

Structurally-informed Bayesian functional connectivity analysis

M. Hinne^{a,b}, L. Ambrogioni^a, R. J. Janssen^a, T. Heskes^b, M. A. J. van Gerven^a

^a*Radboud University Nijmegen, Donders Institute for Brain, Cognition and Behaviour, Nijmegen, The Netherlands*

^b*Radboud University Nijmegen, Institute for Computing and Information Sciences, Nijmegen, The Netherlands*

Abstract

Functional connectivity refers to covarying activity between spatially segregated brain regions and can be studied by measuring correlation between functional magnetic resonance imaging (fMRI) timeseries. These correlations can be caused either by direct communication via active axonal pathways or indirectly via the interaction with other regions. It is not possible to discriminate between these two kinds of functional interaction simply by considering the covariance matrix. However, the non-diagonal elements of its inverse, the precision matrix, can be naturally related to direct communication between brain areas and interpreted in terms of partial correlations. In this paper, we propose a Bayesian model for functional connectivity analysis which allows estimation of a posterior density over precision matrices, and, consequently, allows one to quantify the uncertainty about estimated partial correlations. In order to make model estimation feasible it is assumed that the sparseness structure of the precision matrices is given by an estimate of structural connectivity obtained using diffusion imaging data. The model was tested on simulated data as well as resting-state fMRI data and compared with a graphical lasso analysis. The presented approach provides a theoretically solid foundation for quantifying functional connectivity in the presence of uncertainty.

Keywords: functional connectivity, structural connectivity, G -Wishart prior, Bayesian inference

1. Introduction

One of the oldest and most influential ideas in cognitive neuroscience is that the brain, and in particular the cortex, can be divided into specialized functional regions (Friston, 2011). In recent times, the neuroscience community has become increasingly interested in determining how these regions are organized as large functional networks and how their modulation reflects ongoing cognitive processing (Bullmore and Sporns, 2009). The organization of these functional networks can be described using the umbrella term ‘functional connectivity’, defined as the deviations from statistical independence between distributed and often spatially remote neuronal units (Friston, 1994; Craddock et al., 2013). Despite the indirect nature of the blood oxygenation level dependent (BOLD) signal, functional magnetic resonance imaging (fMRI) has proven to be able to extract patterns of co-activation between clusters of voxels (Lowe et al., 2000).

The easiest way to operationalize the notion of functional connectivity is to calculate a covariance matrix which, in case of standardized variables, is equivalent to the correlation structure between brain regions. However, this approach is not able to identify direct (monosynap-

tic) functional connections as it is also sensitive to indirect (polysynaptic) functional interactions. For example, if regions A and B as well as regions B and C display correlated activity, then A and C will also show correlated activity even if they are not directly connected (Smith, 2012; Varoquaux and Craddock, In press).

In contrast, the precision matrix, defined as the inverse of the covariance matrix, captures conditional independence between brain regions (Lauritzen, 1996; Whittaker, 2009). That is, elements of the precision matrix are related to partial rather than full correlations and zero elements of the precision matrix imply an absence of direct functional connectivity. Therefore, sparse precision matrices provide us with valuable information about how different regions interact, though the estimates need to be interpreted with care (Marrelec and Benali, 2009; Friston, 2011; Woolrich and Stephan, In press; Hutchison et al., In press).

A common approach to obtain a point estimate for a sparse precision matrix is by means of the graphical lasso (Friedman et al., 2008; Varoquaux et al., 2010; Smith et al., 2011), which achieves sparseness through ℓ_1 regularization. Although the graphical lasso provides a reasonable point estimate, it is biased due to the induced shrinkage of the partial correlations. Furthermore, it does not directly provide a measure of uncertainty regarding the partial correlation estimates. This could lead to possibly erroneous conclusions about functional connectivity.

From a Bayesian perspective we are interested in the posterior density of the precision matrix given observed

*Corresponding author at Radboud University Nijmegen, Faculty of Science, Institute for Computing and Information Sciences, P. O. Box 9010, 6500 GL Nijmegen, The Netherlands. Tel.: +31 24 365 21 72, Fax: +31 24 365 27 28

Email address: mhinne@cs.ru.nl (M. Hinne)

data. Ultimately, this should lead to more reliable inferences about a subject’s cognitive state. In order to facilitate the estimation problem, we will not resort to shrinkage, as in the graphical lasso. Rather, we assume that the conditional independence structure between brain regions is given by an independent estimate of structural connectivity.

Structural connectivity refers to the presence of white matter tracts between spatially segregated brain regions (Hagmann et al., 2008). In humans, these tracts can be estimated in vivo by diffusion weighted imaging (DWI) which measures the anisotropy in the diffusion of water molecules (Le Bihan et al., 2001). The final result is usually a binary undirected graph which reports whether or not two areas are structurally connected. Clearly two brain regions can be directly functionally coupled only if they are physically connected, therefore the concepts of functional and structural connectivity are intimately related (Damoiseaux and Greicius, 2009). The idea is to infer structural connectivity from DWI data and use it as an additional constraint in our Bayesian model. The validity of this approach is supported by several recent experimental studies which found a substantial overlap between structural and functional networks both inside specific cortical areas (Koch et al., 2002) and on a whole brain scale (Hagmann et al., 2008; Honey et al., 2009, 2007; Damoiseaux and Greicius, 2009; Greicius et al., 2009; Cabral et al., 2012). Related approaches have been used before in the context of functional and effective connectivity analysis (Stephan et al., 2009; Deligianni et al., 2011; Ng et al., 2012).

In the following we present a new Bayesian framework for estimating functional connectivity. The framework, which we refer to as Bayesian functional connectivity (BFC) analysis, makes use of a G -Wishart prior (Roverato, 2002). This prior allows the sparseness structure of estimated precision matrices to be determined by a graph G , corresponding to structural connectivity. BFC analysis then amounts to computing a posterior density over sparse precision matrices. This posterior may then be used to compute marginal densities for partial correlations of interest. Our approach is compared with existing approaches using both simulated data and empirical data. We show that our approach provides robust partial correlation estimates while at the same time quantifying the uncertainty about functional connectivity.

2. Materials and Methods

2.1. Conventional functional connectivity estimation

Traditionally, functional connectivity estimation has relied on estimating covariance structure between p brain regions from timeseries data $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_N)$. Each vector $\mathbf{x}_n = (x_{n1}, \dots, x_{np})$ reflects neuronal activity (e.g. BOLD responses) for p brain regions. Without loss of generality, we assume that data is standardized to have zero mean

and unit standard deviation such that covariance coincides with correlation. It is assumed that the data are generated according to a zero-mean multivariate Gaussian density

$$p(\mathbf{X} | \mathbf{\Omega}) = \prod_n \mathcal{N}(\mathbf{x}_n | \mathbf{0}, \mathbf{\Omega}) \propto |\mathbf{\Omega}|^{N/2} \exp \left[-\frac{1}{2} \langle \mathbf{S} \mathbf{\Omega} \rangle \right] \quad (1)$$

with precision (inverse covariance) matrix $\mathbf{\Omega} = \mathbf{\Sigma}^{-1}$, scatter matrix $\mathbf{S} = \mathbf{X} \mathbf{X}^T$ and trace operator $\langle \cdot \rangle$. The choice of this distribution is justified by the fact that it is the maximum entropy distribution among all distributions with a specified mean and covariance (Cover and Thomas, 2006). Alternatively, the likelihood may be characterized in terms of the scatter matrix \mathbf{S} which follows a Wishart distribution, i.e.

$$p(\mathbf{S} | \mathbf{\Sigma}, N) = \mathcal{W}_p(\mathbf{S}, N) \propto |\mathbf{S}|^{-N/2} \exp \left[-\frac{1}{2} \langle \mathbf{S} \mathbf{\Sigma}^{-1} \rangle \right]. \quad (2)$$

This perspective can be applied more easily for distributions with a mean different from zero (Anderson, 1984).

We focus on estimating the precision matrix $\mathbf{\Omega} = \mathbf{\Sigma}^{-1}$ rather than the covariance matrix. As mentioned before, zero elements in $\mathbf{\Omega}$ reflect the absence of direct interactions. More formally, the sparseness structure of $\mathbf{\Omega}$, represented in terms of an undirected graph G where $V(G)$ is a set of nodes and $E(G)$ is a set of undirected edges between nodes, is equivalent to the conditional independence structure of a Gaussian Markov random field (Lauritzen, 1996; Whittaker, 2009). In other words, in the context of connectivity analysis, $\omega_{ij} = 0$ corresponds to the absence of structural connectivity between brain regions i and j .

In order to estimate the precision matrix $\mathbf{\Omega}$ of a zero-mean multivariate Gaussian density from data \mathbf{X} one may maximize the log likelihood

$$\log p(\mathbf{X} | \mathbf{\Omega}) = \frac{1}{2} [N \log |\mathbf{\Omega}| - \langle \mathbf{S} \mathbf{\Omega} \rangle]$$

which gives the maximum likelihood estimate (MLE):

$$\hat{\mathbf{\Omega}} = \arg \max_{\mathbf{\Omega} \in M^+} [N \log |\mathbf{\Omega}| - \langle \mathbf{S} \mathbf{\Omega} \rangle] = N \mathbf{S}^{-1} \quad (3)$$

where the maximization is constrained to precision matrices in the family of $p \times p$ positive definite matrices M^+ .

