluent hypergeometric function (Kummer, 1836)

$$M(a,c,\kappa) = \sum_{l\geq 0} \frac{a^{\bar{l}}\kappa^l}{c^{\bar{l}}l!},$$

where $a^{\bar{l}} = a(a+1)\cdots(a+l-1), l \geq 1$, $a^{\bar{0}} = 1$ is the rising factorial (or Pochammer symbol). Kummer's confluent hypergeometric function is an infinite series that is non-trivial to evaluate (Gautschi, 1977; Gil et al., 2007). When κ is positive, the density concentrates isotropically around the mean direction $\boldsymbol{\mu}$. When κ is negative, the density concentrates around the great circle orthogonal to $\boldsymbol{\mu}$.

Maximum likelihood estimation involves the log-likelihood function (Sra & Karp, 2013)

$$\mathcal{L}_{\mathrm{W}}\left(\boldsymbol{\mu},\kappa|\mathbf{u}_1,\ldots,\mathbf{u}_n\right) = n\left(\kappa\boldsymbol{\mu}^\top\mathbf{G}\boldsymbol{\mu} - \ln M\left(\tfrac{1}{2},\tfrac{p}{2},\kappa\right) + \delta\right),$$

where $\mathbf{G} = \frac{1}{n}\sum_{t=1}^{n}\mathbf{u}_t\mathbf{u}_t^\top$ is the sample scatter matrix and δ is a constant term unrelated to parameter estimation. Maximizing this function with respect to $\boldsymbol{\mu}$ yields the parameter estimate

$$\hat{\boldsymbol{\mu}} = \mathbf{g}_1 \quad \text{if} \quad \hat{\kappa} > 0, \quad \hat{\boldsymbol{\mu}} = \mathbf{g}_p \quad \text{if} \quad \hat{\kappa} < 0,$$

where $\mathbf{g}_1$ and $\mathbf{g}_p$ are the leading and trailing eigenvectors of $\mathbf{G}$, respectively. Solving for the precision parameter κ involves the non-linear optimization of the function

$$\frac{M'\left(\frac{1}{2},\frac{p}{2},\kappa\right)}{M\left(\frac{1}{2},\frac{p}{2},\kappa\right)} - \boldsymbol{\mu}^\top\mathbf{G}\boldsymbol{\mu} = 0.$$

The derivative of Kummer's function is given by the recurrence relation $M'(a,c,z) = \frac{a}{c}M(a+1,c+1,z)$ (https://dlmf.nist.gov/13.3). Several studies have made optimization of the precision parameter easier using bounds (Bijral et al., 2007; Mardia & Dryden, 1999; Sra & Karp, 2013). In this work, we implemented the lower, middle, and upper bounds presented by (Sra & Karp, 2013). We define $a = \frac{1}{2}$, $b = \hat{\boldsymbol{\mu}}^\top\mathbf{G}\hat{\boldsymbol{\mu}}$, and $c = \frac{p}{2}$. The bounds are

$$L(b) = \frac{bc-a}{b(1-b)}\left(1 + \frac{1-b}{c-a}\right),$$

$$B(b) = \frac{bc-a}{2b(1-b)}\left(1 + \sqrt{1 + \frac{4(c+1)b(1-b)}{a(c-a)}}\right),$$

$$U(b) = \frac{bc-a}{b(1-b)}\left(1 + \frac{b}{a}\right).$$

For $a/c < b < 1$ the solution is within the bounds $L(b) < \kappa(b) < B(b)$. For $0 < b < a/c$ the solution is within $B(b) < \kappa(b) < U(b)$. If $b = a/c$, then $\kappa(b) = 0$.

