HHS Public Access

Author manuscript

Neuroimage. Author manuscript; available in PMC 2019 May 15.

Published in final edited form as:

Neuroimage. 2018 May 15; 172: 478–491. doi:10.1016/j.neuroimage.2018.01.029.

Improved estimation of subject-level functional connectivity using full and partial correlation with empirical Bayes shrinkage

Amanda F. Mejia[a,*], **Mary Beth Nebel**[b,c], **Anita D. Barber**[d], **Ann S. Choe**[e,f], **James J. Pekar**[e,f], **Brian S. Caffo**[g], and **Martin A. Lindquist**[g]

[a]Department of Statistics, Indiana University, USA

[b]Center for Neurodevelopmental and Imaging Research, Kennedy Krieger Institute, USA

[c]Department of Neurology, Johns Hopkins University, USA

[d]Center for Psychiatric Neuroscience, Feinstein Institute for Medical Research, USA

[e]Russell H. Morgan Department of Radiology and Radiological Science, Johns Hopkins University School of Medicine, USA

[f]F.M. Kirby Research Center for Functional Brain Imaging, Kennedy Krieger Institute, USA

[g]Department of Biostatistics, Johns Hopkins University, USA

Abstract

Reliability of sub ject-level resting-state functional connectivity (FC) is determined in part by the statistical techniques employed in its estimation. Methods that pool information across subjects to inform estimation of subject-level effects (e.g., Bayesian approaches) have been shown to enhance reliability of subject-level FC. However, fully Bayesian approaches are computationally demanding, while empirical Bayesian approaches typically rely on using repeated measures to estimate the variance components in the model. Here, we avoid the need for repeated measures by proposing a novel measurement error model for FC describing the different sources of variance and error, which we use to perform empirical Bayes shrinkage of subject-level FC towards the group average. In addition, since the traditional intra-class correlation coefficient (ICC) is inappropriate for biased estimates, we propose a new reliability measure denoted the mean squared error intra-class correlation coefficient (ICC_{MSE}) to properly assess the reliability of the resulting (biased) estimates. We apply the proposed techniques to test-retest resting-state fMRI data on 461 subjects from the Human Connectome Project to estimate connectivity between 100 regions identified through independent components analysis (ICA). We consider both correlation and partial correlation as the measure of FC and assess the benefit of shrinkage for each measure, as well as the effects of scan duration. We find that shrinkage estimates of subject-level FC exhibit substantially greater reliability than traditional estimates across various scan durations, even for the most reliable connections and regardless of connectivity measure. Additionally, we find partial

*Corresponding author. afmejia@iu.edu.

correlation reliability to be highly sensitive to the choice of penalty term, and to be generally worse than that of full correlations except for certain connections and a narrow range of penalty values. This suggests that the penalty needs to be chosen carefully when using partial correlations.

Keywords

functional connectivity; connectome; partial correlation; reliability; Bayesian statistics; shrinkage; measurement error; resting-state fMRI

1 Introduction

Measurement reliability is a persistent concern in psychological science (Button et al., 2013; Munafò et al., 2014; Collaboration, 2015). Functional connectivity (FC) of the brain, as measured using resting-state functional magnetic resonance imaging (rs-fMRI), is no exception (Shehzad et al., 2009). Driven by the growing role of subject-level FC estimates in fingerprinting (Finn et al., 2015; Airan et al., 2016), precision functional connectomics (Gordon et al., 2017), brain-behavior studies (Smith et al., 2015), and surgical planning (Tie et al., 2014), determining the best practices for reliable estimation of FC is an important and ongoing topic of research (e.g., Anderson et al., 2011; Birn et al., 2013; Laumann et al., 2015; Noble et al., 2017b). An analysis technique that has been shown to improve reliability of subject-level FC and related measures is *shrinkage*, a statistical estimation method in which individual observations "borrow strength" from a larger group of observations (Su et al., 2008; Varoquaux et al., 2010; Shou et al., 2014; Mejia et al., 2015; Dai et al., 2016; Chong et al., 2017; Rahim et al., 2017).

