

INTRODUCTION

Functional magnetic resonance imaging (fMRI) is a brain imaging technique that measures fluctuations in blood-oxygen associated with neuronal activity. These measurements can be used to quantify voxel-wise functional connectivity relationships [Biswal et al., 1995, 2010; Lowe et al., 1998], which are commonly defined in terms of linear covariance estimates, or equivalently, correlation for standardized time-series. As fMRI measurements are often noisy and exhibit considerable inter-subject variability, the empirical estimation of covariance (or correlation) can be highly unreliable. This prohibits the analysis and interpretation of measured functional connectivity structure, particularly when trying to estimate generalizable group-level connectivity. This issue is further intensified when correcting for interdependency, for example, when estimating partial correlations [Marrelec et al., 2006].

To address this issue, two approaches are commonly employed to reduce noise and inter-subject variability: reducing the connectivity analysis to anatomically predefined brain regions by averaging voxels [Achard et al., 2006; Bassett and Bullmore, 2006; Biswal et al., 1995; Fox et al., 2005; Greicius et al., 2003; Smith et al., 2010], and filtering noise by compressing the high dimensional spatial representation into a low dimensional subspace using multivariate component models [Calhoun et al., 2001b; Friston et al., 1993; McKeown et al., 1998]. In this article, we focus on the latter approach, which identifies sets of brain regions with similar temporal dynamics. Although it is the more challenging modeling approach, there are significant benefits to “learning” the salient features of functional connectivity, without being biased by *a priori* assumptions regarding the significance and functional organization of cortical regions.

A variety of different component models have been developed for modeling multi-subject fMRI data [Beckmann and Smith, 2005; Calhoun et al., 2001a, 2009; Damoiseaux et al., 2006; Esposito et al., 2005; Guo and Pagnoni, 2008; Schmithorst and Holland, 2004]. Typical approaches include Principal Component Analysis (PCA) which maximizes explained variance, and Independent Component Analysis (ICA) models, where component bases are usually linear rotations of the PCA space that maximize the independence of higher-order moments [Bell and Sejnowski, 1995]. These models can be used for both individual analysis and the modeling of subject groups. For groups, the extracted spatial subspace is assumed to be consistent across subjects, and components are estimated on data that are temporally concatenated across subjects (e.g., “group ICA” and “concat-ICA” approaches [Calhoun et al., 2001b]). As a compromise between individual and group-level models, dual regression techniques [Filippini et al., 2009] use group-level spatial components to estimate subject-specific time-courses, which are then used as regressors to estimate the within-subject brain maps that best approximate the group-level spatial components.

When modeling functional connectivity relationships in multi-subject data, it may be useful to assume that interactions between components in the extracted subspace are consistent across subjects, to improve model generalization and simplify the interpretation of model parameters. This is the assumption behind the multi-way PARAFAC2 model [Harshman, 1972; Kiers et al., 1999]. This model identifies a component basis that maximizes the explained data variance, under the constraint of a fixed connectivity network, with subject-dependent network scaling parameters. The PARAFAC2 model has been widely used in chemometrics [Bro et al., 1999], with recent applications in EEG [Weis et al., 2010] and integration of EEG-fMRI data [Ferdowsi et al., 2013], although it has not been previously used to quantify whole-brain functional connectivity structure in fMRI.

Other data-driven techniques have also been developed to compress fMRI data and summarize structure in functional data. For example, clustering models are among the most widely used alternatives to component modeling. Unlike component models, which represent voxels as a linear combination of independent sources, these techniques identify homogeneous brain regions that constitute functional units [Cohen et al., 2008; Goutte et al., 1999]. In recent years clustering methods have also been extended to perform group-level modeling in neuroimaging data, for example, Craddock et al. [2012]. However, to date, there has been little comparison between component-based and clustering models, in terms of their ability to robustly model group-level functional connectivity.

An important goal when modeling functional brain networks is to identify reliable network structure in multi-subject data, by leveraging shared structure across subjects. This is a non-trivial challenge, and there has been little investigation into which component models are best suited for this purpose. In particular, component models can be defined with varying levels of flexibility, ranging from fully subject-specific PCA and ICA models, to moderately constrained PARAFAC2 models (e.g., consistent covariance but some flexible parameters), to highly constrained representations of “average” covariance in the group. The optimal model will depend on which level of flexibility best captures functional connectivity features within the group.

To evaluate the effectiveness of component models at defining whole-brain functional connectivity, we may ask how generalizable the results are. That is, how well do they predict between-voxel functional connectivity values in independent datasets? To address this question, we implemented a cross-validation framework that performs within-subject splitting, to estimate prediction accuracy of whole-brain connectivity for individual subjects [Hansen et al., 1999; Strother et al., 2002; Thirion et al., 2014]. We used three different measures of covariance generalization, and compared a range of different component models including: within-subject analysis, concatenated group-

level models, dual regression, and constrained-covariance models based on the PARAFAC2 framework. In addition, we compared the performance of the group-level clustering model developed by [Craddock et al., 2012]. These models were evaluated using both simulated and experimental resting-state fMRI datasets, thereby highlighting how well different component models quantify whole-brain connectivity.

As a secondary issue, researchers are often concerned with the interpretation of parameter estimates produced by component models, in particular the anatomical regions represented in spatial components, and the functional connectivity between individual components. These parameter estimates must be stable for the population of interest, otherwise they do not allow for a meaningful interpretation of functional structure, as there is no reason to choose one parameter representation over another. For experimental data, we also evaluated the reproducibility and interpretability of both the spatial components and between-components connectivity measures within our cross-validation framework [Himberg et al., 2004; Strother et al., 2002; Thirion et al., 2014]. This was evaluated for the two classes of models that optimally predict voxel-wise functional connectivity, including PARAFAC2 and group PCA/ICA. These analyses demonstrate that models with comparable generalization may still significantly differ in the reproducibility and interpretability of model parameters.

MATERIALS AND METHODS

Component Models for the Quantification of Multi-Subject Functional Connectivity

Consider an fMRI data matrix $\mathbf{X}_s \in \mathbb{R}^{V \times T}$, consisting of V brain voxels, sampled at T time-points for subject s . If the voxel (row) means are subtracted, total functional connectivity in the brain is given by the empirical “outer” $V \times V$ covariance matrix $\frac{1}{T} \mathbf{X}_s \mathbf{X}_s^\top$, where T^* is the effective degrees of freedom in \mathbf{X}_s ; typically, $T^* < T$ due to the subtraction of row means and regression-based preprocessing (e.g., controlling for residual motion artifact and physiological noise). Component-based models seek a linear reduced-rank representation of the data, to improve the generalizability of covariance estimates, and reduce the data to a smaller, more interpretable set of features.

The singular value decomposition (SVD) provides the optimal representation of data in terms of explained variance. If we decompose a single subject’s data by SVD, we obtain

$$\mathbf{X}_s \approx \mathbf{U}_s^{V \times C} \boldsymbol{\Sigma}_s^{C \times C} \mathbf{V}_s^{T \times C^\top},$$

for $C \leq T^*$, where \mathbf{U}_s and \mathbf{V}_s are orthonormal matrices. We thereby model subject-specific voxel-wise covariance by $\frac{1}{T} \mathbf{X}_s \mathbf{X}_s^\top \approx \frac{1}{T} \mathbf{U}_s \boldsymbol{\Sigma}_s^2 \mathbf{U}_s^\top$.

Neuroscience studies often use component models to estimate the most consistent functional connectivity structure within a group of subjects. This is often done by concatenating the data of each subject into the data matrix $[\mathbf{X}_1, \dots, \mathbf{X}_s]$ and performing SVD:

$$[\mathbf{X}_1, \dots, \mathbf{X}_s] \approx \mathbf{U} \boldsymbol{\Sigma} [\mathbf{V}_1^\top, \dots, \mathbf{V}_s^\top]^\top.$$

In this case, the modeled covariance $\frac{1}{T} [\mathbf{X}_1, \dots, \mathbf{X}_s] [\mathbf{X}_1, \dots, \mathbf{X}_s]^\top \approx \frac{1}{T} \mathbf{U} \boldsymbol{\Sigma}^2 \mathbf{U}^\top$ forms a single “consensus” estimate of voxel covariance across all subjects, whereas the covariance of each subject based on this model typically differ since $\frac{1}{T} \mathbf{X}_s \mathbf{X}_s^\top \approx \frac{1}{T} \mathbf{U} \boldsymbol{\Sigma} \mathbf{V}_s^\top \mathbf{V}_s \boldsymbol{\Sigma} \mathbf{U}^\top \neq \frac{1}{T} \mathbf{U} \boldsymbol{\Sigma} \mathbf{V}_s^\top \mathbf{V}_s \boldsymbol{\Sigma} \mathbf{U}^\top \approx \frac{1}{T} \mathbf{X}_s \mathbf{X}_s^\top$. The inequality follows from $\mathbf{V}_s^\top \mathbf{V}_s \neq \mathbf{V}_s'^\top \mathbf{V}_s'$ in general. If we wish to correctly estimate a consensus representation of subject covariance, we must constrain $\mathbf{V}_1^\top \mathbf{V}_1 = \mathbf{V}_2^\top \mathbf{V}_2 = \dots = \mathbf{V}_s^\top \mathbf{V}_s = \mathbf{I}$, for example,

$$[\mathbf{X}_1, \dots, \mathbf{X}_s] \approx \mathbf{U} \boldsymbol{\Sigma} [\mathbf{Q}_1^\top, \dots, \mathbf{Q}_s^\top]^\top,$$

where all $\mathbf{Q}_s^\top \mathbf{Q}_s = \mathbf{I}$. This model enforces a consistent covariance representation across subjects that maximizes explained variance, under the constraint that components be orthogonal in both space and time, parameterized by spectral decomposition $\mathbf{U} \boldsymbol{\Sigma}^2 \mathbf{U}^\top$. We can relax this strict constraint by permitting consistent but non-orthogonal covariance between components $\mathbf{V}_s^\top \mathbf{V}_s = \mathbf{R}^\top \mathbf{R}$. In this case, we re-parameterize the model by adding a consistent network term \mathbf{R} :

$$[\mathbf{X}_1, \dots, \mathbf{X}_s] \approx \mathbf{U} \boldsymbol{\Sigma} \mathbf{R}^\top [\mathbf{Q}_1^\top, \dots, \mathbf{Q}_s^\top]^\top,$$

Although more flexible, the above representation does not permit components of the underlying network to vary in strength across subjects. This can be alleviated by replacing $\boldsymbol{\Sigma}$ with a subject-specific (non-negative) diagonal matrix $\boldsymbol{\Sigma}_s$, so that we approximate the s th subject’s data matrix by:

$$\mathbf{X}_s \approx \mathbf{U} \boldsymbol{\Sigma}_s \mathbf{R}^\top \mathbf{Q}_s^\top.$$

This latter model has been well studied in the psychometrics and chemometrics literature, where it is denoted the PARAFAC2 model [Bro et al., 1999; Harshman, 1972; Kiers et al., 1999]. We note that \mathbf{U} is considered orthonormal in all discussed models for consistency; however, this is not assumed in the most general formulation of the PARAFAC2 model.

There are thus a number of different approaches to modeling group functional connectivity in fMRI, with tradeoffs between model simplicity and flexibility of fitting individual subjects. In Table I, we define the above five different modeling approaches to quantify subject specific voxel-wise functional connectivity matrices \mathbf{M}_s in group data, ordered by increasing flexibility in their ability to represent subject-specific effects.