Complex Watson distribution

The multivariate complex Watson distribution (Mardia & Dryden, 1999) is parameterized by a complex unit norm mean direction $\boldsymbol{\mu} \in \mathbb{CP}^{p-1}$ and a real precision parameter $\kappa \in \mathbb{R}$ and is characterized by the density

$$f_{\mathrm{CW}}\left(\mathbf{u}|\boldsymbol{\mu}, \kappa\right) = c_p(\kappa) e^{\kappa \left|\boldsymbol{\mu}^{\mathsf{H}}\mathbf{u}\right|^2}, \ \mathbf{u} \in \mathbb{CP}^{p-1},$$

with normalization constant

$$c_p(\kappa) = \frac{\Gamma(p)}{2\pi^p M(1,p,\kappa)}.$$

The maximum likelihood estimator for the complex Watson distributions is identical to the one presented for the real Watson distribution when replacing the vector transpose with complex conjugate transpose and setting $a = 1$ and $c = p$.

Angular central Gaussian distribution

While the Watson distribution models equal variance in all dimensions, a more expressive alternative that allows modeling full covariance structures is the angular central Gaussian (ACG) distribution with density (Tyler, 1987)

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

Here, $\boldsymbol{\Psi} \in \mathbb{R}^{p\times p}$ is a symmetric positive definite matrix identifiable up to a positive scale factor ($\boldsymbol{\Psi}$ may be scaled such that $\mathrm{tr}(\boldsymbol{\Psi}) = p$ 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., $\mathbf{x} \sim N(\mathbf{0}, \boldsymbol{\Psi}), \quad \frac{\mathbf{x}}{||\mathbf{x}||} \sim \mathrm{ACG}(\boldsymbol{\Psi})$ (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 $\boldsymbol{\Psi}$ involves solving the log-likelihood function

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

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 $\mathrm{tr}(\hat{\boldsymbol{\Psi}}) = p$ (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.

Matrix angular central Gaussian distribution

The matrix angular central Gaussian (MACG) distribution is the ACG equivalent on the Grassmann manifold. Its density function is characterized by (Chikuse, 1990)

$$f_{\mathrm{MACG}}(\mathbf{U}|\mathbf{\Psi}) = |\mathbf{\Psi}|^{-\frac{q}{2}}|\mathbf{U}^\top\mathbf{\Psi}^{-1}\mathbf{U}|^{-\frac{p}{2}},\ \mathbf{U} \in \mathrm{Gr}(p,q).$$

Similar to the ACG, $\mathbf{\Psi} \in \mathbb{R}^{p\times p}$ is a symmetric positive-definite matrix identifiable up to a positive scale factor, and is usually normalized to $\mathrm{tr}(\mathbf{\Psi}) = p$. Samples from a MACG distribution can be generated by first sampling from a centered matrix-variate normal distribution with uncorrelated columns, $\mathbf{X} \sim N(\mathbf{0}, \mathbf{\Psi}, \mathbf{I}_q)$ and then computing the corresponding orthonormal basis $\mathbf{X}\left(\mathbf{X}^\top\mathbf{X}\right)^{-1/2} \sim \mathrm{MACG}(\mathbf{\Psi})$ (You, 2022).

Maximum likelihood estimation for the MACG distribution involves optimizing the log-likelihood function

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

The solution is given by the following iterative fixed point update

$$\hat{\mathbf{\Psi}}_{i+1} = \frac{p}{qn}\sum_{t=1}^{n}\mathbf{U}_t\left(\mathbf{U}_t^\top\hat{\mathbf{\Psi}}_i^{-1}\mathbf{U}_t\right)^{-1}\mathbf{U}_t^\top,$$

followed by trace-normalization $\hat{\mathbf{\Psi}} = p\frac{\hat{\mathbf{\Psi}}}{\mathrm{tr}(\hat{\mathbf{\Psi}})}$, iterating until convergence (see Supplementary Information for derivations).