Shrinkage belongs to the more general family of Bayesian approaches. Fully Bayesian approaches, such as that proposed by Warnick et al. (2017), use a latent variable model in which the unknown connectivity for each subject gives rise to the unobserved "true" time series plus random noise, and subjects are drawn from some population distribution. In this framework, prior distributions are assumed on the parameters controlling the population distribution and the random noise, including the variance within and across subjects. Bayesian computation techniques like Markov chain Monte Carlo (MCMC) or variational Bayes (VB) are used to estimate or sample from the posterior distribution of each parameter and latent variable in the model. The posterior distribution of the connectivity for each subject can then be used to obtain estimates through the posterior mode as well as inference through the posterior quantiles.

Since fully Bayesian approaches tend to be computationally intensive, empirical Bayesian approaches are often employed as an efficient and convenient alternative. In the empirical Bayesian framework, certain parameters are estimated a-priori using the data or prior knowledge obtained from existing studies. Given these parameter estimates, the desired posterior quantities often have a closed-form solution, greatly facilitating computation. Empirical Bayes shrinkage estimators are an example of this approach and result from assuming a measurement error model on a set of estimates. For example, in our case of a Gaussian population prior with independent Gaussian errors, the empirical Bayes shrinkage estimates are weighted combinations of the subject-level observation and the group average,

where the degree of shrinkage towards the group average that gives rise to the posterior mean and minimizes mean squared error (MSE) relative to the truth is equal to the ratio of within-subject variance to total (within-subject plus between-subject) variance (James and Stein, 1961; Efron and Morris, 1975). Therefore, lower within-subject variance combined with higher between-subject variance leads to less shrinkage of subject-level estimates toward the group, while higher within-subject variance and lower between-subject variance leads to greater shrinkage.

Previous work has clearly illustrated the benefits of shrinkage, with 25–30% gain in reliability of subject-level connectivity (Varoquaux et al., 2010; Shou et al., 2014; Dai et al., 2016; Rahim et al., 2017) and parcellations (Mejia et al., 2015; Chong et al., 2017). However, estimating the relevant variance components to determine the degree of shrinkage has typically relied on having access to repeated measures through test-retest fMRI data. This limits the applicability of shrinkage methods, since in many studies only a single rs-fMRI session is available for most if not all subjects, and even if multiple sessions were available one would want to utilize the full data available for each subject to improve estimation.

In this work, we propose a novel method to compute empirical Bayes shrinkage estimates of FC, where the degree of shrinkage is determined using single-session fMRI data. Previous work has proposed using “pseudo test-retest” data, in which a single scanning session is split into two contiguous sub-sessions, as a proxy for inter-session variance (Mejia et al., 2015; Mueller et al., 2015). However, this will tend to overestimate the sampling variance of FC, since fewer time points are used in its estimation. In Mejia et al. (2015), we proposed using an empirically determined adjustment factor to correct for this, but the generalizability of such an approach is limited. Here, we instead propose a measurement error model for FC and describe how this model can be used to estimate within-subject variance of FC using single-session fMRI data. Leveraging recent developments in the study of moment-to-moment changes in FC, this model assumes that within-subject variance of FC comes not only from sampling error, but from changes in true FC over time, i.e. dynamic connectivity (Allen et al., 2014). The measurement error model, resulting shrinkage estimator, and variance component estimation techniques are described in Section 2.1.

Assessing the reliability of shrinkage estimates is also a challenge, since the intra-class correlation coefficient (ICC), a commonly used and interpretable reliability metric, is not appropriate for biased estimators. Therefore, most reliability studies for shrinkage estimates have relied on mean squared error (MSE) using simulations or test-retest fMRI data to illustrate the gains in reliability due to shrinkage. However, MSE is sensitive to measurement scale and lacks the convenient interpretation of ICC, which ranges from 0 to 1 and represents the proportion of variance in the observations due to true between-subject differences rather than within-subject error or deviation. We therefore propose combining ICC and MSE into a novel reliability measure for biased or unbiased estimators, ICC_{MSE}. ICC_{MSE} is equal to ICC for unbiased estimators but is also appropriate for biased estimators and allows for fair and intuitive comparison between shrinkage and traditional estimators. We motivate and describe this measure in Section 2.2.