Shown in Table I, \mathbf{R} has effectively $K(K-1)/2$ free parameters, as $\mathbf{R}^\top \mathbf{R}$ can be defined in terms of an upper triangular matrix representation for \mathbf{R} given by the Cholesky decomposition, reduced by K parameters, as all the diagonal

TABLE I. The five different models considered for parameterizing the de-noised covariance M_s , as well as their numbers of free parameters F

	Parameterization of voxel-wise covariance	Number of model parameters F , as well as subject specific estimated correlation C_s between estimated components
Average	$M_s = \frac{1}{T^*} \mathbf{U} \Sigma^2 \mathbf{U}^\top$	$F = K(V + T^*S + 1) - (S + 1)K(K + 1)/2 + S$ $F_{IC} = K(T^*S + 1) - SK(K + 1)/2 + S$ $C_s = \mathbf{I}$
All subject covariance matrices are approximated by the same spectral decomposition, based on the model $[\mathbf{X}_1, \dots, \mathbf{X}_s] \approx [\mathbf{Z}_1, \dots, \mathbf{Z}_s] = \mathbf{U} \Sigma [\mathbf{Q}_1^\top, \dots, \mathbf{Q}_s^\top]$		
Consistent	$M_s = \frac{1}{T^*} \mathbf{U} \Sigma (\mathbf{R}^\top \mathbf{R}) \Sigma \mathbf{U}^\top$	$F = K(V + T^*S + 1) + K(K - 1)/2 - (S + 1)K(K + 1)/2 + S$ $F_{IC} = K(T^*S + 1) + K(K - 1)/2 - SK(K + 1)/2 + S$ $C_s = \mathbf{R}^\top \mathbf{R}$
All subject covariance matrices are approximated by assuming the covariance between the extracted subject specific time-series are consistent based on the model $[\mathbf{X}_1, \dots, \mathbf{X}_s] \approx [\mathbf{Z}_1, \dots, \mathbf{Z}_s] = \mathbf{U} \Sigma \mathbf{R}^\top [\mathbf{Q}_1^\top, \dots, \mathbf{Q}_s^\top]$		
Scaled consistent	$M_s = \frac{1}{T^*} \mathbf{U} \Sigma_s (\mathbf{R}^\top \mathbf{R}) \Sigma_s \mathbf{U}^\top$	$F = K(V + T^*S + S) + K(K - 1)/2 - (S + 1)K(K + 1)/2 + S$ $F_{IC} = K(T^*S + S) + K(K - 1)/2 - SK(K + 1)/2 + S$ $C_s = \mathbf{R}^\top \mathbf{R}$
Relaxes the constraint of fixed component scaling across subjects based on the model $\mathbf{X}_s \approx \mathbf{Z}_s = \mathbf{U} \Sigma_s \mathbf{R}^\top \mathbf{Q}_s^\top$ corresponding to PARAFAC2.		
Common subspace	$M_s = \frac{1}{T^*} \mathbf{U} \Sigma \mathbf{V}_s^\top \mathbf{V}_s \Sigma \mathbf{U}^\top$	$F = K(V + T^*S + 1) - K(K + 1) + S$ $F_{IC} = K(T^*S + 1) - K(K + 1)/2 + S$ $C_s = \mathbf{V}_s^\top \mathbf{V}_s$
Covariance between components is individual defined for each subject based on the group PCA/ICA model $[\mathbf{X}_1, \dots, \mathbf{X}_s] \approx [\mathbf{Z}_1, \dots, \mathbf{Z}_s] = \mathbf{U} \Sigma [\mathbf{V}_1^\top, \dots, \mathbf{V}_s^\top]$		
Individual	$M_s = \frac{1}{T^*} \mathbf{U}_s \Sigma_s^2 \mathbf{U}_s^\top$	$F = S(K(V + T^* + 1) - K(K + 1) + 1)$ $F_{IC} = S(K(T^* + 1) - K(K + 1)/2 + 1)$ $C_s = \mathbf{I}$
Each subject data is modelled individually by PCA/ICA, based on the model $\mathbf{X}_s \approx \mathbf{Z}_s = \mathbf{U}_s \Sigma_s \mathbf{V}_s^\top$		

K , V , S , and T^* respectively denote the number of components, voxels, subjects and effective degrees of freedom in the time-series. \mathbf{Z}_s and M_s denotes the models' de-noised estimates of the signal and covariance, respectively, for subjects. The columns of \mathbf{R} and \mathbf{V} are unit normalized. F_{IC} denotes the number of free parameters used in calculating the Bayesian/Akaike information criteria, when considering the spatial mode \mathbf{U} (largest dimensionality) marginalized in the training likelihood.

elements of $\mathbf{R}^\top \mathbf{R}$ are fixed to one. Thus, the *Scaled Consistent* model (PARAFAC2) has the same number of free parameters as the *Common Subspace* model (group PCA/ICA) when $S=2$, and becomes more constrained as S increases. As the voxel mode is much larger than the number of subjects and time points, the *Individual* model has substantially more parameters than the other models; the spatial mode \mathbf{U}_s , is defined individually with a total of SKV parameters.

We obtained solutions for the different models, computed in Matlab, as follows: *Common Subspace* and *Individual* models were computed using a standard SVD, which provides the PCA solution; we obtained both spatial and temporal ICA (sICA and tICA) by applying the FastICA algorithm with "pow3" nonlinearity and the deflation approach to decorrelate components (code from: <http://research.ics.aalto.fi/ica/fastica/>) to spatial and temporal components, respectively [Hyvärinen, 1999]. We estimated the *Scaled Consistent* PARAFAC2 solution using the software provided at (<http://www.models.life.ku.dk/algorithms>), which uses the modified Alternating Least Squares (ALS) solution proposed in [Bro et al., 1999]. The *Average* and *Consistent* solutions were estimated using modified implementations of the PARAFAC2 solution. The

five component models were also compared with dual regression (*Dualreg*), which was implemented in Matlab. This article examines dual regression on a PCA basis, as ICA methods did not show a significant difference in generalization. We also evaluated the clustering model of [Craddock et al., 2012] (*Clustering*), with results shown for the temporal-model clustering, as the spatial-mode clustering did not show significant differences in generalization. Cluster-model networks were obtained using the *Common Subspace* approach, but replacing component maps with the unit-normalized binary clustering maps. When evaluating the performance of the clustering approach, it should be noted that clustering is inherently less flexible as it does not allow scaling of the voxel contributions.

In this section, we have defined three features of component-based models that are important for characterizing functional connectivity in group data: (1) the "outer" $V \times V$ covariance matrix, which captures all functional connectivity relationships in the brain; (2) the spatial orthonormal basis \mathbf{U} of brain patterns showing reliable covariance relationships in the data; and (3) the "inner" correlation matrix C_s , which defines the functional connectivity between individual components of \mathbf{U} .

A Cross-Validation Framework for Evaluating Component Models

To compare between component models, we must quantify their ability to detect generalizable features of covariance in fMRI data. With increasing model flexibility, we obtain a richer description of group functional connectivity, but potentially at the expense of decreased model stability and generalizability. In the absence of a “ground truth” for experimental fMRI data, we have developed a cross-validation approach based on the NPAIRS framework [Strother et al., 2002], that uses multiple performance metrics. The following sections review the testing framework, and key features of the metrics. Of primary interest is the ability of the component representation of the $V \times V$ covariance matrix to predict between-voxel connectivity in independent test data. However, we also consider metrics quantifying the reproducibility of the spatial basis \mathbf{U} , and of the correlation matrix \mathbf{C}_s between time series of extracted components. Because model performance may depend on the chosen metric of prediction and reproducibility, we tested multiple different prediction and reproducibility measures.

Prediction of the Voxel-Wise Covariance

We evaluated prediction accuracy of the fitted component model, based on how well the full $V \times V$ subject covariance matrix generalizes to independent test-data covariance $\frac{1}{T^*} \mathbf{X}_s^* \mathbf{X}_s^{*\top}$. Covariance is expected to vary across subjects; our goal is to determine what level of model flexibility gives the most accurate prediction of individual-subject covariance. To perform model evaluation, we defined a within-subject cross-validation framework for group data. For the dataset $\{\mathbf{X}_s\}$ ($s=1 \dots S$), each matrix was split into two split-halves containing an equal number of time-points. The network model was constructed using split-half 1, and the estimated parameters were used to predict covariance $\frac{1}{T^*} \mathbf{X}_s \mathbf{X}_s^\top$ of split-half 2. We then exchanged splits and performed modeling on split-half 2 and prediction on split-half 1, before computing average prediction across splits. The following three measures of prediction were evaluated.

Indirect fitting of voxel covariance (IF)

The IF is given by the sum of squared error (SSE) between the training model and test data, over all elements of the $V \times V$ voxel covariance matrix. For subject s , we computed the Frobenius distance of the modeled covariance matrix \mathbf{M}_s , relative to test data \mathbf{X}_s^* :

$$\text{SSE}(\mathbf{X}_s^*) = \left\| \frac{1}{T^*} \mathbf{X}_s^* \mathbf{X}_s^{*\top} - \mathbf{M}_s \right\|_F^2$$

Total model prediction error was then computed as the sum of subject error terms $\text{IF} = \sum_s \text{SSE}(\mathbf{X}_s^*)$. This metric is

based on the indirect fitting objective developed for PAR-AFAC2. [Harshman, 1972; Kiers, 1993]. Although a statistical theory has not been developed around IF, it has the advantage of minimal modeling assumptions, and is robust in the case of ill-posed covariance matrices, where $V \gg T^*$. Because the IF measures the squared error of variance, it is expected to be highly sensitive to misspecification of variance scaling. This metric has been previously examined as a goodness-of-fit estimator in the context of structural equation modeling [Bentler and Bonett, 1980; MacCallum, 1996] and economic forecasting [Ledoit and Wolf, 2004], but has not yet been applied to functional neuroimaging. Note that this metric includes the main diagonal terms of the covariance matrix which are related to the signal power. To test whether this had an effect, we also calculated the IF while excluding the main diagonal, which led to values that were indistinguishable from the original IF metric.

Covariance generalization error (GE)

This is a probabilistic estimate of generalization error, similar to [Hansen et al., 1999], but used to directly estimate covariance matrices. For subject data matrix \mathbf{X}_s (with row means subtracted), we consider the K -dimensional signal subspace representation \mathbf{M}_s with eigen-decomposition $\mathbf{M}_s = \mathbf{S} \mathbf{A} \mathbf{S}^\top$, and assume that noise σ_{noise}^2 is isotropic Gaussian, that is, $\sigma_{\text{noise}}^2 = \frac{1}{V-K} \text{trace}(\frac{1}{T^*} \mathbf{X}_s \mathbf{X}_s^\top - \mathbf{S} \mathbf{A} \mathbf{S}^\top)$. Given total modeled covariance $\hat{\mathbf{\Sigma}}_{\text{tot}} = \mathbf{M}_s + \mathbf{I}_V \sigma_{\text{noise}}^2$, we can measure the Wishart likelihood of the test-data scatter matrix $\mathbf{X}_s^* \mathbf{X}_s^{*\top}$ with T^* degrees of freedom. Because the test scatter matrix is singular, model generalization was inferred by the negative log-likelihood of the singular Wishart distribution [Uhlig et al., 1994]:

$$L_{\text{GE}}(\mathbf{X}_s^*) = \frac{1}{2} (VT^* - T^{*2}) \log(\pi) + \frac{VT^*}{2} \log(2) + \log \Gamma_{T^*} \left(\frac{T^*}{2} \right) - \frac{1}{2} (T^* - V - 1) \log |\mathbf{X}_s^* \mathbf{X}_s^{*\top}|_+ + \frac{T^*}{2} |\hat{\mathbf{\Sigma}}_{\text{tot}}| + \frac{1}{2} \text{trace} \left(\hat{\mathbf{\Sigma}}_{\text{tot}}^{-1} \mathbf{X}_s^* \mathbf{X}_s^{*\top} \right)$$

where $|\mathbf{X}_s^* \mathbf{X}_s^{*\top}|_+$ denotes the pseudo-determinant, given by the product of all non-zero singular values, and the following terms are defined as per [Hansen et al., 1999]:

$$\begin{aligned} |\hat{\mathbf{\Sigma}}_{\text{tot}}| &= \sum_{k=1}^K \log(\lambda_k) + (V-K) \log(\sigma_{\text{noise}}^2) \\ \text{trace} \left(\hat{\mathbf{\Sigma}}_{\text{tot}}^{-1} \mathbf{X}_s^* \mathbf{X}_s^{*\top} \right) &= \frac{\sum_{t=1}^T \mathbf{x}_s^*(t) \mathbf{x}_s^*(t)^\top - \sum_{k=1}^K \frac{\lambda_k - \sigma_{\text{noise}}^2}{\lambda_k} \sum_{t=1}^T [\mathbf{x}_s^*(t)^\top \mathbf{s}_k]^2}{\sigma_{\text{noise}}^2} \end{aligned}$$

Model prediction error was then obtained by summing subject-specific log-likelihood terms: $\text{GE} = \sum_s L_{\text{GE}}(\mathbf{X}_s^*)$. This model is cast in a multivariate probabilistic framework, which is consistent with standard fMRI network modeling assumptions. However, it makes the simplifying

assumption of isotropic noise variance over voxels, which may be unrealistic for fMRI data.