Singular Wishart distribution

The centered singular Wishart distribution for positive semidefinite symmetric matrices (Khatri, 1970; Srivastava & Khatri, 1979; Uhlig, 1994) has a density characterized by (Bodnar & Okhrin, 2008; Srivastava, 2003)

$$f_{\mathrm{SW}}\left(\mathbf{S}|\mathbf{\Psi}\right) \frac{\pi^{q(q-p)/2} 2^{-pq/2}}{\Gamma_q\left(\frac{q}{2}\right)|\Psi|^{q/2}} |\mathbf{S}_{11}|^{(q-p-1)/2} \operatorname{etr}\left(-\frac{1}{2}\mathbf{\Psi}^{-1}\mathbf{S}\right),$$

where $\Gamma_q(d) = \pi^{q(q-1)/4} \prod_{j=1}^{q} \Gamma(d-(j-1)/2)$. Here, $\mathbf{S} = \mathbf{X}\mathbf{X}^\top$ is a singular (non-full-rank) covariance matrix of samples drawn from a centered matrix-variate normal distribution with uncorrelated columns $\mathbf{X} \sim N\left(\mathbf{0}, \mathbf{\Psi}, \mathbf{I}_q\right)$, and $\mathbf{\Psi} \in \mathbb{R}^{p\times p}$ is a symmetric positive definite matrix. Given the eigendecomposition $\mathbf{S} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top$, the functionally independent elements (given that $\mathbf{S}$ is rank-q) are $\mathbf{S}_{11} = \mathbf{U}_1\mathbf{\Lambda}\mathbf{U}_1^\top$, where $\mathbf{U} = (\mathbf{U}_1, \mathbf{U}_2)$ is a partition of the subspace into a $q \times q$ matrix $\mathbf{U}_1$ and a $p-q \times q$ matrix $\mathbf{U}_2$.

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

$$\mathcal{L}_{\mathrm{SW}}\left(\mathbf{\Psi}|\mathbf{S}_1, \ldots, \mathbf{S}_n\right) = -\frac{nq}{2}\ln\left(\mathbf{\Psi}\right) - \frac{1}{2}\sum_{t=1}^{n} \operatorname{tr}\left(\mathbf{\Psi}^{-1}\mathbf{S}_t\right) + \delta,$$

where δ is a term unrelated to parameter optimization. The maximization with respect to $\mathbf{\Psi}$ yields the estimate

$$\hat{\mathbf{\Psi}} = \frac{1}{qn}\sum_{t=1}^{n} \mathbf{S}_t = \frac{1}{qn}\sum_{t=1}^{n} \left(\mathbf{U}_t\mathbf{\Lambda}_t^{1/2}\right)\left(\mathbf{U}_t\mathbf{\Lambda}_t^{1/2}\right)^\top,$$

where $\mathbf{U}_t$ and $\mathbf{\Lambda}_t$ are the eigenvectors and eigenvalues of $\mathbf{S}_t$, respectively (see Supplementary Information for derivations).

Probabilistic mixture modeling

Mixture modeling involves computing centroids for K components, each corresponding to a unique probability distribution mixed with component weights $\mathbf{w} = \{w_k\}_{k=1}^{K}$, $\sum_k w_k = 1$. Previous studies have explored mixtures of Watson distributions using various estimation frameworks (Bijral et al., 2007; Hasnat et al., 2014; Olsen et al., 2023; Sra & Karp, 2013; Taghia & Leijon, 2016) as well as mixtures of complex Watson distributions (Tran Vu & Haeb-Umbach, 2010). Mixtures of ACG distributions have only been approached through a numerical optimization framework (Olsen et al., 2023), while mixtures of complex ACG distributions have been covered using EM inference (Ito et al., 2016). Mixtures of MACG and Singular Wishart distributions have not to our knowledge been explored previously.

The mixture density for each observation $\mathbf{y}_t \in \{\mathbf{u}_t, \mathbf{U}_t, \mathbf{U}_t\boldsymbol{\Lambda}_t^{1/2}\}$ is defined as

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

where $\boldsymbol{\phi}_k$ are the parameters of the kth component, i.e., $\boldsymbol{\phi}_k = (\boldsymbol{\mu}_k, \kappa_k)$ for Watson and complex Watson, or $\boldsymbol{\phi}_k = \boldsymbol{\Psi}_k$ for ACG, complex ACG, MACG, Singular Wishart, and zero-mean Gaussian.