Finally, we explore the role of scan duration in reliability of both shrinkage and traditional estimates of FC, as the effect of scan duration on reliability of FC is a topic of much recent interest (e.g., Shehzad et al., 2009; Van Dijk et al., 2010; Anderson et al., 2011; Birn et al., 2013; Laumann et al., 2015; Noble et al., 2017a). Several recent studies have also observed substantial differences in reliability across connections. For example, connections within the default mode network (DMN) have been found to exhibit particularly high reliability even for short scan duration, while connections involving the motor network tend to exhibit poor reliability (Shehzad et al., 2009; Van Dijk et al., 2010; Anderson et al., 2011; Laumann et al., 2015; Mueller et al., 2015; Finn et al., 2015). Therefore, we also explore the relationship between scan duration and reliability for connections within different resting-state networks. In addition, most of the extant literature on the relationship between scan duration and FC reliability uses Pearson correlation coefficients as the primary measure of the degree of connectivity between brain regions. The issue of sufficient scan duration deserves greater investigation in the context of partial correlations, which are becoming increasingly popular for their ability to distinguish between brain regions that are directly versus indirectly correlated (Smith et al., 2011; Varoquaux and Craddock, 2013; Smith et al., 2015; Wang et al., 2016).

We perform a reliability analysis using data from the Human Connectome Project (HCP) to examine the role of scan duration, shrinkage, and connectivity measure (full and partial correlation) on reliability of functional connectivity. The HCP is ideal for this analysis due to its large sample size and relatively long duration of rs-fMRI scans. We assess reliability at multiple levels: omnibus reliability over all connections, reliability of within-network connections, reliability of all connections with a particular seed, and reliability of individual connections. This multi-resolution approach provides a more complete picture of reliability and illustrates that reliability of FC is more compex than a single measure can detect. For estimating partial correlations through ridge regression, we first perform a reliability study to assess the impact of the regularization parameter, ρ. We find that certain values of ρ lead to partial correlation estimates with improved reliability for *particular* connections but worse reliability overall compared with full correlations. Notably, we find that common choices of ρ, such as 0.01, lead to partial correlations with much worse reliability than full correlations. The reliability study is described in Section 3, and we conclude with a discussion in Section 4.

2 Methods

2.1 Empirical Bayes shrinkage using single-session data

Let $q = 1, \ldots, Q$ index the nodes (voxels, vertices or regions) between which we wish to estimate pairwise connectivity, and consider a single pair of nodes q and q'. The overall estimate of FC obtained from an fMRI session is often referred to as *static* connectivity, in contrast to *dynamic* connectivity across the session. As we describe in the next section, considering static connectivity as an average over a dynamic connectivity time series enables us to use the central limit theorem to model and estimate the relevant variance components. Hereafter, the measure of FC is assumed to be Fisher-transformed correlation or partial correlation.

2.1.1 A measurement error model for functional connectivity—Assume here that all subjects have the same scan duration, and let $\mathcal{T} = \{1, \ldots, T\}$ index the fMRI time series for each subject (see Appendix A for the case when subjects have differing scan duration). Let $x_{i,t}(q, q')$ be the *unobserved true* connectivity between q and q' at time $t \in \mathcal{T}$, which we can write as $x_{i,t}(q, q') = \mu_i(q, q') + \delta_{i,t}(q, q')$, where $\mu_i(q, q')$ is the long-term average connectivity for subject i, and $\delta_{i,t}(q, q')$ represents dynamic fluctuations in connectivity over time. Since fMRI data contains noise, the *observed* or *estimated* connectivity $w_{i,t}(q, q')$ at time t can be written as the truth $x_{i,t}(q, q')$ plus noise $\varepsilon_{i,t}(q, q')$, that is

$$w_{i,t}(q,q') = x_{i,t}(q,q') + \varepsilon_{i,t}(q,q') = \mu_i(q,q') + \delta_{i,t}(q,q') + \varepsilon_{i,t}(q,q').$$

We assume that $\mu_i(q,q') \sim N\left\{\mu(q,q'), \sigma^2_\mu(q,q')\right\}$, where $\mu(q, q')$ is the average population-level connectivity between q and q' and $\sigma^2_\mu(q,q')$ is its between-subject variance, and that $\delta_{i,t}(q,q') \sim \left\{0, \sigma^2_\delta(q,q')\right\}$ and $\varepsilon_{i,t}(q,q') \sim \left\{0, \sigma^2_\varepsilon(q,q')\right\}$. Since the connectivity measures are Fisher-transformed correlations, it is reasonable to assume that $\delta_{i,t}(q, q')$ and $\varepsilon_{i,t}(q, q')$ are independent, although each may exhibit autocorrelation due to temporal dependence in fMRI data.