Heteroscedastic covariance generalization error (HGE)

As an alternative to GE, we may assume that the component signal space is independent of noise, and that noise variance is voxel-specific (heteroscedastic). Consider again the modeled K -dimensional subspace representation of data covariance M_s with eigen-decomposition $M_s = S\Lambda S^\top$, but now with voxel-wise noise variance $\text{diag}(\sigma_{\text{noise}}^2)$, estimated on the residual of the training data after subtracting the signal subspace $X_{\text{resid}} = X_s - U\Sigma_s V_s^\top$. We again measured negative log-likelihood of the test-data scatter matrix $X_s^* X_s^{*\top}$ with T^* degrees of freedom, based on the singular Wishart distribution, and the $L_{\text{HGE}}(X_s^*)$ takes the same form as $L_{\text{GE}}(X_s^*)$ above. However, we total covariance is now defined as $\hat{\Sigma}_{\text{tot}} = M_s + \text{diag}(\sigma_{\text{noise}}^2)$, with the following substitutions:

$$|\hat{\Sigma}_{\text{tot}}| = |I_K + A^{1/2} S^\top \text{diag}(\sigma_{\text{noise}}^{-2}) S A^{1/2}| + \sum_{v=1}^V \sigma(v)_{\text{noise}}^2$$

$$\text{trace}(\hat{\Sigma}_{\text{tot}}^{-1} X_s^* X_s^{*\top}) = \sum_{t=1}^T x_s^*(t)^\top [\text{diag}(\sigma_{\text{noise}}^2) - \text{diag}(\sigma_{\text{noise}}^{-2}) S (A^{-1} + S^\top \text{diag}(\sigma_{\text{noise}}^{-2}) S) S^\top \text{diag}(\sigma_{\text{noise}}^{-2})] x_s^*(t)$$

The above expressions make use of Woodbury's matrix inversion lemma as well as the matrix determinant lemma. The model prediction error is again obtained by summing over subject-specific negative log-likelihood terms: $\text{HGE} = \sum_s L_{\text{HGE}}(X_s^*)$. This probabilistic model employs a more flexible spatially-varying noise estimate than GE, under the assumption that the estimated signal subspace is noise-independent.

Total prediction error was plotted for each metric and all component models, as a function of model order K . For all prediction curves, we also generated confidence bounds on total prediction error by Bootstrapping on subjects and computing standard error on the distribution of IF, GE, and HGE values.

Information Criteria

In addition to the generalization error, we evaluated model evidence via the Bayesian information criterion (BIC) [Schwarz, 1978] and the Akaike Information Criterion (AIC) [Akaike, 1974]. To obtain the training likelihood for this purpose we consider all the models represented in the form $\tilde{X} \approx U\mathbf{L}$, where \tilde{X} is a $V \times (T^*S)$ concatenated matrix of subject data in a compressed temporal representation, U is the orthonormal voxel-projection matrix, and L is the associated low-dimensional data representation. The full-rank representation (T^* temporal dimensions per subject) can be obtained as a set of subject-specific SVDs, concatenated on the temporal dimension. We approximate this data representation with the generative model:

$$u_v \sim \mathcal{N}(0, V^{-1}I)$$

$$\tilde{x}(v) \sim \mathcal{N}(u_v L, \Psi)$$

where Ψ is a diagonal noise variance matrix with T^*S elements; the T^* elements corresponding to subject s are given by residual noise variance σ_s^2 , averaged across all voxels and temporal dimensions.

The marginalized likelihood is obtained by integrating out u_v , using Woodbury's matrix inversion lemma:

$$P(\tilde{X}|L) = \prod_v \int \mathcal{N}(\tilde{x}(v)|u_v L, \Psi) \mathcal{N}(u_v|0, V^{-1}I) du_v$$

$$= \prod_v \mathcal{N}(\tilde{x}(v)|0, \Psi + V^{-1}L^\top L)$$

This likelihood can then be used to obtain the BIC, which is a large-sample approximation to the model evidence:

$$\text{BIC} = V \log |\Psi + V^{-1}L^\top L| + \sum_v \tilde{x}(v)^\top (\Psi + V^{-1}L^\top L)^{-1} \tilde{x}(v) + F_{\text{IC}} \log(SVT^*).$$

Similarly for AIC, we obtain:

$$\text{AIC} = V \log |\Psi + V^{-1}L^\top L| + \sum_v \tilde{x}(v)^\top (\Psi + V^{-1}L^\top L)^{-1} \tilde{x}(v) + 2F_{\text{IC}},$$

where the number of free parameters, F_{IC} , corresponds to the number of free parameters in L and σ_s^2 . These numbers are given in Table I. Note that for the *Individual* model, likelihood is computed by marginalizing with respect to u_v for each subject, and thus the information criteria are calculated individually, and then summed across subjects. Model evidence was plotted for all component models, as a function of model order K .

Reproducibility of Spatial Parameters

We further examined the reproducibility of the estimated network spatial basis U , since a common goal of component modeling is to interpret spatial component features. Components must be stable in order to have a valid interpretation which, at minimum, requires that there be no model ambiguity. For PCA-based models, this is fulfilled as long as all singular values of its eigen-decomposition are distinct. For models enforcing consistent group covariance, uniqueness is less clear. However, uniqueness conditions exist for the PARAFAC2 model [Berge and Kiers, 1996; Harshman et al., 1996] and our model is further constrained by spatial orthonormality. In general, we do not expect reproducibility to be hampered by non-identifiability issues.

When model ambiguity is not a concern, reproducibility may be evaluated separately from prediction: due to bias-variance tradeoffs in model parameterization, it is possible to have a highly predictive model with unstable voxel features. Reproducibility was evaluated using the cross-

validation procedure in the previous sub-section, and we compared fitted parameters for the component models trained on split-half 1 relative to split-half 2. We evaluated two different metrics of spatial reproducibility:

Amari distance metric (AD)

This metric was developed and is widely used in the context of comparing ICA source-separation models [Amari et al., 1996]. For two split-half matrices X and Y with columns corresponding to the spatial component vectors, we define cross-product matrix $Q = XY^{-1}$ (where Y^{-1} denotes the pseudo-inverse of Y for rectangular matrices). We then compute the “distance” between subspaces estimated for the two split-halves by:

$$AD(X, Y) = \frac{1}{K} \left(\sum_{i=1}^K \left[\left(\sum_{j=1}^K |Q_{ij}| / \max_j |Q_{ij}| \right) - 1 \right] + \sum_{j=1}^K \left[\left(\sum_{i=1}^K |Q_{ij}| / \max_i |Q_{ij}| \right) - 1 \right] \right)$$

Because this is a non-symmetric measure, we computed the average of $AD(X, Y)$ and $AD(Y, X)$. This metric quantifies the consistency of the basis vectors in X and Y . Thus, for two matrices spanning the same subspace, it will assign a lower AD value if the individual components of matrices X and Y have high pairwise correlation.

RV coefficient (RV)

The RV coefficient, first established by [Escoufier, 1973], is an alternative metric of reproducibility between spatial subspaces. It is the multivariate generalization of the Pearson correlation to square, positive semi-definite matrices and is applied to rectangular matrices X and Y , by measuring the RV coefficient on scatter matrices $S_x = XX^T$ and $S_y = YY^T$. The RV coefficient is computed as:

$$RV(X, Y) = \frac{\text{trace}\{(X^T Y)^2\}}{\sqrt{\text{trace}\{(X^T X)^2\}} \sqrt{\text{trace}\{(Y^T Y)^2\}}}$$

This metric is of particular interest because most multivariate analysis models can be defined as maximizing the RV coefficient, subject to additional constraints, including PCA, partial least squares, and canonical correlation analysis [Robert et al., 2007]. Unlike AD, this similarity metric is rotationally invariant, and does not depend on the orientation of the individual components in X and Y but rather the overlap of their spanning subspace.

Reproducibility of the Functional Network Model

A second parameter set of interest is the connectivity between components; this feature may be as important as

the spatial components in studies of functional neuroimaging. In this article, we focused on estimation of both the average between-cluster connectivity C and the partial correlation matrix P , which fully accounts for dependencies between variables [Marrelec et al., 2006]. In general, connectivity values cannot be directly compared between the split-halves, as they are estimated for non-identical component bases. Instead, we developed a within-split bootstrapping approach. For this procedure, we initially fit the component model of interest on all subject data, to obtain spatial basis U . Subjects were then sampled with replacement for $b=1, \dots, 1,000$ iterations, and for each sample, we re-estimated the average between-cluster connectivity C_b on the fixed basis U . For the *Common Subspace* model, C_b was obtained by taking the average of all subject connectivity matrices $C_s = V_s^T V_s$ in the bootstrap sample scaled to unit diagonal to reflect correlations. For the *Scaled Consistent* model this was done by re-fitting the PARAFAC2 model on the bootstrap sample with fixed U , to generate a new estimate $C_b = R_b^T R_b$. For each connectivity element (i, j) , we converted the bootstrapped correlations into normally-distributed variables via Fisher’s Z-transform, defined as $c_{ij,b}^* = \frac{1}{2} \log((1+c_{ij,b})/(1-c_{ij,b}))$. Using the transformed values $\{c_{ij,b}^*\}_{b=1}^{1000}$, we computed the bootstrap ratio as the mean divided by the standard error, which provides z-scored measure of reliability for all elements of C [Efron and Tibshirani, 1986], and allows us to estimate P -values based on a two-tailed normal distribution. A similar approach was used to estimate z-score significance for all elements of P , by converting bootstrapped C_b matrices into P_b .

We obtained a robust measure of overall reliability for the connectivity matrices, by taking the median of (absolute) bootstrap ratio values for all connectivity elements (i, j) . The median bootstrap ratio was then compared for component models as a function of model order K . We also compared the reliability of lagged models of maximum connectivity, as suggested in Arbabshirani et al. [2013], by projecting subject data onto the fixed spatial basis $V_s = X_s^T U$, then selecting the time-lag that maximizes each pairwise correlation between columns of V_s within a ± 2 TR window. The bootstrapping procedure was then repeated for time-lagged estimates of subject connectivity matrices C_s .

Visualization and Interpretation of Component Models

While the metrics of prediction and reproducibility are used to guide model selection, the ultimate goal of group functional connectivity studies is the visualization and interpretation of the results in the context of neuroscience research. To this end, we must be able to display spatial component maps U and between-component correlation matrix C , and assess them in terms of statistical significance.