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 $\mathcal{Y} = \{\mathbf{y}_t\}_{t=1}^{n}$

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

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 in conjunction with automatic differentiation. Although the E-step is not required for the 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 for the ACG, MACG, and Singular Wishart distributions, 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 and MACG. Similar derivations for the Watson distribution can be found elsewhere (Sra & Karp, 2013; Tran Vu & Haeb-Umbach, 2010). The cluster weights are for all the EM-based estimation procedures updated analytically as

$$\hat{w}_k = \frac{1}{n}\sum_{t=1}^{n} \beta_{kt}.$$

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 $\alpha_{1,k}^{(s)} = w_k f(\mathbf{y}_1^{(s)} | \boldsymbol{\phi}_k)$. The log-likelihood to be maximized numerically is

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

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 parameters for the ACG, MACG, and Singular Wishart models can be challenging due to the 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 ACG and MACG distributions corresponds to the trace-normalization, i.e., $\gamma = \frac{p}{\mathrm{tr}\left(\mathbf{M}\mathbf{M}^{\mathsf{H}} + \mathbf{I}\right)} = \frac{1}{o/p+1}$ where $o = ||\mathbf{M}||_F^2$. For the Singular Wishart and zero-mean Gaussian distributions, which are 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 Singular Wishart and Gaussian) as detailed in the Supplementary Information. 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, MACG, Singular Wishart, 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 $\mathrm{LR} = 0.1$, which decreased to $\mathrm{LR} = 0.01$ upon reaching a plateau, 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: Watson components were constrained to unit-norm mean directions $\boldsymbol{\mu} = \frac{\tilde{\boldsymbol{\mu}}}{\|\tilde{\boldsymbol{\mu}}\|}$ while $\tilde{\boldsymbol{\mu}}$ was allowed to vary freely. 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 Singular Wishart and Gaussian distributions to ensure non-negativity. The transition probability matrix was treated as a row-wise probability mass function implemented using the softmax function.

For the Watson distribution, the Kummer function was evaluated until an absolute tolerance of 10^{-10} and for the EM algorithm, estimation of the precision parameter κ was performed using scipy's bounded scalar optimization with an absolute tolerance of 10^{-10} which we verified worked as expected (Supplementary Fig. S8). Analytical fixed-point updates for (complex) ACG, MACG, Singular Wishart, and Gaussian distributions were evaluated until an absolute tolerance of 10^{-10}. Mixture models, both using analytical and numerical optimization, were evaluated until a *relative* tolerance $\mathrm{tol} = 10^{-10}$. While the EM algorithm typically guarantees monotone convergence, this is not the case for the low-rank-plus-diagonal approximations, as well as the Watson distribution, since these rely on iterative optimization procedures in the M-step. Thus, estimation stopped when, for the last 10 updates, the difference between the maximum and second-highest log-likelihood relative to the maximum was less than the tolerance.

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, 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 Information.

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. S2). Probabilistic mixture models were initialized using centroids derived from K-means. For ACG, complex ACG, MACG, Singular Wishart, 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

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

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 $\boldsymbol{\pi}$ and transition probability matrices $\mathbf{T}$ for HMMs were initialized based on the empirical transition probability matrix from the initialization partition. For the singular Wishart and Normal distributions, γ was initialized to the same value as ϵ.

We also investigated a scenario where high-rank ACG, MACG, or Singular Wishart 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. S3).

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. We also evaluated NMI between consecutive runs of the same model, i.e., $\{\mathrm{NMI}(\mathcal{M}_1, \mathcal{M}_2), \mathrm{NMI}(\mathcal{M}_2, \mathcal{M}_3), \ldots, \mathrm{NMI}(\mathcal{M}_M, \mathcal{M}_1)\}$, given M runs, as well as between runs of different models.