For analyses of static connectivity, we do not actually estimate the dynamic time series of connectivity, but rather its average across the fMRI session $\mathcal{T}$, which can be written

$$w_{i,\mathcal{T}}(q,q') = \mu_i(q,q') + \delta_{i,\mathcal{T}}(q,q') + \varepsilon_{i,\mathcal{T}}(q,q'), \quad (1)$$

where by the central limit theorem for dependent observations, $\delta_{i,\mathcal{T}}(q,q')$ and $\varepsilon_{i,\mathcal{T}}(q,q')$ are approximately distributed

$$\delta_{i,\mathcal{T}}(q,q') = \frac{1}{T}\sum_{t=1}^{T} \delta_{i,t}(q,q') \sim N\left\{0, T_\delta^{-1}\sigma^2_\delta(q,q')\right\} \text{and}$$
$$\varepsilon_{i,\mathcal{T}}(q,q') = \frac{1}{T}\sum_{t=1}^{T} \varepsilon_{i,t}(q,q') \sim N\left\{0, T_\varepsilon^{-1}\sigma^2_\varepsilon(q,q')\right\}.$$

Here, $T_\delta = T/\tau_\delta$ is the effective sample size (ESS) of the timeseries $\left\{\delta_{i,t}(q,q')\right\}_{t \in \mathcal{T}}$, and $T_\varepsilon = T/\tau_\varepsilon$ is the ESS of $\left\{\varepsilon_{i,t}(q,q')\right\}_{t \in \mathcal{T}}$, where τ_δ and τ_ε are the unknown *autocorrelation times* of $\left\{\delta_{i,t}(q,q')\right\}_{t \in \mathcal{T}}$ and $\left\{\varepsilon_{i,t}(q,q')\right\}_{t \in \mathcal{T}}$, respectively (Kass et al., 1998; Thompson, 2011; Gong and Flegal, 2016). Various techniques exist for estimating autocorrelation time, and in the case of an autoregressive (AR) process, there is a closed-form solution involving only the partial autocorrelation function of the process (Thompson, 2011); however, the proposed shrinkage methods described below do not require estimation of τ_δ or τ_ε.

Equation (1) describes a measurement error model for static connectivity estimated from an fMRI time series of length T. The variance components are $T_\delta^{-1}\sigma_\delta^2(q,q')$, the *within-subject, across-session variance* of true connectivity; $T_\varepsilon^{-1}\sigma_\varepsilon^2(q,q')$, the *within-subject sampling variance*; and $\sigma_\mu^2(q,q') = :\sigma_{between}^2(q,q')$, the *between-subject variance.* The total within-subject variance is therefore $\sigma_{within}^2(q,q'): = T_\delta^{-1}\sigma_\delta^2(q,q') + T_\varepsilon^{-1}\sigma_\varepsilon^2(q,q')$. The empirical Bayes shrinkage estimator of $\mu_i(q, q')$ is given as in Morris (1983) by

$$\widetilde{w}_{i,\mathcal{T}}(q,q') = \lambda(q,q')\overline{w}_{\mathcal{T}}(q,q') + \{1-\lambda(q,q')\}w_{i,\mathcal{T}}(q,q'),$$

where $\overline{w}_{\mathcal{T}}(q,q') = \frac{1}{n}\sum_{i=1}^{n} w_{i,\mathcal{T}}(q,q')$ is the estimated population average connectivity, and the degree of shrinkage $\lambda(q, q')$ is equal to the ratio of within-subject variance to total (within-subject plus between-subject) variance of $w_{i,\mathcal{T}}(q,q')$, given by

$$\lambda(q,q') = \frac{\sigma_{within}^2(q,q')}{\sigma_{within}^2(q,q') + \sigma_{between}^2(q,q')} = \frac{T_\delta^{-1}\sigma_\delta^2(q,q') + T_\varepsilon^{-1}\sigma_\varepsilon^2(q,q')}{T_\delta^{-1}\sigma_\delta^2(q,q') + T_\varepsilon^{-1}\sigma_\varepsilon^2(q,q') + \sigma_\mu^2(q,q')}, \quad (2)$$

We now describe how to estimate the relevant variance components to achieve the optimal degree of shrinkage, which minimizes MSE of the resulting estimates relative to the truth.