Representing spatial components

We obtained consensus spatial maps that are Z-scored based on their reproducibility, using the procedure established by [Strother et al., 2002]. Because the individual spatial components of matrices $\mathbf{U}^{(1)}$ (split-1) and $\mathbf{U}^{(2)}$ (split-2) may have arbitrary ordering and sign, we first matched components between splits using a restricted Procrustes matching procedure (i.e., admitting only re-ordering and sign-flipping of components) to minimize a total sum-of-squares cost function. For each pair of matched components $\mathbf{u}_k^{(1)}$ and $\mathbf{u}_k^{(2)}$, we normalized them to mean zero and variance one, and constructed the scatterplot of $\mathbf{u}_k^{(1)}$ versus $\mathbf{u}_k^{(2)}$ pairwise voxel values. The projection $\mathbf{u}^{(e)} = (\mathbf{u}_d^{(1)} + \mathbf{u}_d^{(2)})/\sqrt{2}$ (1st PC of the scatterplot) measures reliable voxel values across splits, and $\mathbf{u}_d^n = (\mathbf{u}_d^{(1)} - \mathbf{u}_d^{(2)})/\sqrt{2}$ (2nd PC of the scatterplot) measures uncorrelated, non-reproducible signal. This was used to obtain spatially Z-scored maps for each component as $\mathbf{z}_d = \mathbf{u}_d^{(e)} / \sigma_{\mathbf{u}_d^{(n)}}$, where $\sigma_{\mathbf{u}_d^{(n)}}$ is a single global noise estimate.

Representing functional connectivity

We employed the bootstrapping procedure defined above in “Reproducibility of the functional network model,” in order to visualize reliable functional connectivity values. Based on bootstrap ratio values, empirical P -values were obtained for elements of the partial correlation matrix \mathbf{P} , and between-component connectivity plots were then displayed for the different component models, after correcting for multiple comparisons using Bonferroni correction over all possible connections.

Simulation Data

To evaluate component models in the presence of ground truth, we simulated 2D “brain slices,” of dimensions 55×80 pixels. Slices consisted of 36 square functional parcels ranging in size from 5×5 to 20×20 pixels, where each parcel has relative scaling from 0 (edge) to 1 (center) defined by a cosine basis function. Data were generated for 20 simulated subjects, with 100 time-points per split-half. We used PARAFAC2 (“scaled consistent” model in Table I) as a generative model for simulation signal, as this can be converted into the other subspace models by altering parameter constraints and/or noise contributions.

For a fixed $K=20$ component subspace, spatial basis \mathbf{U} was generated by first producing 20 maps via random Gaussian weightings on the 36 parcels. An orthonormal basis \mathbf{U} was then obtained by computing an SVD on the maps and taking the resulting 20 eigen-images. Connectivity matrix \mathbf{R} was generated as a random normal matrix (with unit-normalized columns), and orthonormal subject basis \mathbf{Q} s obtained by generating random normal vectors and convolving with AFNI’s “SPMG1” function (afni.nimh.nih.gov/pub/dist/doc/program_help/3dDeconvolve.html; for $\text{TR} = 2.0\text{s}$) to simulate a smooth hemodynamic response.

Linear independence between vectors was enforced via Graham-Schmidt orthogonalization. Realistic network scaling matrices Σ_s were simulated by applying SVD to each experimental subject dataset (described in the next section), and taking the average 20-component eigenspectrum.

Noise and variability were simulated at three different levels of the model. At the voxel level, we added Gaussian noise to each pixel and time-point, with noise amplitude v_{noise} defined as the ratio of noise variance relative to signal variance, at each pixel. Between-split, within-subject network variability was simulated by adding uniformly-distributed noise to elements of Σ_s , sampled independently per split. Between-subject variability was also simulated, by adding the same uniformly-distributed noise to elements of Σ_s for both splits. For these variability measures, the sampling range is defined as the fraction of the largest eigenvalue v_{split} and v_{subject} . For example, $v_{\text{split}}=0.5$ indicates additive noise sampled from the interval $[0, 0.5\sigma_1]$ is added to each element of Σ_s . Results are shown for combinations of the parameter settings $v_{\text{noise}}=\{0.10, 0.50\}$, $v_{\text{split}}=\{0.00, 0.50\}$, and $v_{\text{subject}}=\{0.00, 0.50\}$, to explore model performance under different levels of noise and signal variability. We compared generalization error of the component models in Table I, as well as their ability to correctly identify optimal model order $K=20$.

We also examined the impact of spatial and temporal noise autocorrelation on model prediction, as a possible explanation for observed differences between experimental and simulated data. For spatial autocorrelation, each 2D slice image had its map of Gaussian noise smoothed by convolving with a 2D Gaussian kernel with full width at half maximum (FWHM) of ± 1 pix. For temporal autocorrelation, at each pixel the Gaussian noise time-series was convolved with the same “SPMG1” hemodynamic response function. These results are briefly discussed in Results section “Prediction of experimental whole-brain functional connectivity,” and depicted in Supporting Information Figure S1.

Experimental Data and Preprocessing

The component models were evaluated on a set of resting-state fMRI data that were used in a previous study [Andersen et al., 2014]. This dataset consisted of 30 healthy subjects (median age 45, range 22–69; 15 female). The Data were acquired using EPI (Echo-planar imaging), with $\text{TR} = 2.49$ s, $\text{TE} = 30$ ms, $\text{FA} = 90^\circ$; 42 slices, acquired in a 64×64 matrix, with 3mm isotropic voxels on a 3T (Siemens Magnetom Trio) MR scanner. For each subject run, 480 time points (~ 20 min of resting-state data) were acquired.

Data were preprocessed using the following steps in the SPM8 package (revision 5236): rigid-body motion correction to volume mean, slice-timing with sinc interpolation, normalization to an EPI template using affine warping and a DCT basis [Ashburner and Friston, 1999], and spatial smoothing with a isotropic 6 mm FWHM Gaussian kernel. We then corrected for head motion spikes using the

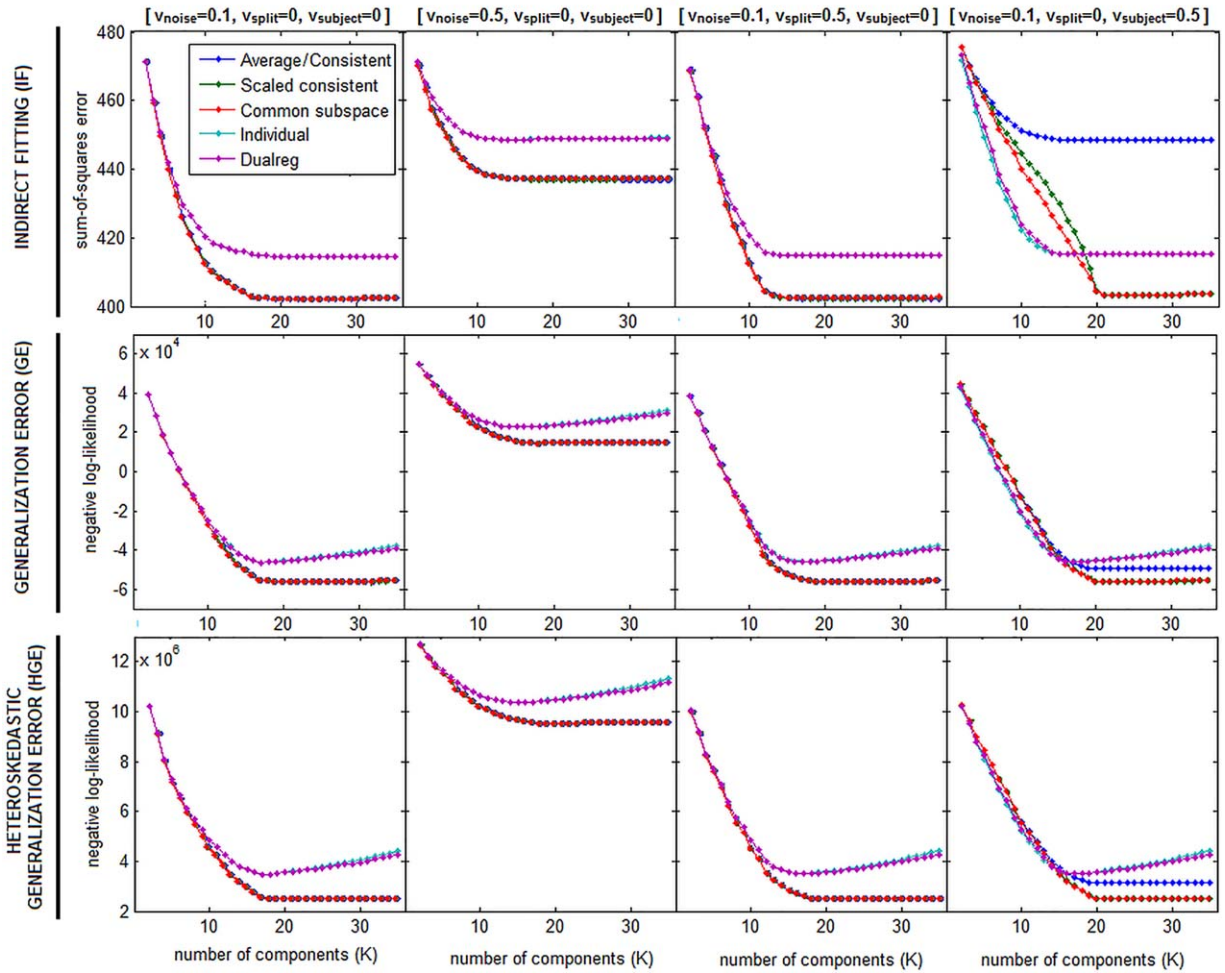


Figure 1.

Prediction error curves for simulated resting-state data. Average and Consistent models are combined due to near-identical error curves. Each column shows prediction values for a set of fixed simulation parameters, including voxel noise scaling V_{noise} , between-split network variability V_{split} , and between-subject network variability V_{subject} . [Color figure can be viewed at wileyonlinelibrary.com]

algorithm established in [Campbell et al., 2013]: we obtained a sliding-window estimate of signal change at each brain volume and identified outliers ($P < 0.05$, fitted gamma distribution). These volumes were replaced by interpolated values from adjacent runs, to avoid temporal discontinuities. We removed global signal by regressing out the first PC of each run [Carbonell et al., 2011], which is a moderately conservative approach that only removes global signal orthogonal to other BOLD fluctuations. We also performed motion parameter regression, by applying a PCA to the six rigid-body motion parameters [Woods et al., 1999] and removing the first two PCs (consistently account for $>85\%$ of variance), and we removed low-frequency drifts by regressing out third-order Legendre polynomial basis (chosen using AFNI's selection heuristic; refer to: afni.nimh.nih.gov/pub/dist/doc/program_help/

3dDetrend.html). Finally, we removed vascular, CSF and white matter signal using the PHYCAA+ algorithm [Churhill and Strother, 2013]. For analysis purposes, each subject's 480 time-point run was split into two equal-size blocks of 240 scans. All regression-based preprocessing was performed separately for the two splits, to maximize independence between training/test splits and to avoid potential processing-dependent model biases.