Code 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

Experimental data

We synthesized a dataset of $K = 2$ true clusters of dimensionality $p = 3$, each cluster with $n_k = 1000$ points (Fig. 2B). Data points were constructed by adding random uniform noise, scaled by a factor σ, to phase vector $\boldsymbol{\theta}_1 = \begin{bmatrix} 0 & \pi/2 & \pi \end{bmatrix}^\top$ for cluster 1 and $\boldsymbol{\theta}_2 = \begin{bmatrix} 0 & \pi & \pi/2 \end{bmatrix}^\top$ for cluster 2. Subsequently, we constructed the cosine and complex phase coherence maps and their eigendecomposition (Supplementary Fig. S1).

We used functional magnetic resonance imaging (fMRI) data from 255 unrelated healthy subjects (same subjects as in (Pang et al., 2023)) from the openly available Human Connectome Project (Elam et

al., 2021; Van Essen et al., 2013). Imaging protocols may be found elsewhere (Barch et al., 2013; Smith et al., 2013). We used the MSM-All preprocessed version of the resting-state data (Glasser et al., 2013). Additionally, we bandpass-filtered the data in the range $[0.03, 0.07]Hz$ and spatially downsampled the data to $p = 116$ cortical and subcortical regions using the Schaefer and Tian atlases (Schaefer et al., 2018; Tian et al., 2020). Of the 255 subjects, a randomly selected 155 individuals were used for train and 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 the within-subject test set. The "motor" and "social" tasks were each split into two blocks, randomly assigned to either the train or within-subject test set.

To evaluate performance upon increasing model complexity, we used phase-controlled 116-dimensional 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 200 time points and the first 23 regions, cluster 2 corresponded to time points 201 to 400 and regions 24-46 and so on. Then we computed, for all regions, and temporal windows, the Fourier spectrum and if the region j 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. Fourier transform edge effects were avoided by prepending each signal with the first 10% of itself, and appending the signal with a flipped version of the last 10% of itself.

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* (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

Bijral, A. S., Breitenbach, M., & Grudic, G. (2007). Mixture of Watson distributions: A generative model for hyperspherical embeddings. *Journal of Machine Learning Research*, *2*, 35–42.

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

Bodnar, T., & Okhrin, Y. (2008). Properties of the singular, inverse and generalized inverse partitioned Wishart distributions. *Journal of Multivariate Analysis*, *99*(10), 2389–2405. https://doi.org/10.1016/j.jmva.2008.02.024

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

Calhoun, V. D., Miller, R., Pearlson, G., & Adali, T. (2014). The Chronnectome: Time-Varying Connectivity Networks as the Next Frontier in fMRI Data Discovery. *Neuron*, *84*(2), 262–274. https://doi.org/10.1016/j.neuron.2014.10.015

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

Chang, C., & Glover, G. H. (2010). Time-frequency dynamics of resting-state brain connectivity measured with fMRI. *NeuroImage*, *50*(1), 81–98. https://doi.org/10.1016/j.neuroimage.2009.12.011

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

Farinha, M., Amado, C., Morgado, P., & Cabral, J. (2022). Increased Excursions to Functional Networks in Schizophrenia in the Absence of Task. *Frontiers in Neuroscience*, *16*, 1–12. https://doi.org/10.3389/fnins.2022.821179

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

Gautschi, W. (1977). Anomalous Convergence of a Continued Fraction for Ratios of Kummer Functions. *Mathematics of Computation*, *31*(140), 994–999. https://doi.org/10.2307/2006129

Gil, A., Segura, J., & Temme, N. M. (2007). *Numerical Methods for Special Functions*. Society for Industrial and Applied Mathematics. https://doi.org/10.1137/1.9780898717822

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

Hasnat, Md. A., Alata, O., & Trémeau, A. (2014). Unsupervised Clustering of Depth Images Using Watson Mixture Model. *2014 22nd International Conference on Pattern Recognition*, 214–219. https://doi.org/10.1109/ICPR.2014.46