2.1.2 Variance component estimation—The total variance on the denominator of $\lambda(q, q')$ is easily estimated as $Var_i\left\{w_{i,\mathcal{T}}(q,q')\right\}$. If a second fMRI session $\mathcal{T}'$ the same length as $\mathcal{T}$ were available for each subject, the within-subject variance could also be estimated by taking the differences for each pair of subject-level measurements, $w_{i,\mathcal{T}'}(q,q') - w_{i,\mathcal{T}}(q,q')$, since

$$\begin{aligned} Var_i\left\{w_{i,\mathcal{T}'}(q,q') - w_{i,\mathcal{T}}(q,q')\right\} &= Var_i\left\{\delta_{i,\mathcal{T}'}(q,q') + \varepsilon_{i,\mathcal{T}'}(q,q') - \delta_{i,\mathcal{T}}(q,q') - \varepsilon_{i,\mathcal{T}}(q,q')\right\} \\ &= 2\mathrm{Var}\left\{\delta_{i,\mathcal{T}}(q,q')\right\} + 2\mathrm{Var}\left\{\varepsilon_{i,\mathcal{T}}(q,q')\right\} \\ &= 2T_\delta^{-1}\sigma_\delta^2(q,q') + 2T_\varepsilon^{-1}\sigma_\varepsilon^2(q,q')\Big\} \\ &= 2\sigma_{within}^2(q,q'). \end{aligned}$$

We call this the "oracle" estimate of within-subject variance. However, since such repeated measurements are often not available in practice, we propose the following estimation procedure for within-subject variance using single-session fMRI data.

Assume without loss of generality that T is even, and define the two sub-time series $\mathcal{T}_1 = \left\{1, \ldots, \frac{T}{2}\right\}$ and $\mathcal{T}_2 = \left\{\frac{T}{2}+1, T\right\}$.[1] Consider $w_{i,\mathcal{T}_1}(q,q')$ and $w_{i,\mathcal{T}_2}(q,q')$, the estimates of

[1] If T is odd, simply let $h = \lfloor T/2 \rfloor$ and let $\mathcal{T}_1 = \{1, \ldots, h\}$ and $\mathcal{T}_2 = \{T-h+1, \ldots, T\}$.

static connectivity within each sub-time series. Applying the measurement error model in equation (1) we can write

$$w_{i,\mathcal{T}_1}(q,q') = \mu_i(q,q') + \delta_{i,\mathcal{T}_1}(q,q') + \varepsilon_{i,\mathcal{T}_1}(q,q') \text{and}$$
$$w_{i,\mathcal{T}_2}(q,q') = \mu_i(q,q') + \delta_{i,\mathcal{T}_2}(q,q') + \varepsilon_{i,\mathcal{T}_2}(q,q').$$

Note that $w_{i,\mathcal{T}_1}(q,q')$ and $w_{i,\mathcal{T}_2}(q,q')$ have $\mu_i(q,q')$ in common, since it represents long-term average connectivity for subject i. We may assume that $\delta_{i,\mathcal{T}_1}(q,q')$ and $\delta_{i,\mathcal{T}_2}(q,q')$ (and $\varepsilon_{i,\mathcal{T}_1}(q,q')$ and $\varepsilon_{i,\mathcal{T}_2}(q,q')$) are approximately independent, as long as $T/2$ is large relative to the autocorrelation in the time series.[2] Note that since $\mathcal{T}_1$ and $\mathcal{T}_2$ have the same autocorrelation structure as $\mathcal{T}$, the autocorrelation time for $\delta_{i,\mathcal{T}_j}(q,q'), j = 1,2$, is also τ_δ. Hence, the ESS of $\left\{\delta_{i,t}(q,q')\right\}_{t\in\mathcal{T}_j}$ is $(T/2)/\tau_\delta = T_\delta/2$. Similarly, the autocorrelation time for $\varepsilon_{i,\mathcal{T}_j}(q,q'), j = 1,2$, is τ_ε, and its ESS is $T_\varepsilon/2$. Therefore, for $j = 1, 2$,

$$\delta_{i,\mathcal{T}_j}(q,q') \sim N\left\{0, \left(T_\delta/2\right)^{-1}\sigma^2_\delta(q,q')\right\} \text{and}$$
$$\varepsilon_{i,\mathcal{T}_j}(q,q') \sim N\left\{0, \left(T_\varepsilon/2\right)^{-1}\sigma^2_\varepsilon(q,q')\right\}.$$