In practice, however, this empirical estimate does not contain zero elements. Furthermore, in case $N < p$, the maximum likelihood solution does not exist since \mathbf{S}/N becomes singular. Even in case $N > p$, the MLE is often poorly behaved, and regularization is called for (Pourahmadi, 2011). The graphical lasso (Friedman et al., 2008) regularizes the preceding MLE through sparsification by solving

$$\hat{\mathbf{\Omega}} = \arg \max_{\mathbf{\Omega} \in M^+} \left[\log |\mathbf{\Omega}| - \frac{1}{N} \langle \mathbf{S} \mathbf{\Omega} \rangle - \lambda \|\mathbf{\Omega}\|_1 \right]. \quad (4)$$

The employed ℓ_1 regularizer encourages sparse precision matrices as determined by the regularization parameter

λ. This maximization problem can be solved using established coordinate descent methods (Friedman et al., 2008). The graphical lasso has been proposed as the method of choice for functional connectivity estimation (Varoquaux et al., 2010; Smith et al., 2011; Varoquaux and Craddock, In press).

Even though the graphical lasso is commonly used to estimate sparse precision matrices, it suffers from two issues. First, since the graphical lasso employs shrinkage, pushing precision values towards zero, the resulting functional connectivity estimate is biased. Second, the graphical lasso produces a point estimate which does not directly allow inferences to be drawn about the uncertainty in our estimates arising from sampling noise and finite sample size.

2.2. Bayesian functional connectivity estimation

In order to tackle the aforementioned issues, we developed a Bayesian framework for inferring functional connectivity which does not rely on shrinkage but rather assumes that the sparseness structure G of Ω is given (Dempster, 1972). Specifically, we assume that the graph G is given by the structural connectivity as estimated from DWI data.

We start by assuming a G -Wishart distribution as the conjugate prior on precision matrices Ω . The G -Wishart is defined for the cone $M^+(G)$ of positive-definite symmetric matrices with off-diagonal elements $\omega_{ij} = 0$ whenever $(i, j) \notin E(G)$. A zero-constrained random matrix Ω has the G -Wishart distribution $\mathcal{W}_G(\delta_0, \mathbf{D})$ if its density is (Wang and Li, 2012):

$$p(\Omega | G) = \frac{|\Omega|^{(\delta_0-2)/2}}{Z_G(\delta_0, \mathbf{D})} \exp\left(-\frac{1}{2}\langle \mathbf{D}\Omega \rangle\right) 1_{\{\Omega \in M^+(G)\}}$$

where δ are the prior degrees of freedom, \mathbf{D} a symmetric positive definite prior scatter matrix, and $Z_G(\delta, \mathbf{D})$ the normalizing constant. The indicator function 1_x evaluates to 1 if its argument x is true and to 0 if its argument is false. In our experiments, we set $\delta_0 = 3$ and choose $\mathbf{D} = \mathbf{I}_{p \times p}$ (Moghaddam et al., 2009). This amounts to a vague prior for the precision matrix in (1), except that its support is restricted by G . We may now use Bayes' rule to obtain the posterior density for Ω according to

$$\begin{aligned} p(\Omega | \mathbf{X}, G) &\propto p(\mathbf{X} | \Omega)p(\Omega | G) \\ &= \frac{|\Omega|^{(\delta_n-2)/2}}{Z_G(\delta_n, \mathbf{B})} \exp\left(-\frac{1}{2}\langle \mathbf{B}\Omega \rangle\right) 1_{\{\Omega \in M^+(G)\}} \cdot \end{aligned} \quad (5)$$

Because the G -Wishart prior is conjugate to the likelihood, the resulting posterior once again follows a G -Wishart distribution with $\delta_n = \delta_0 + N$ the posterior degrees of freedom and $\mathbf{B} = \mathbf{D} + \mathbf{S}$ the posterior scatter matrix.

In order to approximate this posterior density we employed an efficient edgewise block Gibbs sampler described in detail in Wang and Li (2012). The algorithm is similar to a Bayesian iterative proportional scaling algorithm,

but instead of updating Ω in large blocks (e.g. maximum cliques in G , as is used in (Lenkoski and Dobra, 2011)), it is updated per edge. This can be done efficiently, as the authors show that this only requires 2×2 matrix inversions. Source code implementing the Gibbs sampler can be freely obtained from the author's website.¹

In practice, we find it more convenient to express functional connectivity in terms of partial correlation rather than precision. This can be easily achieved via the following transformation

$$\rho_{ij|\mathbf{Z}} = \begin{cases} -\frac{\omega_{ij}}{\sqrt{\omega_{ii}\omega_{jj}}} & \text{for } i \neq j \\ 1 & \text{for } i = j \end{cases} \quad (6)$$

reflecting the correlation between brain regions i and j when we condition on all other brain regions $\mathbf{Z} = \{1, \dots, N\} \setminus \{i, j\}$. Observe that, given this transformation, the interpretation in terms of conditional independence still holds. Hence, in practice, we use the posterior density

$$p(\mathbf{R} | \mathbf{X}, G) \quad (7)$$

for the partial correlation matrix \mathbf{R} as our estimate of Bayesian functional connectivity. This density is computed by applying the transformation (6) to each of the samples of the precision matrix as generated by the Gibbs sampler.

2.3. Experimental validation

We compared conventional functional connectivity analysis with Bayesian functional connectivity analysis using both simulated data and empirical data.

For the conventional analysis, we examined the maximum likelihood estimate (3) and the graphical lasso (4). In order to obtain an optimal estimate for the graphical lasso, we used a five-fold cross-validation procedure where the log likelihood computed on hold-out data was used to fine-tune the regularization parameter, as proposed in (Friedman et al., 2008).

For the BFC analysis, we computed the posterior density (7) using Gibbs sampling, where we discarded the first 2000 burn-in samples and stored the subsequent 5000 samples while using the transformation (6). The mode of this posterior (maximum a posteriori (MAP) estimate) was used to compare with point estimates obtained using conventional analyses. Both this MAP estimate as well as the graphical lasso estimates were computed using fast optimization procedures due to Schmidt (2010).²

In the remainder of this section, we describe the empirical data which were used to validate our approach.

2.4. Data acquisition

Six healthy volunteers were scanned after giving informed written consent in accordance with the guidelines

¹<http://www.stat.sc.edu/~wang345>

²<http://www.di.ens.fr/~mschmidt/Software>

Table 1: Left-hemisphere AAL regions used in functional connectivity analysis. Right-hemisphere AAL regions 46–90 follow the same labeling.

#	region	#	region	#	region	#	region	#	region
1	precentral	10	supp_motor	19	hippocampus	28	fusiform	37	putamen
2	frontal_sup	11	olfactory	20	parahippocampal	29	postcentral	38	pallidum
3	frontal_sup_orb	12	frontal_sup_medial	21	amygdala	30	parietal_sup	39	thalamus
4	frontal_mid	13	frontal_mid_orb	22	calcarine	31	parietal_inf	40	heschl
5	frontal_mid_orb	14	rectus	23	cuneus	32	supramarginal	41	temporal_sup
6	frontal_inf_oper	15	insula	24	lingual	33	angular	42	temporal_pole_sup
7	frontal_inf_tri	16	cingulum_ant	25	occipital_sup	34	precuneus	43	temporal_mid
8	frontal_inf_orb	17	cingulum_mid	26	occipital_mid	35	paracentral_lobule	44	temporal_pole_mid
9	rolandic_oper	18	cingulum_post	27	occipital_inf	36	caudate	45	temporal_inf

of the local ethics committee. These subjects represent a subset of the data previously used in (Hinne et al., 2013). T1 structural scans, resting-state functional data and diffusion-weighted images were obtained using a Siemens Magnetom Trio 3 T system at the Donders Centre for Cognitive Neuroimaging, Radboud University Nijmegen, The Netherlands. Resting-state fMRI data were acquired at 3 Tesla by using a multi-echo echo-planar imaging (ME-EPI) sequence (voxel size 3.5 mm isotropic, matrix size 64×64, TR=2000 ms, TEs=6.9, 16.2, 25, 35 and 45 ms, 39 slices, GRAPPA factor 3, 6/8 partial Fourier). A total of 1030 volumes were obtained. An optimized acquisition order described by Cook et al. (2007) was used in the DWI protocol (voxel size 2.0 mm isotropic, matrix size 110×110, TR=13,000 ms, TE=101 ms, 70 slices, 256 directions at b=1500 s/mm²).

Tools from the Oxford FMRIB Software Library (FSL, FMRIB, Oxford, UK) were used to preprocess the data. For each subject’s anatomical scan, brain extraction was performed using FSL BET (Smith, 2002), probabilistic brain tissue segmentation was performed using FSL FAST (Zhang et al., 2001) and sub-cortical structures were segmented using FSL FIRST (Patenaude et al., 2011).

2.5. Preprocessing of diffusion imaging data

DWI data were preprocessed using FSL FDT (Behrens et al., 2003) and consisted of correction for eddy currents and estimation of the diffusion parameters. A measure of white matter connectivity was obtained by FDT Protrackx 2.0 (Behrens et al., 2003, 2007). The voxels that live on the boundary between white matter and gray matter were used as seeds for tractography, gray matter was used as a target mask. For each seed voxel 5000 streamlines were drawn with a maximum length of 2000 steps. The streamlines were restricted by the fractional anisotropy to prevent them from wandering around in gray matter. Streamlines in which a sharp angle ($> 80^\circ$) occurred were discarded.

This procedure resulted in a matrix providing voxel-to-voxel connectivity for each subject s . Each matrix was reduced to a 90×90 streamline count matrix C^s by summing the streamline counts over voxels belonging to 90 distinct brain regions, as defined by the automatic anatomical labelling (AAL) atlas (Tzourio-Mazoyer et al., 2002) while

excluding cerebellar structures. The regions defined by the AAL template are displayed in Table 1 for ease of reference. Finally, edges of the structural connectivity graph G were determined by the streamline count matrices as follows:

$$(i, j) \in E(G) \leftrightarrow \bigcap_{s \in \{1, \dots, 6\}} (c_{ij}^s > 0 \wedge c_{ji}^s > 0). \quad (8)$$

This particular definition corresponds to the following intuition: at the coarse scale that is given by the AAL template, we require that each subject shows support for a particular connection. Per subject we are lenient and threshold at zero. Alternatively, using the union of the thresholded networks for each subject would result in extremely dense estimates. The structural graph G was used as a constraint for Bayesian functional connectivity analysis.