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

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., Handwerker, D. A., Keilholz, S., Kiviniemi, V., Leopold, D. A., de Pasquale, F., Sporns, O., Walter, M., & Chang, C. (2013). Dynamic functional connectivity: Promise, issues, and interpretations. *NeuroImage*, *80*, 360–378. https://doi.org/10.1016/j.neuroimage.2013.05.079

Ito, N., Araki, S., & Nakatani, T. (2016). Complex angular central Gaussian mixture model for directional statistics in mask-based microphone array signal processing. *2016 24th European Signal Processing Conference (EUSIPCO)*, 1153–1157. https://doi.org/10.1109/EUSIPCO.2016.7760429

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

Kummer, E. E. (1836). Über die hypergeometrische Reihe ... . *Journal für die reine und angewandte Mathematik*, *15*, 39–83.

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

Lurie, D., Kessler, D., Bassett, D., Betzel, R., Breakspear, M., Keilholz, S., Kucyi, A., Liégeois, R., Lindquist, M., McIntosh, A. R., Poldrack, R., Shine, J., Thompson, W. H., Bielczyk, N., Douw, L., Kraft, D., Miller, R., Muthuraman, M., Pasquini, L., … Calhoun, V. (2018). *On the nature of time-varying functional connectivity in resting fMRI*. 1–34. https://doi.org/10.31234/osf.io/xtzre

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., & Dryden, I. L. (1999). The Complex Watson Distribution and Shape Analysis. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, *61*(4), 913–926.

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

Pang, J. C., Aquino, K. M., Oldehinkel, M., Robinson, P. A., Fulcher, B. D., Breakspear, M., & Fornito, A. (2023). Geometric constraints on human brain function. *Nature*, *618*(7965), Article 7965. https://doi.org/10.1038/s41586-023-06098-1

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

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

Sra, S., & Karp, D. (2013). The multivariate watson distribution: Maximum-likelihood estimation and other aspects. *Journal of Multivariate Analysis*, *114*(1), 256–269. https://doi.org/10.1016/j.jmva.2012.08.010

Srivastava, M. S. (2003). Singular Wishart and multivariate beta distributions. *The Annals of Statistics*, *31*(5), 1537–1560. https://doi.org/10.1214/aos/1065705118

Srivastava, M. S., & Khatri, C. G. (1979). *An Introduction to Multivariate Statistics*. North-Holland/New York.

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

Taghia, J., & Leijon, A. (2016). Variational Inference for Watson Mixture Model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *38*(9), 1886–1900. https://doi.org/10.1109/TPAMI.2015.2498935

Thomas Yeo, B. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., Roffman, J. L., Smoller, J. W., Zöllei, L., Polimeni, J. R., Fisch, B., Liu, H., & Buckner, R. L. (2011). The organization of the human cerebral cortex estimated by intrinsic functional

connectivity. *Journal of Neurophysiology*, *106*(3), 1125–1165. https://doi.org/10.1152/jn.00338.2011

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

Tran Vu, D. H., & Haeb-Umbach, R. (2010). Blind speech separation employing directional statistics in an Expectation Maximization framework. *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, 241–244. https://doi.org/10.1109/ICASSP.2010.5495994

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

Uhlig, H. (1994). On Singular Wishart and Singular Multivariate Beta Distributions. *The Annals of Statistics*, *22*(1), 395–405. https://doi.org/10.1214/aos/1176325375

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

Watson, G. S. (1965). Equatorial Distributions on a Sphere. *Biometrika*, *52*(1/2), 193. https://doi.org/10.2307/2333824

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

You, K. (2022). *A Note on Angular Central Gaussian Distribution and its Matrix Variant*. https://kisungyou.com/posts/note003-angular-gaussian

Zalesky, A., Fornito, A., Cocchi, L., Gollo, L. L., & Breakspear, M. (2014). Time-resolved resting-state brain networks. *Proceedings of the National Academy of Sciences*, *111*(28), 10341–10346. https://doi.org/10.1073/pnas.1400181111

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