RESULTS AND DISCUSSION

Prediction of Simulated Whole-Brain Functional Connectivity

Figure 1 displays prediction curves of simulated data based on measures of IF, GE, and HGE, for all component

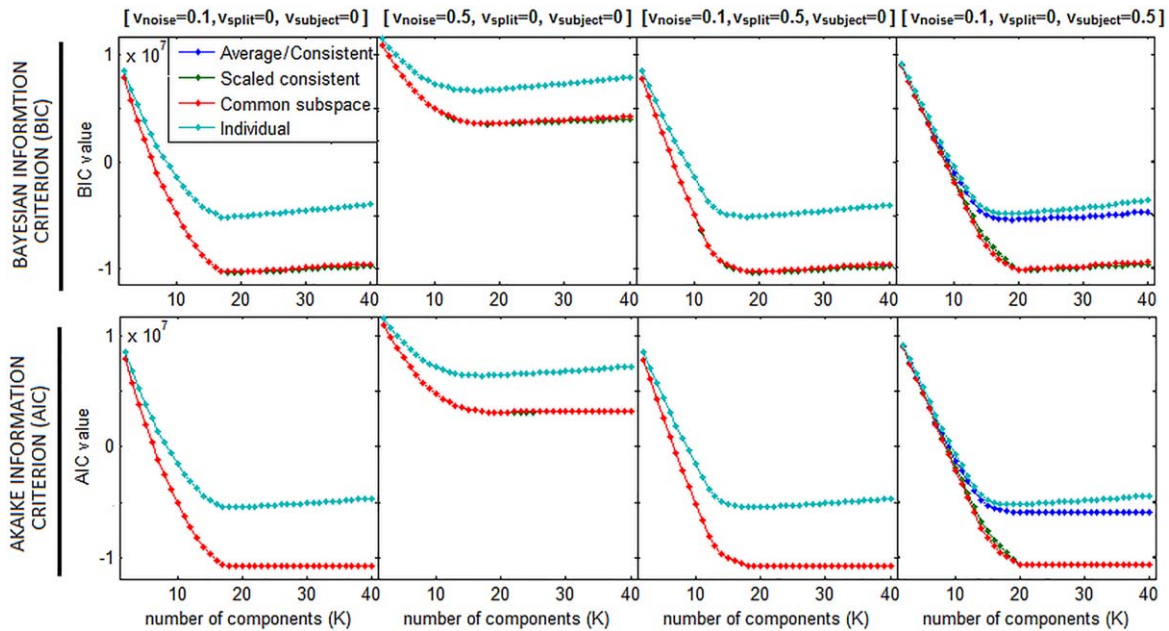


Figure 2.

Information criterion curves for simulated resting-state data, including Bayesian Information Criterion (BIC) and Akaike Information Criterion (AIC). *Average* and *Consistent* models are combined due to near-identical curves. Each column shows

information criterion values for a set of fixed simulation parameters, including voxel noise scaling v_{noise} , between-split network variability v_{split} and between-subject network variability v_{subject} . [Color figure can be viewed at wileyonlinelibrary.com]

models. For all simulations and prediction measures, error decreases until saturating near the true data dimensionality of $K=20$. Going from left to right, column-1 depicts low-noise data. All models have comparable performance at low K , but the highly flexible *Individual* and *Dualreg* models show consistently greater error at high K , whereas group-level models retain comparable generalization. Column-2 shows increased prediction error for higher voxel-wise noise ($v_{\text{noise}}=0.5$). The prediction curves now saturate at a lower K , indicating a tendency to underestimate model order in the presence of voxel-level noise. The IF metric is particularly sensitive to noise effects, saturating near $K=12$ components. In comparison, Column-3 shows that the effect of adding between-split network variability ($v_{\text{split}}=0.5$) has a relatively small effect on prediction error compared with voxel noise. Finally, Column-4 shows that adding inter-subject variability ($v_{\text{subject}}=0.5$) has the greatest impact on relative model performance. The constrained *Average* and *Consistent* models produce higher prediction error than the *Scaled Consistent* and *Common Subspace* models. In addition, *Individual* and *Dualreg* models now exhibit “cross-over” behavior, where they outperform group-level models at low K , but are consistently worse at higher K . The prediction measures diverge in one key respect: for IF, the *Average/Consistent* models exhibit higher error than *Individual* and *Dualreg* models, whereas the opposite is seen for GE and HGE measures.

Figure 2 plots information criterion curves for simulated data, showing similar curves for both BIC and AIC measures. In general, results are consistent with cross-validation measures, with all models saturating near the true $K=20$ dimensionality, although for group-level models the BIC begins to increase again beyond $K=20$. This is a consequence of BIC having a larger penalty weighting on the number of free parameters F_{IC} , compared to AIC.

As with cross-validation measures, higher voxel-wise noise ($v_{\text{noise}}=0.5$) tends to increase both BIC and AIC values and leads to under-estimation of model order, whereas adding between-split network variability ($v_{\text{split}}=0.5$) has a minimal effect on model evidence. The addition of inter-subject variability ($v_{\text{subject}}=0.5$) also increases between-model differences, with *Average/Consistent* models showing worse performance than *Scaled Consistent* and *Common Subspace* models. As expected, the probabilistic information criterion curves more closely resemble the probabilistic GE and HGE metrics, compared with the non-probabilistic IF measure.

Prediction of Experimental Whole-Brain Functional Connectivity

Figure 3 displays prediction curves of experimental data based on measures of IF, GE, and HGE, for all component models. Highly constrained *Average/Consistent* models

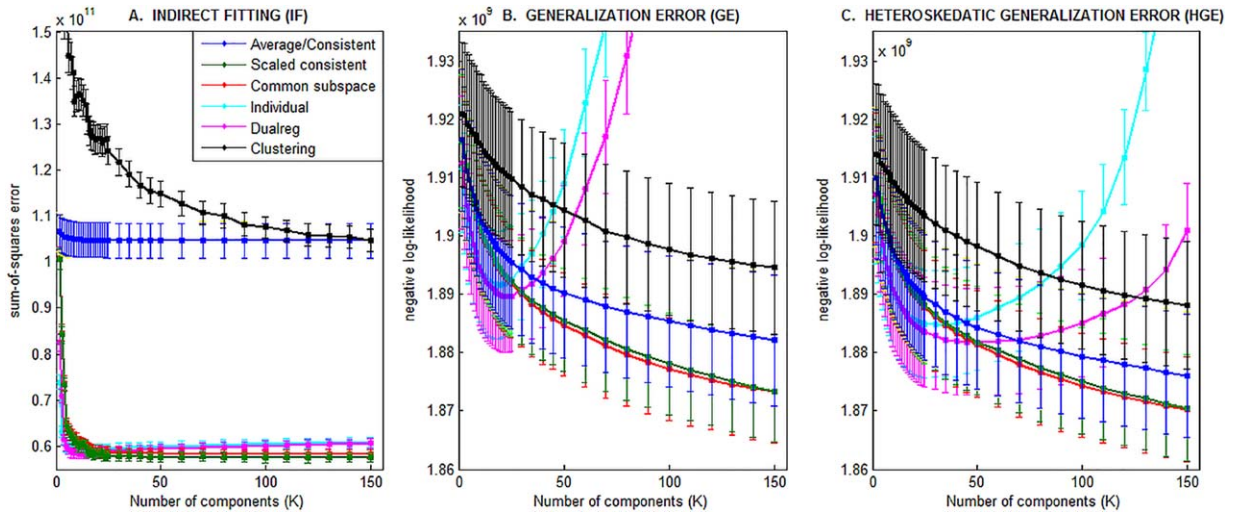


Figure 3.

Prediction error curves for experimental resting-state data, along with bootstrapped standard error bars. *Average* and *Consistent* models are combined due to near-identical error curves. [Color figure can be viewed at wileyonlinelibrary.com]

show consistently greater error than the more flexible *Scaled Consistent* and *Common Subspace* models. The latter two models show consistently optimal prediction, where *Scaled Consistent* has lowest error for the IF metric, and *Common Subspace* has lowest error for GE and HGE metrics. The *Clustering* model shows consistently higher error than all group-level component models. Conversely, the

highly flexible *Individual* and *Dualreg* models exhibit cross-over behavior seen in simulations (Fig. 1), where they outperform group-level models at low K , but are consistently worse at higher K . The *Dualreg* approach shows the benefits of group-level component estimation, as it has consistently better prediction than *Individual*. Thus, the *Scaled Consistent* and *Common Subspace*

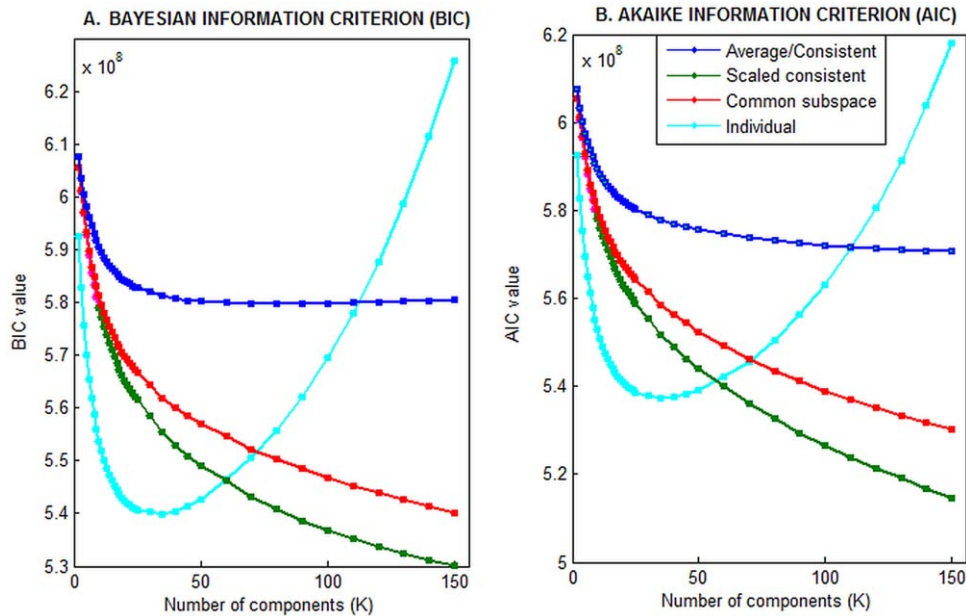


Figure 4.

Information criterion curves for simulated resting-state data, including Bayesian Information Criterion (BIC) and Akaike Information Criterion (AIC). *Average* and *Consistent* models are combined due to near-identical curves. [Color figure can be viewed at wileyonlinelibrary.com]

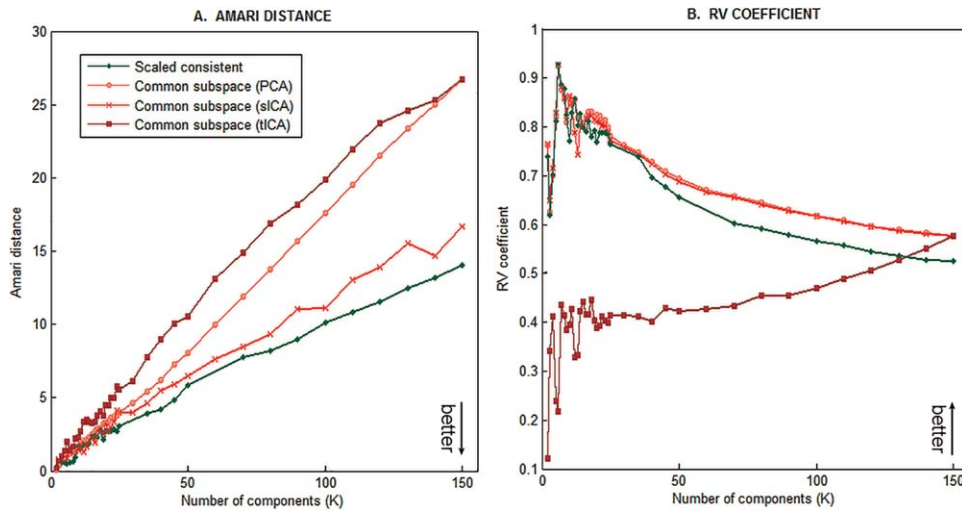


Figure 5.