2.6. Preprocessing of functional data

Multi-echo images obtained using the rs-fMRI acquisition protocol were combined using a custom Matlab script (MATLAB 7.7, The MathWorks Inc., Natick, MA, USA) which implements the procedure described in (Poser et al., 2006) and also incorporates motion correction using functions from the SPM5 software package (Wellcome Department of Imaging Neuroscience, University College London, UK). Of the 1030 combined volumes, the first six were discarded to allow the system to reach a steady state. The data was then factorized into 77 independent components using FSL MELODIC (Beckmann and Smith, 2004), of which 37 ± 1 were retained by manually removing artifact components. After preprocessing, the fMRI data were parcellated according to the AAL atlas. For these regions, the functional data was summed and standardized to have zero mean and unit standard deviation. The resulting BOLD timeseries for all six subjects were used in subsequent functional connectivity analyses.

3. Results

3.1. Simulation study

We performed a simulation study in order to compare the performance of the G -Wishart approach with the maximum likelihood estimate and the graphical lasso. We assumed the structural estimate defined in (8) as the ground

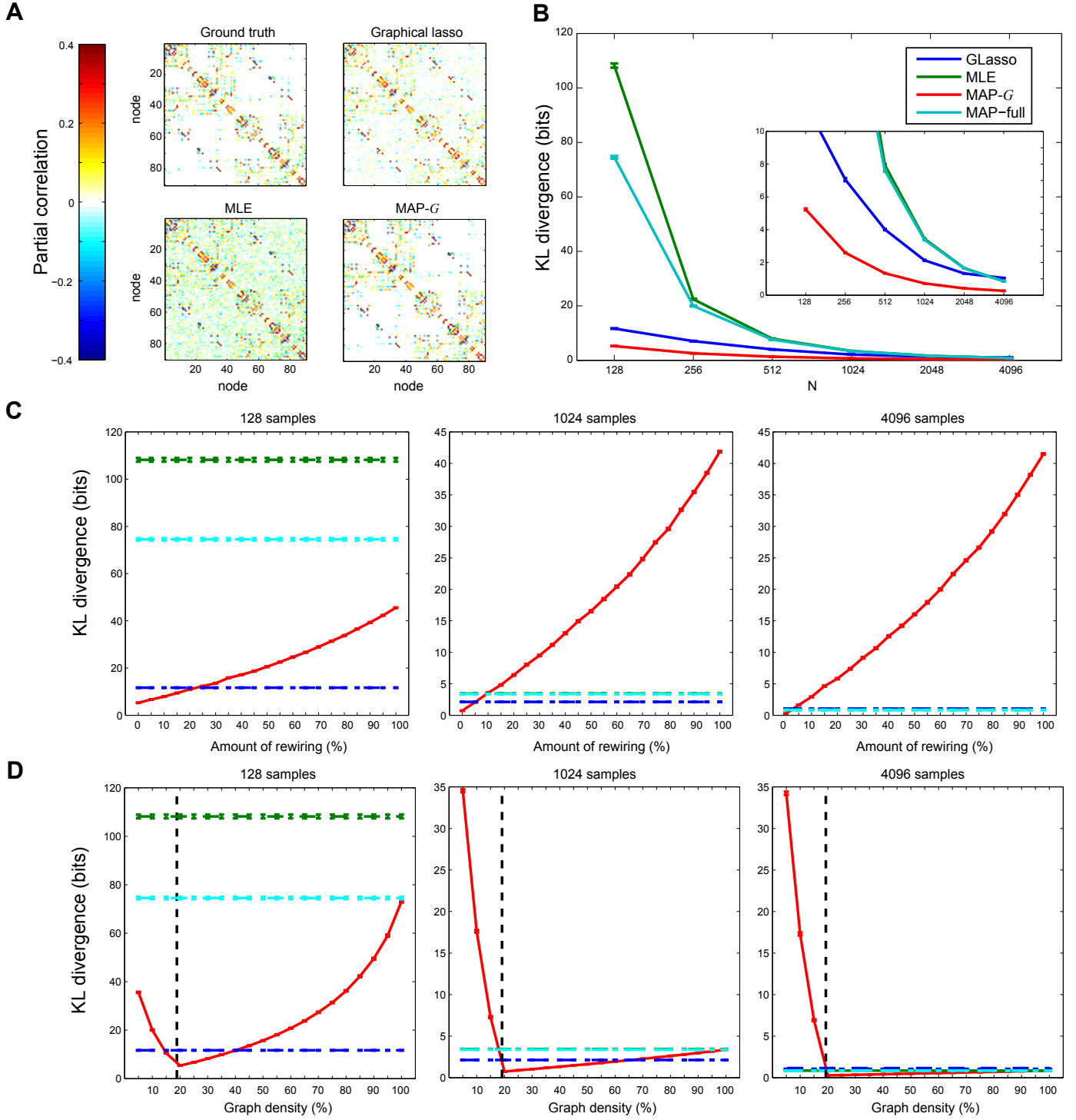


Figure 1: **A**. The simulated precision matrix Ω_s , and reconstructions Ω_r using the graphical lasso, MLE and MAP-G solution, for $N = 1024$. Note that the diagonal elements have been set to zero to increase the visibility of the differences between the off-diagonal elements. The reconstruction for the MAP estimate using a full graph has been omitted, as it was indistinguishable from the MLE solution. **B**. The Kullback-Leibler divergence for the four different methods as a function of the number of samples. The inset shows a detailed view. **C**. The Kullback-Leibler divergence for the four different methods as a function of the amount of rewiring applied to the ground truth graph. Note that the MLE, the graphical lasso and the MAP estimate using the full graph are unaffected by this rewiring; the dotted lines are provided as a visual aid. **D**. The Kullback-Leibler divergence for the four different methods as a function of false positives and false negatives. The vertical dashed line indicates the actual density. The effect of false negatives is shown on the left side of this line (until the graph is near empty) whereas the right side shows the effect of false positives (until the graph is the complete graph). Similar to **C**, the other methods are unaffected by the changes in G and are only shown for convenience.

truth. Using this structure, 100 random precision matrices $\mathbf{\Omega}_s$ were drawn from the posterior $p(\mathbf{\Omega}|\mathbf{X}, G)$ with $\delta_n = 3 + n$ and $\mathbf{B} = \mathbf{I}_{p \times p} + \mathbf{S}$, with \mathbf{S} the mean covariance matrix of the six subjects. The consequence of this approach was that the used precision matrices were ensured to be positive definite, follow the structure of G , and have comparable conditioning as the MAP estimates in our experiments using real data. From the sampled precision matrices $\mathbf{\Omega}_s$, an $N \times p$ data matrix $\mathbf{X}_s \sim \mathcal{N}(\mathbf{0}, \mathbf{\Omega}_s^{-1})$ was generated, with $N = \{128, 256, \dots, 4096\}$. For each ground truth precision matrix $\mathbf{\Omega}_s$ and the different sample sizes, we created reconstructions $\mathbf{\Omega}_r$ of $\mathbf{\Omega}_s$ using four strategies: the graphical lasso estimate, the maximum likelihood estimate (MLE), the maximum a posteriori (MAP) estimate using the G -Wishart prior and G_s as structural estimate and the MAP estimate using the Wishart prior, which corresponds to a G -Wishart prior with a full graph. An example precision matrix, as well as the different reconstructions, is shown in Fig. 1A.

We quantified the quality of the reconstruction of the ground truth precision using the Kullback-Leibler (KL) divergence:

$$D_{\text{KL}}(\mathbf{\Omega}_s || \mathbf{\Omega}_r) = \frac{1}{2 \log 2} \left[\log \frac{|\mathbf{\Omega}_s|}{|\mathbf{\Omega}_r|} + \langle \mathbf{\Omega}_r, \mathbf{\Omega}_s^{-1} \rangle - p \right].$$

The results of the simulation are provided in Fig. 1B. The figure shows that compared to the maximum likelihood estimator and the MAP estimate using the Wishart distribution, the G -Wishart approach does not require many samples in order to reconstruct $\mathbf{\Omega}_s$. Also, it outperforms the graphical lasso estimates. However, this critically depends on the quality of G . To see how well the G -Wishart MAP estimate approximates the ground truth, we ran additional simulation runs in which the structural graph G was increasingly perturbed. We considered three cases. In the first, the graph is rewired by removing edges at random and simultaneously connecting randomly selected disconnected nodes, thus keeping the density of G constant. The results of this rewiring are shown in Fig. 1C. In the second and third case we either removed edges, representing false negatives, or we added edges, representing false positives. The results of these perturbations are shown in Fig. 1D. The results show that the performance of the G -Wishart approach deteriorates linearly with increasing rewiring. In the low-powered setting of $N = 128$, we observe that the approach outperforms the alternatives for up to 20% of the edges replaced, but for the high-powered setting of $N = 4096$ this percentage drops to 5%. For false negative errors in the structural estimate, the number of samples has a clear effect on the performance. For few samples, the effect of mere sparsity keeps the Kullback-Leibler divergence lower than the MLE or the Wishart MAP, although the graphical lasso obviously performs better than the (near) empty structural graph. Once more samples are acquired, false negatives have an even stronger negative influence on the performance. Less troublesome are false positives, for which we observe the converse behavior.

For few samples, false positives provide difficulty for the G -Wishart approach, but once a large number of samples is obtained, an increase in density of the structural estimate of up to 90% still shows the G -Wishart approach as the most successful.