Reproducibility of spatial component bases for the optimal group-level models, identified based on generalization error (Fig. 3) and information criteria (Fig. 4), including *Scaled Consistent* and *Common Subspace* models. For the *Common Subspace* approach, we also show different component representations including Principal Component Analysis (PCA), and temporal

and spatial Independent Component Analysis (tICA and sICA, respectively). As indicated by the arrow in the bottom of each plot, note that for the Amari index lower values indicate better reproducibility, whereas higher values are associated with better reproducibility for the RV metric. [Color figure can be viewed at wileyonlinelibrary.com]

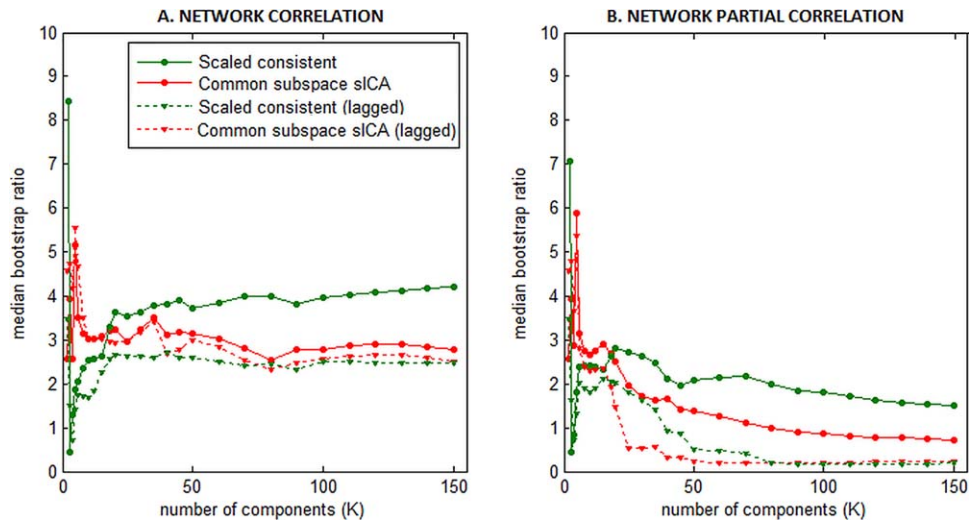


Figure 6.

Bootstrapped reliability of between-component functional connectivity matrices for the optimal group-level models, identified based on generalization error (Fig. 3) and information criteria (Fig. 4), including *Scaled Consistent* and *Common Subspace* models. For the *Common Subspace* method, we depict spatial Independent Component Analysis (sICA), as it is the only tested method that does not assume temporal independence between

components. Lagged connectivity matrices are also estimated by maximizing subject-specific connectivity over a ± 2 TR window. Results are shown for both (A) correlation and (B) partial correlation matrices. For each model, median (absolute) bootstrapped reliability is measured over all connectivity elements, as a function of model order K. [Color figure can be viewed at wileyonlinelibrary.com]

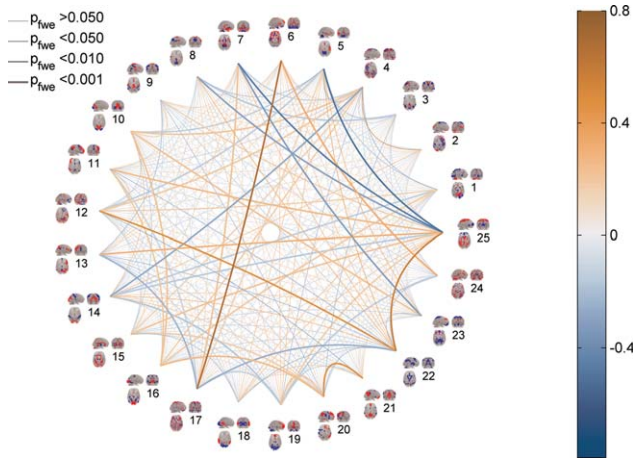


Figure 7.

Scaled Consistent (PARAFAC2) spatial components and network connectivity. The figure shows the spatial components. Each map is displayed as sagittal, coronal and axial projections in neurological convention and are arranged in a circle ordered. For visualization the maps were thresholded to retain the 5% of voxels with largest absolute magnitude, positive super-threshold voxels are displayed in red, whereas negative super-threshold voxels are displayed in blue. The lines that connect each spatial map represent partial correlations as indicated by the colorbar to the right, and the thickness of each line indicate the statistical significance of the connection calculated based bootstrap ratios of each correlation value and corrected for multiple comparisons over all connections by Bonferroni correction. [Color figure can be viewed at wileyonlinelibrary.com]

approaches tend to provide optimal prediction of functional connectivity.

In general, experimental results resembled simulation curves with between-subject network variability (Fig. 1, Column-4). However, unlike simulated data, models do not consistently saturate for experimental data, within the tested range of 1–150 components. While IF consistently saturates for all models by $K=25$, only highly flexible *Individual* and *Dualreg* models show saturation of GE and HGE within the tested range of components. In addition, for experimental data, *Individual* and *Dualreg* models appear much more heavily penalized for GE and HGE metrics at high K . Because these metrics are based on an assumption of independent, identically distributed noise, the discrepancies may be due to an inaccurate error model. In Supporting Information Figure S1, the random noise model (Supporting Information Fig. S1A) is compared with models with spatially and temporally autocorrelated noise. Supporting Information Figure S1B shows that for spatially smooth noise, GE and HGE fail to saturate at the correct $K=20$ dimensionality, for group-level models. Conversely, Supporting Information Figure S1C shows that for temporally smooth noise, GE and HGE are much larger for *Individual* and *Dualreg* models. These results

demonstrate tradeoffs between the different prediction models: IF (which assumes no error model) is more robust to the effects of noise autocorrelation, whereas GE and HGE are less prone to under-estimating dimensionality in the presence of noise (shown in Fig. 1).

In Figure 4, we plot the information criterion curves for experimental data. As with the simulations, both BIC and AIC curves are similar to probabilistic generalization error curves GE and HGE of Figure 3. The highly constrained *Average* and *Consistent* models show highest error for group-level models, followed by *Scaled consistent*, and *Common subspace* models. In addition, the flexible *Individual* model shows initially high generalization, but quickly becomes worse than group-level models with increasing K . However, two distinctions are seen from cross-validated error. The first is that *Scaled consistent* and *Common subspace* models show a greater difference in performance, whereas they are near-indistinguishable for GE and HGE measures. The second difference is that *Individual* models show optimal BIC and AIC for up to $K=50$ components, whereas they become sub-optimal at $K=30$ for GE and HGE. Finally, we note that the shapes of BIC and AIC curves are generally similar, with the exception of *Average* and *Consistent* models that saturate at approximately $K=60$ for the BIC metric. This indicates that the choice of information criterion has a minimal impact when evaluating models at the tested range of K .

Reliability of Model Parameters

Figure 5 plots reproducibility of the spatial component bases for the *Scaled Consistent* (i.e., PARAFAC2) and *Common Subspace* (i.e., PCA and ICA) models, which have

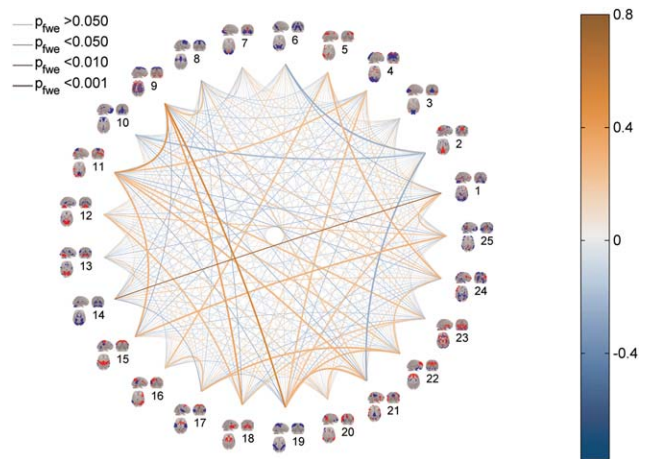
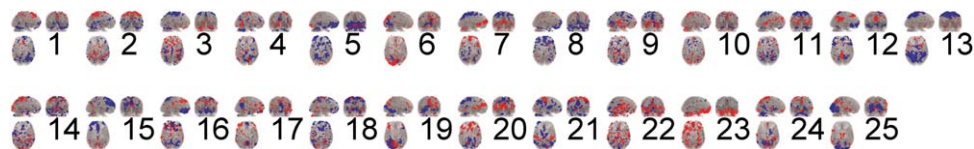


Figure 8.

Spatial ICA components and network connectivity. This figure displays the components extracted by spatial ICA based on the *Common Subspace* model, the formatting of the figure is consistent with Figure 7. [Color figure can be viewed at wileyonlinelibrary.com]

**Figure 9.**

Temporal ICA components. The figure displays the spatial components extracted by temporal ICA based on the *Common Subspace* model. Spatial maps are displayed and thresholded consistently with Figures 7 and 8, but here the maps are arranged in two lines because the orthogonal property of the model does not enable estimation of between component correlations. [Color figure can be viewed at wileyonlinelibrary.com]

lowest prediction error in Figure 2. For both reproducibility metrics, we see comparable performance between component models for low dimensionalities of $K < 25$, but beyond this point they diverge in performance. The AD metric shows optimal performance for PARAFAC2 followed by sICA, PCA, then tICA. As this metric is highly sensitive to pairwise matching of components, this indicates that PARAFAC2 provides more stable individual component patterns at high model dimensionalities. Conversely, the RV coefficient shows near-identical performance between PCA and sICA, but lower performance for PARAFAC2, particularly at higher subspace dimensionalities. This indicates that overall subspace overlap between independent split-halves is higher for the *Common Subspace* models.

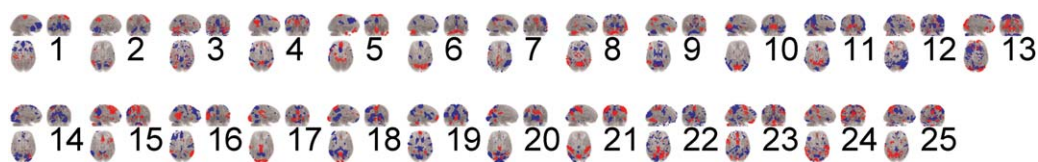
Figure 6 depicts the reliability of the between-component functional connectivity estimates. We compared *Scaled consistent* against the sICA *Common Subspace* approach, since PCA and tICA models assume temporal independence between components. Comparisons included both standard network model and the lagged models of maximal connectivity. For fixed connectivity matrices, *Scaled Consistent* consistently estimates more reliable connectivity values, for both correlation (Fig. 6A) and partial correlation (Fig. 6B) estimates. However, the reliability of *Scaled Consistent* decreases for lagged estimation, becoming comparable with the *Common Subspace* model. This we attribute to the fact that the PARAFAC2 model specifically optimizes for consistency of the zero-lag consistency of the group covariance, while ICA does not explicitly consider the group covariance.

Figures 7–10 display both spatial component patterns and functional connectivity relationships between

components, based on partial correlation estimates; results are shown for the optimal $K=25$ model dimensionality (estimated from the IF prediction metric). For visualization, each of the spatial maps were thresholded to retain the 5% voxels that showed the largest absolute values. For temporally-independent PCA and tICA models, no between-component connectivity structure can be displayed, as the between-component correlations are uniformly zero. Both the *scaled consistent* (PARAFAC2) and *common subspace* (sICA) models show numerous highly-significant partial correlations between components.

The spatial patterns for the PCA and tICA model are difficult to interpret due to the global and distributed nature of the components. Both the sICA and PARAFAC2 models identify several components that are restricted within certain brain regions, allowing easier interpretation. Quantitatively, the sICA model appears to be more interpretable in that it generally captures more spatially sparse brain regions, whereas components estimated by PARAFAC2 model are less interpretable due to less skewed component distributions (i.e. more mixed sign in the components). This was anticipated due to the super-Gaussian source distribution of the sICA model, which enforces spatially sparse and thereby interpretable components.