Finally, we considered the merit of the Bayesian approach by considering how the posterior density of partial correlation changes depending on the number of samples and the quality of the provided structural estimate. We considered three cases: the estimation of partial correlation for a ground truth edge using G , the same estimation using a perturbed graph with 20% of the edges replaced, but with the ground truth edge still intact and finally the estimation of partial correlation for an edge that was absent in the ground truth, yet is present in the perturbed graph. We used the same structural estimate for G as before, and used one random sampled precision matrix as ground truth, from which data was generated for $N = \{128, 1024, 4096\}$. As an example, Fig. 2A and B show the estimated partial correlation between the supplementary motor areas for the true graph and the perturbed graph that contains this connection, respectively. Notably, the distributions become more narrow when more samples are acquired (as expected). In Fig. 2C the partial correlation between the left cuneus and the right temporal pole is shown. Note that this connection is not in our structural estimate. The distributions show that for erroneous edges (i.e. false positives), the G -Wishart approach correctly estimates a partial correlation of zero.

3.2. Empirical validation

We now turn to the empirical validation where we used a structural connectivity graph G as estimated from DWI data to constrain the correlation structure which explains resting-state fMRI data.

3.2.1. Estimated structural connectivity graph

The structural connectivity as estimated via the procedure described in Eq. (8) is shown in Fig. 3, separately for intra-hemispheric connectivity in left and right hemispheres as well as inter-hemispheric connectivity between left and right hemispheres. The region numbers correspond to those displayed in Table 1. A total of 774 structural links were identified of which 329 were left intra-hemispheric connections, 328 were right intra-hemispheric connections and 117 were inter-hemispheric connections. This gives a network density of 0.19. Note that this is denser than the density of 0.11 reported by Gong et al. (2009). However, for our purposes, the BFC analysis can still provide evidence against edges which were erroneously included in the structural graph as we saw in the simulation.

3.2.2. Bayesian functional connectivity analysis

BFC analysis was performed on the resting-state data for each of the six subjects and amounted to computing

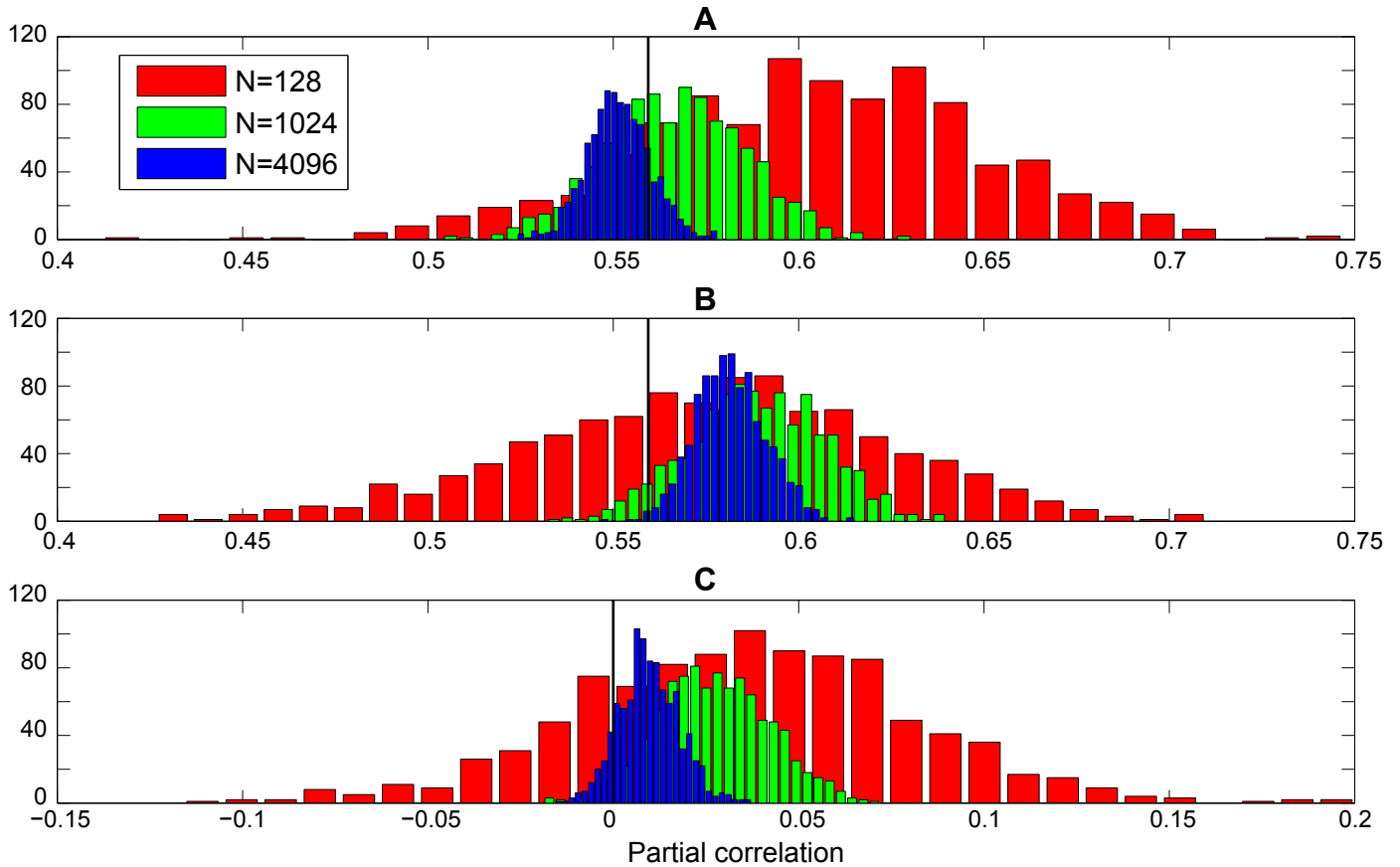


Figure 2: **A.** Partial correlation for an example true edge. **B.** Partial correlation using a perturbed graph that still contained this edge. **C.** Partial correlation for an example edge that was absent in the ground truth. The true partial correlation is indicated with a vertical black line.

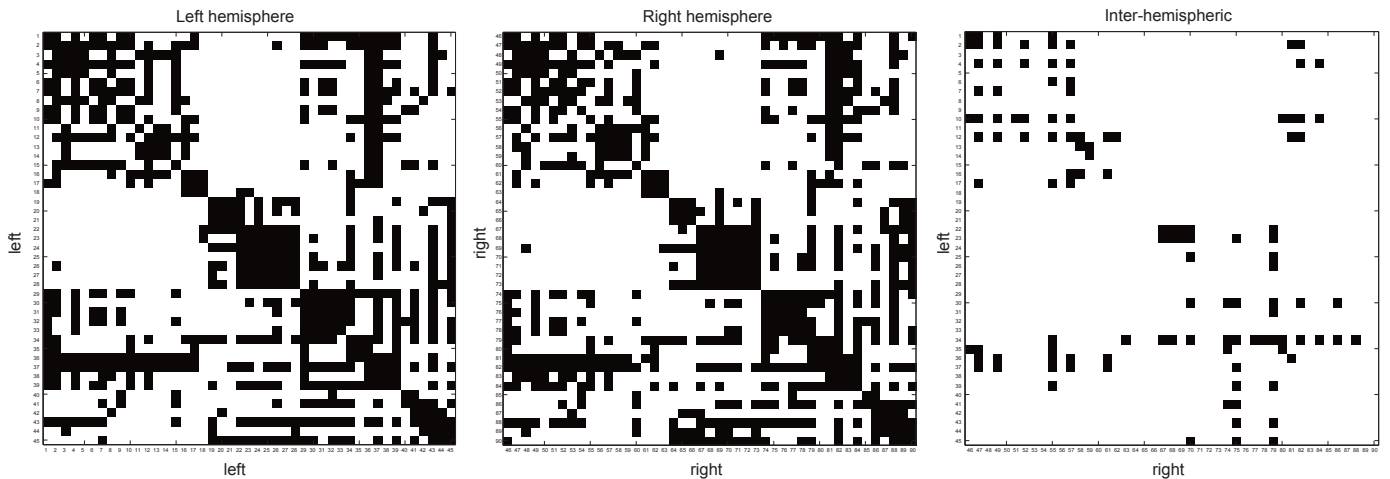


Figure 3: Connectivity of the structural connectivity matrix used as a constraint in BFC estimation. Submatrices representing intra-hemispheric and inter-hemispheric connectivity are shown separately here for ease of reference. Black (white) squares indicate the presence (absence) of a structural connection.

posterior densities $p(\mathbf{R} \mid \mathbf{X}, G)$. Figure 4 shows the mean partial correlations between all brain regions for each of the six subjects. Mean partial correlations between regions ranged from -0.43 to 0.79 . Furthermore, partial correlation estimates had standard deviations ranging up to

0.04 , illustrating the fact that absolute certainty cannot be achieved using a finite amount of data. It can be observed that functionally homologous regions in the left and right hemispheres are strongly partially correlated.

In order to gain more insight into the partial correlations

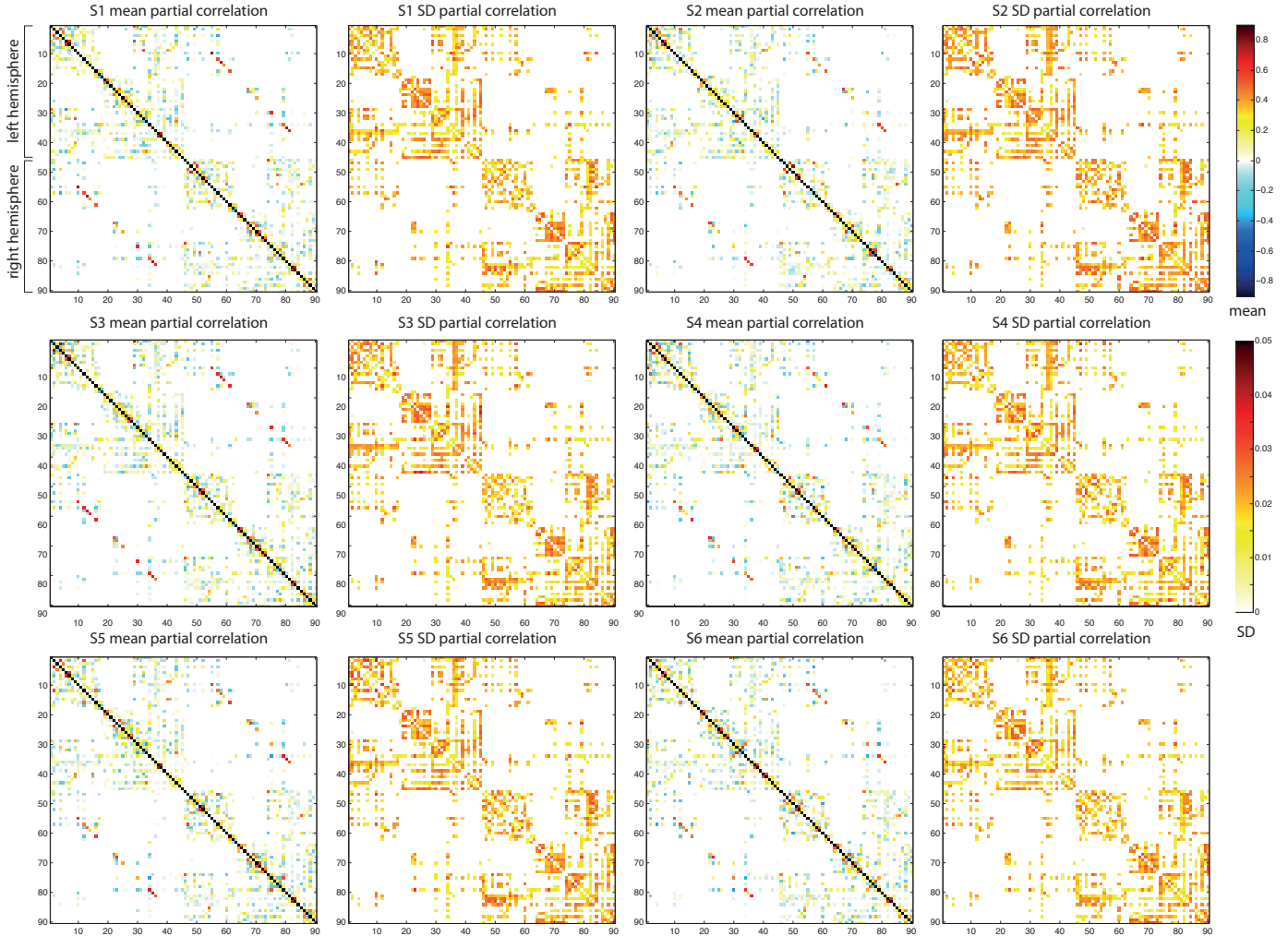


Figure 4: Whole-brain BFC estimates. Mean and standard deviation of the partial correlation matrices are shown for each of the six subjects.

between particular regions, we focus on connections with either high or low mean partial correlations. Since qualitatively different behavior can be observed for inter- and intra-hemispheric connections, we handle them separately.

Table 2 depicts five inter-hemispheric connections and five intra-hemispheric connections which showed the strongest partial correlations. Evidence for non-zero partial correlations between the regions identified in Table 2 is supported by literature. Specifically, partial correlations between supplementary motor areas (Salvador et al., 2005), cingulate cortices (Salvador et al., 2005), as well as prefrontal areas (Damoiseaux et al., 2006; Beckmann et al., 2005), have been identified before. The strong partial correlation between putamen and pallidum can be explained by monosynaptic feedforward and feedback pathways between the striatum and the globus pallidus (Bolam et al., 2000; Smith and Bolam, 1990). Structural connectivity between left and right precuneus is also supported by anatomical studies (Cavanna and Trimble, 2006). Likewise, evidence exists for direct connections between left and right caudate (Mensah and Deadwyler, 1974; Medina

and Pazo, 1981).

Table 3 shows five inter-hemispheric connections and five intra-hemispheric connection which showed the weakest partial correlations. A comparison between Tables 2 and 3 suggests that connections with strong partial correlations are supported by many more streamlines compared to the connections that showed weak partial correlations. This is confirmed by highly significant positive correlations between streamline count and absolute partial correlation values ($\rho > 0.40$, $p < 10^{-31}$ in all subjects). This relationship has been observed before (Hermundstad et al., 2013). This indicates that there is a clear correspondence between the information conveyed by the structural connectivity as estimated using DWI and the Bayesian functional connectivity analysis applied to resting-state fMRI data.

3.2.3. Comparison with the graphical lasso

While the Bayesian approach to functional connectivity analysis has been shown to lead to interpretable estimates of partial correlation while at the same time quantifying the uncertainty in these estimates, it is important to com-

Table 2: Five inter-hemispheric and intra-hemispheric connections with highest mean partial correlations. Hemispheres H are indicated by ‘L’ (left hemisphere) and ‘R’ (right hemisphere). Mean and standard deviation of the partial correlation, averaged over subjects, are denoted by μ and σ , respectively. Logarithm of the total number of streamlines between regions, averaged over subjects, is denoted by n .

#	region 1	H	region 2	H	μ	σ	$\log n$
1	supp_motor_area	R	supp_motor_area	L	0.60	0.018	9
2	paracentral_lobule	R	paracentral_lobule	L	0.59	0.018	7
3	cingulum_ant	R	cingulum_ant	L	0.57	0.018	7
4	precuneus	R	precuneus	L	0.57	0.016	7
5	caudate	R	caudate	L	0.56	0.016	4
1	frontal_mid	L	frontal_sup	L	0.65	0.019	10
2	pallidum	R	putamen	R	0.60	0.017	10
3	pallidum	L	putamen	L	0.59	0.016	11
4	frontal_inf_tri	L	frontal_inf_oper	L	0.56	0.019	9
5	frontal_inf_tri	R	frontal_inf_oper	R	0.56	0.019	9

Table 3: Five inter-hemispheric and intra-hemispheric connections with lowest mean partial correlations. Hemispheres H are indicated by ‘L’ (left hemisphere) and ‘R’ (right hemisphere). Mean and standard deviation of the partial correlation, averaged over subjects, are denoted by μ and σ , respectively. Logarithm of the total number of streamlines between regions, averaged over subjects, is denoted by n .

#	region 1	H	region 2	H	μ	σ	$\log n$
1	frontal_sup_medial	R	putamen	L	0.02	0.016	5
2	putamen	R	frontal_mid	L	0.02	0.014	4
3	parietal_sup	R	putamen	L	0.02	0.014	2
4	putamen	R	parietal_sup	L	0.02	0.014	2
5	putamen	R	precuneus	L	0.02	0.015	4
1	putamen	L	calcarine	L	0.01	0.009	5
2	caudate	L	precuneus	L	0.02	0.013	4
3	putamen	R	lingual	R	0.02	0.017	5
4	putamen	L	cingulum_ant	L	0.02	0.019	5
5	lingual	R	frontal_sup_orb	R	0.02	0.007	4

pare its behavior with conventional approaches to functional connectivity analysis. In the following, we compare BFC estimates with estimates produced by the graphical lasso. Note that the cross-validation procedure used to select the value of the regularization parameter for the graphical lasso led to quite dense partial correlation matrices. On average, only 53 region pairs were estimated to have zero partial correlation.

Figure 5A shows the whole-brain point estimates produced by the graphical lasso. A comparison of these estimates with those obtained by BFC analysis, as depicted in Fig. 4, indicate that both approaches show some correspondence in terms of block diagonal structure and strong partial correlations between inter-hemispheric functionally homologous areas. At the same time, results show that the estimated partial correlations tend to be weaker for the graphical lasso.

Figure 5B shows a scatter plot comparing partial correlation estimates obtained using both approaches. The clouds of blue and yellow datapoints indicate that the graphical lasso leads to smaller partial correlation estimates as expected by the shrinkage property. This is confirmed by the fact that 72% of the connections according to G have lower absolute partial correlations for the graphical lasso compared to the Bayesian approach. Green datapoints show non-zero partial correlations for BFC analysis

which were forced to zero by the graphical lasso. Red datapoints signify non-zero partial correlations for the graphical lasso which were forced to be zero according to the structural graph G . The two outliers with non-zero partial correlations above 0.7 according to the graphical lasso reflect connectivity between left and right posterior cingulate cortex in two subjects, which has been reported previously (Salvador et al., 2005).

An alternative to using cross-validation for selecting the regularization parameter λ is to tune it such that a predetermined network density is achieved. Figure 5C compares partial correlations between BFC analysis and graphical lasso estimates obtained with λ set to 0.15. This gives much sparser point estimates that more closely resemble the structural graph G . Two main effects can be observed. First, with increased λ partial correlations are much more affected by shrinkage. That is, 94% of the connections given by the structural graph G have lower absolute partial correlations for the graphical lasso compared to the Bayesian approach. Second, there is a substantial increase in the number of connections implied by the structural graph that are set to zero.