Figure 11 shows a comparison between matched components for the *Scaled consistent* and sICA *Common subspace* models. We identified components reflecting the default-mode network, primary visual network and motor network which had the highest absolute correlation with spatial reference templates. For the default-mode network we used the template from [Franco et al., 2009] provided in the Group ICA toolbox (<http://mialab.mrn.org/software/>)

**Figure 10.**

PCA components. This figure displays spatial components extracted by the *common subspace* PCA model, formatting is consistent with Figure 9. [Color figure can be viewed at wileyonlinelibrary.com]

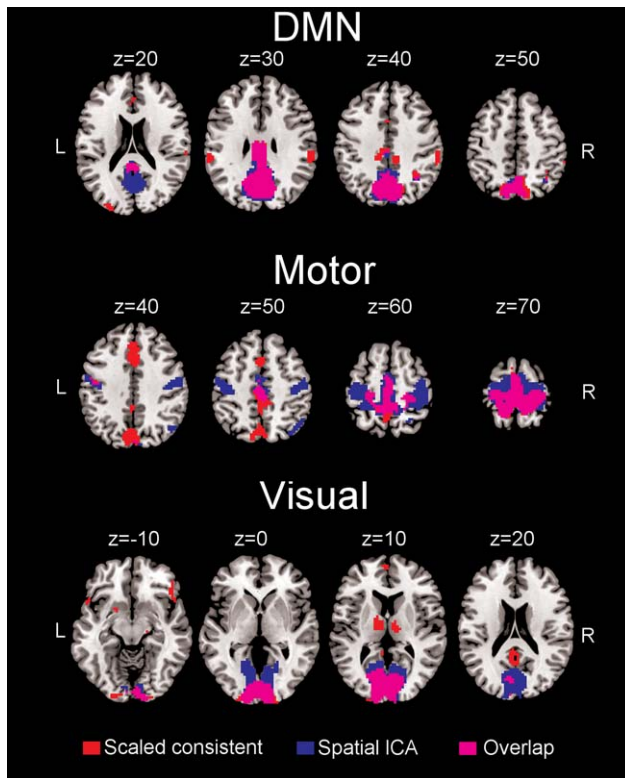


Figure 11.

Comparison of *Scaled consistent* and sICA *Common subspace* components. This figure shows selected axial slices of spatial components matched by absolute spatial correlation across *Scaled consistent* (PARAFAC2) and sICA models. The top row (marked “DMN”) displays the network showing the largest correlation with the default-mode template, whereas the second row (marked “Motor”) displays the component showing the largest correlation with the template motor template. The bottom row (marked “Visual”) shows the component exhibiting the largest correlation with the visual template. Spatial maps are displayed in neurological intervention, as marked by the L-Left and R-Right markers, at axial MNI-space slices according to the displayed z-coordinate. Spatial maps are thresholded at a Z-score of 3, red indicates that the region is above threshold in the *Scaled consistent* model, whereas blue indicates that the region was above threshold for the sICA model. Purple indicates that both models were above threshold. [Color figure can be viewed at wileyonlinelibrary.com]

gift), whereas the motor templates were created by combining Brodmann areas 4a and 4p [Geyer et al., 1996] provided in the SPM Anatomy toolbox [Eickhoff et al., 2005]. Similarly, the visual template was created by combining Brodmann areas 17 and 18 [Amunts et al., 2000]. All three matched components show some overlap indicating that they likely reflect similar networks. However, as also stated above the components identified using sICA appear to be slightly more sparse, leading to easier interpretation.

CONCLUSION

We have established a cross-validation framework for comparing different component subspace models in group fMRI data, which predicts between-voxel functional connectivity relationships. This framework was used to analyze five different component models, with varying levels of flexibility in parameterizing subject-specific variability. The *Average* model ensures that all subject covariance matrices are approximated by the same spectral decomposition, whereas for the *Consistent* model the covariance between the extracted component time-series is consistent across subjects. The *Scaled consistent* model relaxes this assumption by allowing scaling of the components across subjects. For the *Common subspace* model, the covariance between components is individually defined for each subject forming temporal concatenated group PCA/ICA models. Finally, in the *Individual* models each subject is modeled by a separate spectral decomposition (per subject PCA/ICA).

We demonstrated consistent results for three different predictive measures (IF, GE, and HGE), including uniformly poor generalization for highly constrained *Average* and *Consistent* models. We attribute this poor generalization to substantial inter-subject variability, which these models were unable to account for. The superior generalization of the more flexible *Scaled consistent* and *Common subspace* models indicates that inter-subject differences in covariance are reliable across independent data splits, and can thus be consistently estimated.

In contrast, the highly flexible *Individual* model was optimal only for small dimensionalities, and rapidly increased in error with number of estimated components K . This is due to its inability to share information about covariance structure across subjects, producing a noisy, over-fitted structure with poor generalization. Moreover, the dual regression (*Dualreg*) approach which uses group-level spatial components to estimate subject-specific time-courses, improves prediction, but remains highly similar to the *Individual* model in performance. We attribute the lack of improved predictability using dual regression to overfitting, due to the additional parameters added in the dual regression step. Improved reproducibility of spatial maps was observed by [Zuo et al., 2010] when using dual regression. However, we note that our prediction metrics are not directly comparable to the reproducibility of the spatial networks as measured by spatial correction. In addition, a limitation of our within session splitting procedure is that the long resting state sessions (20 min) might be associated with limited reproducibility, whereas [Zuo et al., 2010] used three separate sessions.

With respect to the prediction metrics investigated, the best performing models were the *Common subspace* and *Scaled consistent* models, which were both able to account for subject variability, while exploiting shared structure across subjects. This supports the well-established group-PCA (and equivalently, group-ICA) approaches as viable

models for multi-subject fMRI data. We also demonstrated no significant difference in generalization between the *Common subspace* model (group PCA/ICA) and the constrained *Scaled consistent* model (PARAFAC2). Therefore, we can define a consistent connectivity model across subjects with strong generalization, simply by accounting for subject variability in terms of relative scaling of the components.

Our results also show that the *Scaled consistent* and *Common subspace* models, which have comparable prediction, have significant tradeoffs in stability (and interpretability) of model parameters. While the *Common subspace* model produced greater overlap of spatial subspaces, (measured by the RV coefficient), the *Scaled consistent* model maximized the stability of individual spatial component patterns (measured via AD) and extracted the most consistent subspace network structure C_s . Thus, the added constraints of consistency in PARAFAC2 makes component estimation more stable across independent data splits, potentially providing more viable interpretations of individual component maps and their interactions over less constrained common subspace models. We also expect that reproducibility may be improved by further constraining common subspace modeling approaches, either by regularizing using prior information such as functional templates [Schultz et al., 2014] and atlases [Dhillon et al., 2014], or by imposing additional constraints on inter-subject dependence, as in independent vector analysis [Lee et al., 2008] or group information guided ICA [Du and Fan, 2013]. Similarly, we expect the performance of the *Individual* models may be improved using regularization such as penalizing the deviation from a common subspace [Varoquaux et al., 2011].

When comparing the extracted spatial maps of the group-ICA model to the group-PCA and scaled consistent (PARAFAC2) models, we find that spatial ICA provides sparser, more localized components. This is more interpretable in terms of connectivity relationships between brain regions, compared with group-PCA, PARAFAC2 and temporal ICA. The latter models result in a mixture of highly global and highly sparse components, making it more difficult to interpret brain regions implicated in network interactions. We displayed components for a dimensionality of $K = 25$ (i.e., optimizing the IF metric), as the whole component space can be directly visualized in Figures 7–10. At this dimensionality, the individual components contain several non-contiguous regions and may be regarded as networks themselves [Kiviniemi et al., 2003]. In the analysis of resting-state fMRI data, a much higher number of components are often estimated using group ICA, which typically creates component maps with small, isolated brain regions. Although this enables easier interpretation of network interactions [Kiviniemi et al., 2009], our results in Figures 5 and 6 indicate that both the component maps and functional connectivity values become significantly less stable.

To our knowledge, this is the first article to systematically evaluate generalization and reproducibility of component models for multi-subject fMRI analysis. We believe that this validation framework is a powerful approach for investigating component models in multi-subject fMRI data. Current results point toward group-PCA/ICA and PARAFAC2 as viable frameworks for modeling subject variability, while at the same time identifying consistent generalizable structure. In particular, it seems the more constrained scaled consistent representation offered by the PARAFAC2 modeling framework provides both high model prediction, and the most reproducible results. One remaining issue for PARAFAC2 is component interpretation, which is hampered by the lack of localization, when compared with spatial group ICA. Further research is required to establish extensions of this framework that incorporate spatial constraints, without compromising prediction and network reproducibility. In the current manuscript we considered only resting-state fMRI data, which was a deliberate choice as the models investigated make no assumption about generalization of time-series across subjects, making them particularly suitable for analysis of resting-state fMRI data. Nonetheless, it is an important area of research to investigate whether the studied models are also able to successfully extract task-related patterns of brain activity.

Robust quantification of whole-brain functional connectivity in multi-subject data is important, to accurately and reliably determine intrinsic functional network structure in the brain. We believe that the proposed framework can be used to quantify how well component models account for both group-level and subject-specific covariance relationships between brain regions in multi-subject fMRI datasets. This also has general importance for analyses of covariance in the brain, including sparse inverse estimation [Marrelec et al., 2006; Smith et al., 2010] and time-evolving dynamics [Hutchison et al., 2013].

ACKNOWLEDGMENTS

The Magnetom Trio MR scanner was donated by the Simon Spies Foundation.

REFERENCES

- Achard S, Salvador R, Whitcher B, Suckling J, Bullmore E (2006): A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs. *J Neurosci* 26:63–72.
- Akaike H (1974): A new look at the statistical model identification. *IEEE Trans Automat Contr* 19:716–723.
- Amari S, Cichocki A, Yang HH (1996): A new learning algorithm for blind signal separation. In: Touretzky DS, Mozer MC, and Hasselmo ME, editors. *Proceedings of the 8th International Conference on Neural Information Processing Systems (NIPS'95)*, MIT Press, Cambridge, MA, USA, pp 757–763.