In order to quantify the fit of the different models to the data we used a hold-out scheme where the first 512 samples were used as training data, and the remaining samples as test data. If we compare the G -Wishart approach, the

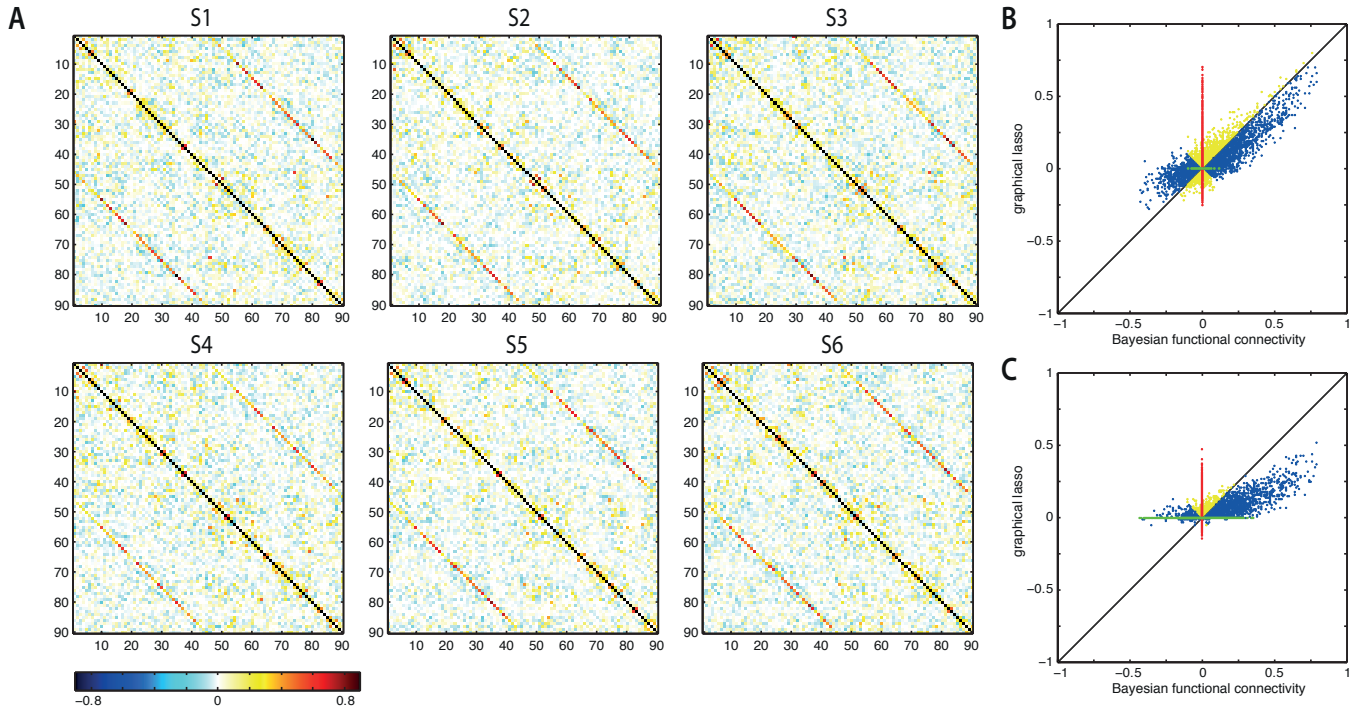


Figure 5: Graphical lasso results. (A) Graphical lasso functional connectivity estimates converted to partial correlations for six subjects. (B) Scatter plot comparing partial correlations for all six subjects as estimated using BFC analysis (mean partial correlation) or the graphical lasso. For the graphical lasso, a point estimate was obtained using nested cross-validation. Blue datapoints denote non-zero partial correlations that are larger in magnitude for BFC than for the graphical lasso. Yellow datapoints denote non-zero partial correlations that are lower in magnitude for BFC than for the graphical lasso. Red datapoints denote partial correlations which were zero for BFC analysis and non-zero for the graphical lasso. Green datapoints denote partial correlations which were zero for the graphical lasso and non-zero for BFC analysis. (C) Same as (B) but now with graphical lasso estimates obtained with λ fixed to 0.15.

cross-validated graphical lasso and the graphical lasso with sparsity matched with G , we find a log-likelihood on the test data of $-4.81 \cdot 10^4 \pm 0.48 \cdot 10^4$, $-5.90 \cdot 10^4 \pm 0.11 \cdot 10^4$ and $-4.36 \cdot 10^4 \pm 0.18 \cdot 10^4$, respectively. From these results we conclude that the sparsity of G is valuable for correct estimation of partial correlations, but that the structural estimate we used was sub-optimal.

3.2.4. Posterior densities

In order to gain additional insight into the posterior densities estimated by BFC analysis and to allow further comparison with the point estimates produced by the graphical lasso, Fig. 6 shows the estimates in each subject for the inter- and intra-hemispheric connections with strongest and weakest partial correlation according to the BFC analysis. For the strong partial correlations, shown in Fig. 6A, the graphical lasso estimates tend to be much smaller than the estimates obtained using the BFC approach. In part, this is likely due to the shrinkage induced by the graphical lasso. MLE estimates are also shown for comparison. The weak partial correlations shown in Fig. 6B are especially interesting. While the graphical lasso makes binary statements about the presence or absence of a non-zero partial correlation, Bayesian functional connectivity analysis quantifies the uncertainty in our estimates, thereby providing a more nuanced view.

For comparison, Fig. 6 also shows the 95% confidence intervals around the MAP estimates of the distribution, which were obtained by applying the Fisher transform to the estimates (Fisher, 1915) with $n-3-p = 1024-3-90 = 931$ degrees of freedom. The distributions and intervals show that for high partial correlations, the 95% confidence interval is enclosed in the MAP distribution, while for low partial correlation the confidence interval is wider than the (bulk of the) distribution.

4. Discussion

We proposed Bayesian functional connectivity analysis as a new approach for analyzing the interactions between BOLD timeseries in multiple brain regions. The approach produces a posterior density for the partial correlation matrix and relies on the use of a G -Wishart distribution. This distribution has been used extensively for analyzing covariance structure in high-dimensional problems in biology (Jones et al., 2005), economics (Carvalho et al., 2007) and epidemiology (Dobra et al., 2011). In a neuroscientific context, we use the G -Wishart distribution as a prior which constrains the partial correlation matrix via a structural graph G , as derived from diffusion imaging data.

As demonstrated using both simulated and empirical data, Bayesian functional connectivity analysis based on

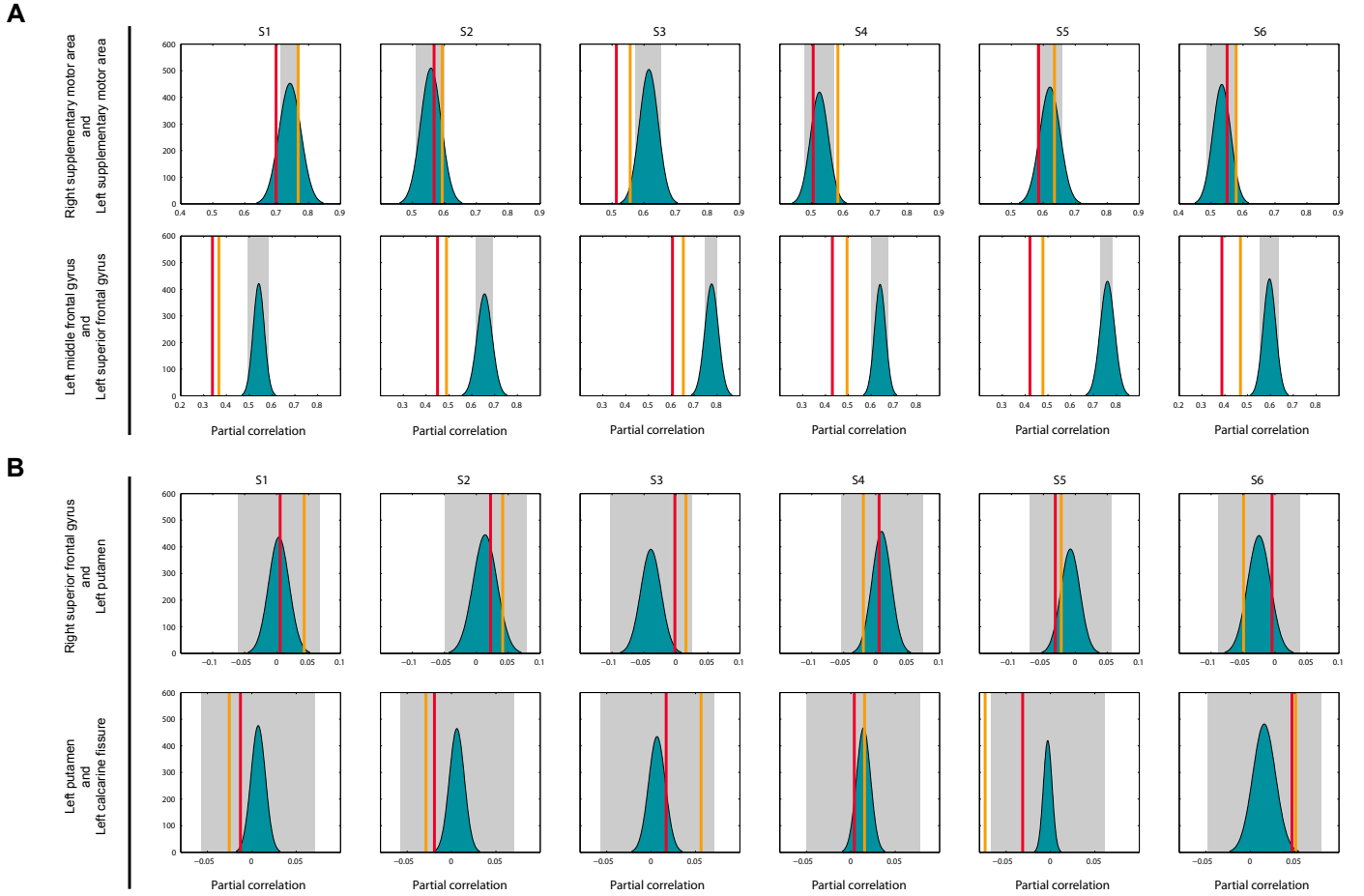


Figure 6: Posterior densities for individual partial correlations as estimated by BFC analysis (blue), point estimates produced by the graphical lasso (red) and the MLE (yellow). The shaded areas show the 95% confidence interval around the MAP estimate. (A) Estimates obtained for the inter- and intra- hemispheric connections with the strongest partial correlation. (B) Estimates obtained for the inter- and intrahemispheric connections with the weakest partial correlation.

the G -Wishart is a promising approach for functional connectivity analysis in cognitive neuroscience. The main advantages of our approach compared to existing approaches are as follows. First, we constrain the functional connectivity estimates using a structural graph G as estimated from diffusion imaging data, thereby effectively achieving multi-modal data fusion (Biessmann et al., 2011; Groves et al., 2011). The constraints imposed by G alleviate the need to impose shrinkage, thereby reducing bias in the partial correlation estimates. Second, we can quantify the uncertainty in our estimates, which allows sound inferences to be drawn about the presence of non-zero partial correlations between BOLD timecourses for multiple regions of interest.

Results based on simulated data show that the G -Wishart approach outperforms the maximum likelihood estimate, the normal Wishart MAP estimate and the graphical lasso for limited data (cf. Fig. 1B). As more data becomes available the different approaches converge to the same estimate. Note however that these results are obtained when the G -Wishart approach has access to the ground truth graph G . Figure 1C and Fig. 1D show that

as the structural graph is perturbed more and more, eventually, the G -Wishart is outperformed by the MLE and the graphical lasso.

Empirical results show that the connections with high partial correlations correspond to anatomical tracts that are known from literature. In our empirical comparison with the graphical lasso, we observed salient differences between respective functional connectivity estimates. This can be due to a number of reasons. First, we used cross-validation for the graphical lasso to select the regularization parameter λ . This led to much denser precision matrices and, hence, different functional connectivity estimates. Second, the graphical lasso employs shrinkage, which can explain in part the observation that graphical lasso partial correlation estimates are typically lower than those obtained with the Bayesian approach. As shown in Fig. 5C, an alternative to cross-validation is to set the regularization parameter to a fixed value, achieving a desired network density. However, in this case, many connections that are implied by the structural graph and achieve relatively high partial correlations under the Bayesian approach will be set to zero. Furthermore, the shrinkage

effect becomes stronger, thereby further underestimating partial correlation values. By comparing the model likelihood on hold-out data for either the G -Wishart approach, the cross-validated graphical lasso and the graphical lasso with a fixed density, we find that the graphical lasso algorithm with fixed density performs best, followed by the G -Wishart approach. Note however, that the fixed lasso density was determined by the structure G and that cross-validation, which would be a standard way to estimate the lasso shrinkage, performed worse. Most likely, improvements in the structural estimate will increase the G -Wishart performance.

In order to constrain the estimation problem, BFC analysis makes use of a structural graph as estimated from diffusion imaging data. Drawn conclusions therefore critically depend on the quality of the employed structural graph. As shown in Fig. 1C and Fig. 1D, errors in the estimated structural matrix can produce serious biases as connections between regions that show high partial correlation can be completely ruled out from the analysis. In other words, false negatives in the structural estimate strongly (negatively) influence the usefulness of the G -Wishart approach. For example, we observed that the potential connection between posterior cingulate cortices (which was found to have a high partial correlation in two subjects when using the graphical lasso) was absent in our estimates. In order to prevent the exclusion of important connections due to biases associated with probabilistic streamlining (Dauguet et al., 2007; Li et al., 2012), a more lenient threshold may be used to estimate the structural graph, preventing false negatives at the expense of false positives. In addition, connections known to be absent may simply be ruled out in the prior. Notwithstanding these caveats, the proposed methodology is theoretically sound and may take advantage of various future developments in diffusion imaging (Dell’Acqua and Catani, 2012).

A potential alternative to the multimodal approach followed in this paper would be to dispense with diffusion imaging data altogether and use functional data to estimate not only a posterior density over partial correlation matrices but also over structural graphs (Atay-Kayis and Massam, 2005). In order to solve this inference problem, sophisticated methods have been developed (Wang and Li, 2012). However, at present, these methods do not scale well with problem size, prohibiting a straightforward application to functional connectivity analysis.

Notwithstanding the strengths and weaknesses of different approaches to functional connectivity analysis, we maintain that, in order to draw valid conclusions about functional connectivity, one needs to employ methods that quantify the uncertainty in our estimates. This holds especially when inferences depend on a small amount of data. One approach would be to use bootstrap procedures together with the graphical lasso (Hastie et al., 2008). Another approach, as demonstrated in this paper, is to use a Bayesian approach. While both structural connectivity and effective connectivity have been tackled

using Bayesian approaches (Hinne et al., 2013; Jbabdi et al., 2007; Daunizeau et al., 2011), research into how whole brain functional connectivity can be estimated using Bayesian approaches has remained scarce. Some notable exceptions are the approach by Venkataraman et al. (2010) who used a forward model in which fMRI and DWI data were combined and the approach by (Marrelec et al., 2006), who used a Bayesian approach to estimate a group partial correlation matrix. As advocated in this paper, we propose that a generative model consisting of a G -Wishart prior and a multivariate Gaussian likelihood term serves as an elegant new approach for Bayesian functional connectivity analysis.

Acknowledgments

The authors gratefully acknowledge the support of the BrainGain Smart Mix Programme of the Netherlands Ministry of Economic Affairs and the Netherlands Ministry of Education, Culture and Science. The authors thank Erik van Oort and David Norris for the acquisition of the rs-fMRI and DWI data.

References

- Anderson, T., 1984. An Introduction to Multivariate Statistical Analysis. Wiley series in probability and mathematical statistics, Wiley.
- Atay-Kayis, A., Massam, H., 2005. A Monte Carlo method for computing the marginal likelihood in nondecomposable Gaussian graphical models. *Biometrika* 92, 317–335.
- Beckmann, C., DeLuca, M., Devlin, J., Smith, S., 2005. Investigations into resting-state connectivity using independent component analysis. *Phil. Trans. R. Soc. B* 360, 1001–1013.
- Beckmann, C.F., Smith, S.M., 2004. Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE Trans. Med. Imaging* 23, 137–152.
- Behrens, T.E.J., Johansen-Berg, H., Jbabdi, S., Rushworth, M.F.S., Woolrich, M.W., 2007. Probabilistic diffusion tractography with multiple fibre orientations: What can we gain? *NeuroImage* 34, 144–155.
- Behrens, T.E.J., Woolrich, M.W., Jenkinson, M., Johansen-Berg, H., Nunes, R.G., Clare, S., Matthews, P.M., Brady, J.M., Smith, S.M., 2003. Characterization and propagation of uncertainty in diffusion-weighted MR imaging. *Magnet. Reson. Med.* 50, 1077–1088.
- Biessmann, F., Plis, S., Meinecke, F.C., Eichele, T., Muller, K., 2011. Analysis of Multimodal Neuroimaging Data. *IEEE Rev. Biomed. Eng.* 4, 26–58.
- Bolam, J.P., Hanley, J.J., Booth, P.A., Bevan, M.D., 2000. Synaptic organisation of the basal ganglia. *J. Anat.* 196, 527–542.
- Bullmore, E., Sporns, O., 2009. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat. Rev. Neurosci.* 10, 186–198.
- Cabral, J., Hugues, E., Kringelbach, M.L., Deco, G., 2012. Modeling the outcome of structural disconnection on resting-state functional connectivity. *NeuroImage* 62, 1342–1353.
- Carvalho, C.M., Massam, H., West, M., 2007. Simulation of hyper-inverse Wishart distributions in graphical models. *Biometrika* 94, 647–659.
- Cavanna, A.E., Trimble, M.R., 2006. The precuneus: a review of its functional anatomy and behavioural correlates. *Brain* 129, 564–583.