- Amunts K, Malikovic A, Mohlberg H, Schormann T, Zilles K (2000): Brodmann's areas 17 and 18 brought into stereotaxic space-where and how variable?. *Neuroimage* 11:66–84.
- Andersen KW, Madsen KH, Siebner HR, Schmidt MN, Mørup M, Hansen LK (2014): Non-parametric Bayesian graph models reveal community structure in resting state fMRI. *Neuroimage* 100:301–315.
- Arbabshirani MR, Kiehl K. a, Pearlson GD, Calhoun VD (2013): Classification of schizophrenia patients based on resting-state functional network connectivity. *Front Neurosci* 7:133.
- Ashburner J, Friston KJ (1999): Nonlinear spatial normalization using basis functions. *Hum Brain Mapp* 7:254–266.
- Bassett DS, Bullmore E (2006): Small-world brain networks. *Neuroscientist* 12:512–523.
- Beckmann CF, Smith SM (2005): Tensorial extensions of independent component analysis for multisubject fMRI analysis. *NeuroImage* 25:294–311.
- Bell AJ, Sejnowski TJ (1995): An information maximization approach to blind source separation and blind deconvolution. *Neural Comput* 7:1129–1159.
- Bentler PM, Bonett DG (1980): Significance tests and goodness of fit in the analysis of covariance structures. *Psychol Bull* 88:588.
- Berge J, Kiers H (1996): Some uniqueness results for PARAFAC2. *Psychometrika* 61:123–132.
- Biswal B, Yetkin FZ, Haughton VM, Hyde JS (1995): Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magn Reson Med* 34:537–541.
- Biswal BB, Mennes M, Zuo X-N, Gohel S, Kelly C, Smith SM, Beckmann CF, Adelstein JS, Buckner RL, Colcombe S, Dogonowski A-M, Ernst M, Fair D, Hampson M, Hoptman MJ, Hyde JS, Kiviniemi VJ, Kötter R, Li S-J, Lin C-P, Lowe MJ, Mackay C, Madden DJ, Madsen KH, Margulies DS, Mayberg HS, McMahon K, Monk CS, Mostofsky SH, Nagel BJ, Pekar JJ, Peltier SJ, Petersen SE, Riedl V, Rombouts SARB, Rypma B, Schlaggar BL, Schmidt S, Seidler RD, Siegle GJ, Sorg C, Teng G-J, Vejjola J, Villringer A, Walter M, Wang L, Weng X-C, Whitfield-Gabrieli S, Williamson P, Windischberger C, Zang Y-F, Zhang H-Y, Castellanos FX, Milham MP (2010): Toward discovery science of human brain function. *Proc Natl Acad Sci U S A* 107:4734–4739.
- Bro R, Andersson CA, Kiers HAL (1999): PARAFAC2-Part II. Modeling chromatographic data with retention time shifts. *J Chemom* 13:295–309.
- Calhoun V, Adali T, Pearlson G, Pekar J (2001a): A method for making group inferences using independent component analysis of functional MRI data: Exploring the visual system. *Neuroimage* 13:S88.
- Calhoun VD, Adali T, Pearlson GD, Pekar JJ (2001b): A method for making group inferences from functional MRI data using independent component analysis. *Hum Brain Mapp* 14:140–151.
- Calhoun VD, Liu J, Adali T (2009): A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data. *Neuroimage* 45:S163–S172.
- Campbell KL, Grigg O, Saverino C, Churchill N, Grady CL (2013): Age differences in the intrinsic functional connectivity of default network subsystems. *Front Aging Neurosci* 5:73.
- Carbonell F, Bellec P, Shmuel A (2011): Global and system-specific resting-state fMRI fluctuations are uncorrelated: Principal component analysis reveals anti-correlated networks. *Brain Connect* 1:496–510.
- Churchill NW, Strother SC (2013): PHYCAA+: an optimized, adaptive procedure for measuring and controlling physiological noise in BOLD fMRI. *Neuroimage* 82:306–325.
- Cohen AL, Fair DA, Dosenbach NUF, Miezin FM, Dierker D, Van Essen DC, Schlaggar BL, Petersen SE (2008): Defining functional areas in individual human brains using resting functional connectivity MRI. *Neuroimage* 41:45–57.
- Craddock RC, James GA, Holtzheimer PE, Hu XP, Mayberg HS (2012): A whole brain fMRI atlas generated via spatially constrained spectral clustering. *Hum Brain Mapp* 33:1914–1928.
- Damoiseaux JS, Rombouts SARB, Barkhof F, Scheltens P, Stam CJ, Smith SM, Beckmann CF (2006): Consistent resting-state networks across healthy subjects. *Proc Natl Acad Sci U S A* 103:13848–13853.
- Dhillion PS, Wolk DA, Das SR, Ungar LH, Gee JC, Avants BB (2014): Subject-specific functional parcellation via prior based eigenanatomy. *Neuroimage* 99:14–27.
- Du Y, Fan Y (2013): Group information guided ICA for fMRI data analysis. *Neuroimage* 69:157–197.
- Efron B, Tibshirani R (1986): Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy. *Stat Sci* 1:54–75.
- Eickhoff SB, Stephan KE, Mohlberg H, Grefkes C, Fink GR, Amunts K, Zilles K (2005): A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *NeuroImage* 25:34.
- Escouffier Y (1973): Le Traitement des Variables Vectorielles. *Biometrics* 29:751–760.
- Esposito F, Scarabino T, Hyvarinen A, Himberg J, Formisano E, Comani S, Tedeschi G, Goebel R, Seifritz E, Di Salle F (2005): Independent component analysis of fMRI group studies by self-organizing clustering. *Neuroimage* 25:193–205.
- Ferdowsi S, Abolghasemi V, Sanei S (2013): EEG-fMRI integration using a partially constrained tensor factorization. In: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing. pp 6191–6195.
- Filippini N, MacIntosh BJ, Hough MG, Goodwin GM, Frisoni GB, Smith SM, Matthews PM, Beckmann CF, Mackay CE (2009): Distinct patterns of brain activity in young carriers of the APOE-epsilon4 allele. *Proc Natl Acad Sci U S A* 106:7209–7214.
- Fox MD, Snyder AZ, Vincent JL, Corbetta M, Van Essen DC, Raichle ME (2005): The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proc Natl Acad Sci U S A* 102:9673–9678.
- Franco AR, Pritchard A, Calhoun VD, Mayer AR (2009): Interrater and intermethod reliability of default mode network selection. *Hum Brain Mapp* 30:2293–2303.
- Friston KJ, Frith CD, Liddle PF, Frackowiak RS (1993): Functional connectivity: The principal-component analysis of large (PET) data sets. *J Cereb Blood Flow Metab* 13:5–14.
- Geyer S, Ledberg A, Schleicher A, Kinomura S, Schormann T, Bürgel U, Klingberg T, Larsson J, Zilles K, Roland PE (1996): Two different areas within the primary motor cortex of man. *Nature* 382:805–807.
- Goutte C, Toft P, Rostrup E, Nielsen F, Hansen LK (1999): On clustering fMRI time series. *Neuroimage* 9:298–310.
- Greicius MD, Krasnow B, Reiss AL, Menon V (2003): Functional connectivity in the resting brain: A network analysis of the default mode hypothesis. *Proc Natl Acad Sci U S A* 100:253–258.
- Guo Y, Pagnoni G (2008): A unified framework for group independent component analysis for multi-subject fMRI data. *Neuroimage* 42:1078–1093.
- Hansen LK, Larsen J, Nielsen FA, Strother SC, Rostrup E, Savoy R, Lange N, Sidtis J, Svarer C, Paulson OB (1999):

- Generalizable patterns in neuroimaging: How many principal components?. *Neuroimage* 9:534–544.
- Harshman RA (1972): PARAFAC2: Mathematical and technical notes. *UCLA Work Pap Phonetics* 22:30–47.
- Harshman RA, Lundy ME, Richard A, Harshman MEL (1996): Uniqueness proof for a family of models sharing features of Tucker's three-mode factor analysis and PARAFAC/candcomp. *Psychometrika* 61:133–154.
- Himberg J, Hyvärinen A, Esposito F (2004): Validating the independent components of neuroimaging time series via clustering and visualization. *Neuroimage* 22:1214–1222.
- Hutchison RM, Womelsdorf T, Allen EA, Bandettini PA, Calhoun VD, Corbetta M, Della Penna S, Duyn JH, Glover GH, Gonzalez-Castillo J, Handwerker DA, Keilholz S, Kiviniemi V, Leopold DA, de Pasquale F, Sporns O, Walter M, Chang C (2013): Dynamic functional connectivity: Promise, issues, and interpretations. *Neuroimage* 80:360–378.
- Hyvärinen A (1999): Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans Neural Netw* 10: 626–634.
- Kiers HAL (1993): An alternating least squares algorithm for PARAFAC2 and three-way DEDICOM. *Comput Stat Data Anal* 16: 103–118.
- Kiers HAL, Ten Berge JMF, Bro R (1999): PARAFAC2-Part I. A direct fitting algorithm for the PARAFAC2 model. *J Chemom* 13:275–294.
- Kiviniemi V, Kantola J-H, Jauhiainen J, Hyvärinen A, Tervonen O (2003): Independent component analysis of nondeterministic fMRI signal sources. *Neuroimage* 19:253–260.
- Kiviniemi V, Starck T, Remes J, Long X, Nikkinen J, Haapea M, Veijola J, Moilanen I, Isohanni M, Zang Y-F, Tervonen O (2009): Functional segmentation of the brain cortex using high model order group PICA. *Hum Brain Mapp* 30:3865–3886.
- Ledoit O, Wolf M (2004): A well-conditioned estimator for large-dimensional covariance matrices. *J Multivar Anal* 88:365–411.
- Lee JH, Lee TW, Jolesz FA, Yoo SS (2008): Independent vector analysis (IVA): multivariate approach for fMRI group study. *Neuroimage* 40:86–109.
- Lowe MJ, Mock BJ, Sorenson JA (1998): Functional connectivity in single and multislice echoplanar imaging using resting-state fluctuations. *Neuroimage* 7:119–132.
- MacCallum RC (1996): PARAFAC2: Mathematical and technical notes. *Psychol Methods* 1:130.
- Marrelec G, Krainik A, Duffau H, Péligrini-Issac M, Lehericy S, Doyon J, Benali H (2006): Partial correlation for functional brain interactivity investigation in functional MRI. *Neuroimage* 32:228–237.
- McKeown MJ, Jung TP, Makeig S, Brown G, Kindermann SS, Lee TW, Sejnowski TJ (1998): Spatially independent activity patterns in functional MRI data during the stroop color-naming task. *Proc Natl Acad Sci U S A* 95:803–810.
- Robert P, Escoufier Y, Abdi H (2007): RV coefficient and congruence coefficient. *Encycl Meas Stat* 849–853.
- Schmithorst VJ, Holland SK (2004): Comparison of three methods for generating group statistical inferences from independent component analysis of functional magnetic resonance imaging data. *J Magn Reson Imaging* 19:365–368.
- Schultz AP, Chhatwal JP, Huijbers W, Hedden T, van Dijk KRA, McLaren DG, Ward AM, Wigman S, Sperling RA (2014): Template based rotation: A method for functional connectivity analysis with a priori templates. *Neuroimage* 102 Pt 2:620–636.
- Schwarz G (1978): Estimating the dimension of a model. *Ann Stat* 6:461–464.
- Smith SM, Miller KL, Salimi-Khorshidi G, Webster M, Beckmann CF, Nichols TE, Ramsey JD, Woolrich MW (2010): Network modelling methods for FMRI. *Neuroimage* 54:875–891.
- Strother SC, Anderson J, Hansen LK, Kjems U, Kustra R, Sidtis J, Frutiger S, Muley S, LaConte S, Rottenberg D (2002): The quantitative evaluation of functional neuroimaging experiments: The NPAIRS data analysis framework. *Neuroimage* 15:747–771.
- Thirion B, Varoquaux G, Dohmatob E, Poline J-B (2014): Which fMRI clustering gives good brain parcellations?. *Front Neurosci* 8:167.
- Uhlig H (1994): On Singular Wishart and Singular Multivariate Beta Distributions. *Source Ann Stat Ann Stat* 22:395–405.
- Varoquaux G, Gramfort A, Pedregosa F, Michel V, Thirion B (2011): Multi-subject dictionary learning to segment an atlas of brain spontaneous activity. *Inf Process Med Imaging* 22:562–573.
- Weis M, Jannek D, Roemer F, Guenther T, Haardt M, Husar P (2010): Multi-dimensional PARAFAC2 component analysis of multi-channel EEG data including temporal tracking. *Conf Proc. Annu Int Conf IEEE Eng Med Biol Soc IEEE Eng Med Biol Soc Annu Conf*. pp 5375–5378.
- Woods RP, Dapretto M, Sicotte NL, Toga AW, Mazziotta JC (1999): Creation and use of a Talairach-compatible atlas for accurate, automated, nonlinear intersubject registration, and analysis of functional imaging data. *Hum Brain Mapp* 8:73–79.
- Zuo X-N, Kelly C, Adelstein JS, Klein DF, Castellanos FX, Milham MP (2010): Reliable intrinsic connectivity networks: Test-retest evaluation using ICA and dual regression approach. *Neuroimage* 49:2163–2177.