- Cook, P.A., Symms, M., Boulby, P.A., Alexander, D.C., 2007. Optimal acquisition orders of diffusion-weighted mri measurements. *J. Magn. Reson. Imaging* 25, 1051–1058.
- Cover, T.M., Thomas, J.A., 2006. *Elements of Information Theory*. Wiley, New Jersey. 2nd edition.
- Craddock, R.C., Jbabdi, S., Yan, C.G., Vogelstein, J.T., Castellanos, F.X., Di Martino, A.e.a., 2013. Imaging human connectomics at the macroscale. *Nat. Meth.* 10, 524–539.
- Damoiseaux, J.S., Greicius, M.D., 2009. Greater than the sum of its parts: a review of studies combining structural connectivity and resting-state functional connectivity. *Brain Struct. Funct.* 213, 525–533.
- Damoiseaux, J.S., Rombouts, S.A.R.B., Barkhof, F., Scheltens, P., Stam, C.J., Smith, S.M., Beckmann, C.F., 2006. Consistent resting-state networks. *Proc. Natl. Acad. Sci. USA* .
- Dauguet, J., Peled, S., Berezovskii, V., Delzescaux, T., Warfield, S.K., Born, R., Westin, C.F., 2007. Comparison of fiber tracts derived from in-vivo DTI tractography with 3D histological neural tract tracer reconstruction on a macaque brain. *NeuroImage* 37, 530–538.
- Daumizeau, J., David, O., Stephan, K.E., 2011. Dynamic causal modelling: A critical review of the biophysical and statistical foundations. *NeuroImage* 58, 312–322.
- Deligianni, F., Varoquaux, G., Thirion, B., Robinson, E., Sharp, D.J., Edwards, A.D., Rueckert, D., 2011. A probabilistic framework to infer brain functional connectivity from anatomical connections. *Inf. Process. Med. Imaging* 22, 296–307.
- Dell’Acqua, F., Catani, M., 2012. Structural human brain networks: hot topics in diffusion tractography. *Curr. Opin. Neurol.* 25, 375–383.
- Dempster, A.P., 1972. Covariance selection. *Biometrics* , 157–175.
- Dobra, A., Lenkoski, A., Rodriguez, A., 2011. Bayesian inference for general Gaussian graphical models with application to multivariate lattice data. *J. Am. Statist. Assoc.* 106, 1418–1433.
- Fisher, R.A., 1915. Frequency distribution of the values of the correlation coefficient in samples of an indefinitely large population. *Biometrika* 10, 507–521.
- Friedman, J., Hastie, T., Tibshirani, R., 2008. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* 9, 432–441.
- Friston, K., 1994. Functional and effective connectivity in neuroimaging: A synthesis. *Hum. Brain Mapp.* 2, 56–78.
- Friston, K.J., 2011. Functional and Effective Connectivity: A Review. *Brain Connect.* 1, 13–36.
- Gong, G., He, Y., Concha, L., Lebel, C., Gross, D., Evans, A., Beaulieu, C., 2009. Mapping anatomical connectivity patterns of human cerebral cortex using in vivo diffusion tensor imaging tractography. *Cereb. Cortex* 19, 524–536.
- Greicius, M.D., Supekar, K., Menon, V., Dougherty, R.F., 2009. Resting-state functional connectivity reflects structural connectivity in the default mode network. *Cereb. Cortex* 19, 72–78.
- Groves, A.R., Beckmann, C.F., Smith, S.M., Woolrich, M.W., 2011. Linked independent component analysis for multimodal data fusion. *NeuroImage* 54, 2198–2217.
- Hagmann, P., Cammoun, L., Gigandet, X., Meuli, R., Honey, C.J., Wedeen, V.J., Sporns, O., 2008. Mapping the structural core of human cerebral cortex. *PLoS Biol.* 6, e159.
- Hastie, T., Tibshirani, R., Friedman, J., 2008. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, New York, NY. 2nd edition.
- Hermundstad, A.M., Bassett, D.S., Brown, K.S., Aminoff, E.M., Clewett, D., Freeman, S., Frithsen, A., Johnson, A., Tipper, C.M., Miller, M.B., Grafton, S.T., Carlson, J.M., 2013. Structural foundations of resting-state and task-based functional connectivity in the human brain. *Proc. Natl. Acad. Sci. USA* 110, 6169–6174.
- Hinne, M., Heskens, T., Beckman, C.F., van Gerven, M., 2013. Bayesian inference of structural brain networks. *NeuroImage* 66, 543–552.
- Honey, C.J., Kötter, R., Breakspear, M., Sporns, O., 2007. Network structure of cerebral cortex shapes functional connectivity on multiple time scales. *Proc. Natl. Acad. Sci. USA* , 10240–10245.
- Honey, C.J., Sporns, O., Cammoun, L., Gigandet, X., Thiran, J.P., Meuli, R., Hagmann, P., 2009. Predicting human resting-state functional connectivity from structural connectivity. *Proc. Natl. Acad. Sci. USA* 106, 2035–2040.
- Hutchison, R.M., Womelsdorf, T., Allen, E.A., Bandettini, P.A., Calhoun, V., Corbetta, M., et al, In press. Dynamic functional connectivity: Promises, issues, and interpretations. *NeuroImage* .
- Jbabdi, S., Woolrich, M.W., Andersson, J.L., Behrens, T.E., 2007. A Bayesian framework for global tractography. *NeuroImage* 37, 116–129.
- Jones, B., Carvalho, C., Dobra, A., Hans, C., Carter, C., West, M., 2005. Experiments in stochastic computation for high-dimensional graphical models. *Stat. Sci.* 20, 388–400.
- Koch, M., Norris, D.G., Hund-Georgiadis, M., 2002. An investigation of functional and anatomical connectivity using magnetic resonance imaging. *NeuroImage* 16, 241–250.
- Lauritzen, S.L., 1996. *Graphical Models*. Oxford University Press, New York.
- Le Bihan, D., Mangin, J.F., Poupon, C., Clark, C., Pappata, S., Molko, N., Chabriat, H., 2001. Diffusion tensor imaging: concepts and applications. *J. Magn. Reson. Im.* 13, 534–546.
- Lenkoski, A., Dobra, A., 2011. Computational aspects related to inference in Gaussian graphical models with the G -Wishart prior. *J. Comp. Graph. Stat.* 20, 140–157.
- Li, L., Rilling, J.K., Preuss, T.M., Glasser, M.F., Hu, X., 2012. The effects of connection reconstruction method on the interregional connectivity of brain networks via diffusion tractography. *Hum. Brain Mapp.* 33, 1894–1913.
- Lowe, M.J., Dzemidzic, M., Lurito, J.T., Mathews, V.P., Phillips, M.D., 2000. Correlations in low-frequency BOLD fluctuations reflect cortico-cortical connections. *NeuroImage* 12, 582–587.
- Marrelec, G., Benali, H., 2009. A theoretical investigation of the relationship between structural equation modeling and partial correlation in functional MRI effective connectivity. *Comput. Intell. Neurosci.* , 369341.
- 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.
- Medina, J.H., Pazo, J.H., 1981. Electrophysiological evidence for the existence of caudate-caudate connections. *Int. J. Neurosci.* 15, 99–101.
- Mensah, P., Deadwyler, S., 1974. The caudate nucleus of the rat: cell types and the demonstration of a commissural system. *J. Anat.* 117, 281–293.
- Moghaddam, B., Marlin, B., Khan, E., Murphy, K., 2009. Accelerating Bayesian structural inference for non-decomposable Gaussian graphical models, in: Bengio, Y., Schuurmans, D., Lafferty, J., Williams, C.K.I., Culotta, A. (Eds.), *Advances in Neural Information Processing Systems* 22. Curran Associates Inc., pp. 1285–1293.
- Ng, B., Varoquaux, G., Poline, J., Thirion, B., 2012. A novel sparse graphical approach for multimodel brain connectivity inference. *Med. Image Comput. Comput. Assist. Interv.* 15, 707–714.
- Patenaude, B., Smith, S., Kennedy, D., , M., J., 2011. A Bayesian model of shape and appearance for subcortical brain. *NeuroImage* 56, 907–922.
- Poser, B., Versluis, M., Hoogduin, J., Norris, D., 2006. BOLD contrast sensitivity enhancement and artifact reduction with multiecho EPI: parallel-acquired inhomogeneity-desensitized fMRI. *Magn. Reson. Med.* 55, 1227–1235.
- Pourahmadi, M., 2011. Covariance Estimation: The GLM and Regularization Perspectives. *Stat. Sci.* 26, 369–387.
- Roverato, A., 2002. Hyper inverse Wishart distribution for non-decomposable graphs and its application to Bayesian inference for Gaussian graphical models. *Scand. J. Stat.* 29, 391–411.
- Salvador, R., Suckling, J., Coleman, M.R., Pickard, J.D., Menon, D., Bullmore, E., 2005. Neurophysiological architecture of functional magnetic resonance images of human brain. *Cereb. Cortex* 15, 1332–1342.
- Schmidt, M., 2010. Graphical model structure learning with l_1 -regularization. Ph.D. thesis. University of British Columbia.

- Smith, A.D., Bolam, J.P., 1990. The neural network of the basal ganglia as revealed by the study of synaptic connections of identified neurones. *Trends Neurosci.* 13, 259–265.
- Smith, S.M., 2002. Fast robust automated brain extraction. *Hum. Brain Mapp.* 17, 143–155.
- Smith, S.M., 2012. The future of fMRI connectivity. *NeuroImage* 62, 1257–1266.
- Smith, S.M., Miller, K.L., Salimi-Khorshidi, G., Webster, M., Beckmann, C.F., Nichols, T.E., Ramsey, J.D., Woolrich, M.W., 2011. Network modelling methods for fMRI. *NeuroImage* 54, 875–891.
- Stephan, K.E., Tittgemeyer, M., Knösche, T.R., Moran, R.J., Friston, K.J., 2009. Tractography-based priors for dynamic causal models. *NeuroImage* 47, 1628–1638.
- Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M., 2002. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage* 15, 273–289.
- Varoquaux, G., Craddock, R.C., In press. Learning and comparing functional connectomes across subjects. *NeuroImage* .
- Varoquaux, G., Gramfort, A., Poline, J., Thirion, B., 2010. Brain covariance selection: better individual functional connectivity models using population prior, in: Zemel, R., Shawe-Taylor, J. (Eds.), *Advances in Neural Information Processing Systems*, John Lafferty, Vancouver, Canada. pp. 2334–2342.
- Venkataraman, A., Rathi, Y., Kubicki, M., Westin, C.F., Golland, P., 2010. Joint generative model for fMRI/DWI and its application to population studies. *Med. Image Comput. Comput. Assist. Interv.* 13, 191–199.
- Wang, H., Li, S.Z., 2012. Efficient Gaussian graphical model determination under G -Wishart distributions. *Elect. J. Stat.* 6, 168–198.
- Whittaker, J., 2009. *Graphical Models in Applied Multivariate Statistics*. Wiley Publishing. 1st edition.
- Woolrich, M.W., Stephan, K.E., In press. Biophysical network models and the human connectome. *NeuroImage* .
- Zhang, Y., Brady, M., Smith, S., 2001. Segmentation of brain MR images through a hidden Markov random field model and the expectation maximization algorithm. *IEEE Trans. Med. Imag.* 20, 45–57.