Taking the difference $w_{i,\mathcal{T}_1}(q,q') - w_{i,\mathcal{T}_2}(q,q')$ for each subject, we can compute

$$\begin{aligned} Var_i\left\{w_{i,\mathcal{T}_1}(q,q') - w_{i,\mathcal{T}_2}(q,q')\right\} &= Var_i\left\{\delta_{i,\mathcal{T}_1}(q,q') + \varepsilon_{i,\mathcal{T}_1}(q,q') - \delta_{i,\mathcal{T}_2}(q,q') - \varepsilon_{i,\mathcal{T}_2}(q,q')\right\} \\ &= 2Var_i\left\{\delta_{i,\mathcal{T}_1}(q,q')\right\} + 2Var_i\left\{\varepsilon_{i,\mathcal{T}_1}(q,q')\right\} \\ &= 4T_\delta^{-1}\sigma^2_\delta(q,q') + 4T_\varepsilon^{-1}\sigma^2_\varepsilon(q,q') \\ &= 4\sigma^2_{within}(q,q'). \end{aligned}$$

Therefore, $\hat{\sigma}^2_{within}(q,q') = \frac{1}{4}Var_i\left\{w_{i,\mathcal{T}_1}(q,q') - w_{i,\mathcal{T}_2}(q,q')\right\}$ serves as a single-session estimate of the within-subject variance of $w_{i,\mathcal{T}}(q,q')$. Using this, along with the estimate of total

[2] To essentially eliminate the small degree of dependence between the time series indexed by $\mathcal{T}_1$ and $\mathcal{T}_2$ induced by auto-correlation, let $h = \lfloor T/2 \rfloor - d$, where d is the number of time points truncated from the end of $\mathcal{T}_1$ and the beginning of $\mathcal{T}_2$. The ESS of $\left\{\delta_{i,t}(q,q')\right\}_{t\in\mathcal{T}_j}$ is then $T_\delta/(T/h)$, and the ESS of $\left\{\varepsilon_{i,t}(q,q')\right\}_{t\in\mathcal{T}_j}$ is $T_\varepsilon/(T/h)$, $j = 1, 2$. Therefore, $\hat{\sigma}^2_{within}(q,q') = (2T/h)^{-1}Var_i\left\{w_{i,\mathcal{T}_1}(q,q') - w_{i,\mathcal{T}_2}(q,q')\right\}$.

variance, $Var_i\left\{w_{i,\mathcal{T}}(q,q')\right\}$, we can determine the optimal degree of shrinkage given in equation (2) for each connection.

2.2 Assessing reliability of biased estimators with ICC_{MSE}

To assess the benefits of shrinkage for enhancing reliability of FC estimates, we need a meaningful way to quantify and compare the accuracy of both traditional and shrinkage estimates. One measure commonly used to quantify reliability of a measure is the intra-class correlation coefficient (ICC). The ICC of an estimate is equal to the ratio of between-subject variance to total (between- plus within-subject) variance. Therefore, the optimal degree of shrinkage of an estimate is equal 1 – ICC of that estimate. This illustrates how the degree of shrinkage is related to the reliability, with very reliable estimates (ICC close to 1) receiving very little shrinkage towards the group mean, and very unreliable estimates (ICC close to 0) receiving almost complete shrinkage to the group mean.

While ICC is a popular measure of reliability due to its interpretability, it is not a meaningful measure of reliability for shrinkage estimates. This is because shrinkage trades bias for a reduction in variance to achieve an overall reduction in mean squared error (MSE), equal to variance plus squared bias. Therefore, only considering the variance of a shrinkage estimate provides an incomplete and overly optimistic picture of its accuracy. We propose a new measure of reliability for both biased and unbiased estimates, ICCmse, which replaces within-subject variance in the ICC formula with within-subject MSE. For unbiased estimates, including traditional estimates of FC with no shrinkage, ICC_{MSE} is the same as ICC, since MSE equals within-subject variance when bias is zero. However, for shrinkage estimates ICC_{MSE} considers both within-subject variance and squared bias to provide an overall picture of reliability, on the same scale as ICC. For the traditional and shrinkage estimates $w_{i,\mathcal{T}}(q,q')$ and $\widetilde{w}_{i,\mathcal{T}}(q,q')$, ICC_{MSE} is computed as

$$\mathrm{ICC}_{MSE}\left\{w_{i,\mathcal{T}}(q,q')\right\} = \frac{\hat{\sigma}^2_{\mu}(q,q')}{\hat{\sigma}^2_{\mu}(q,q') + MSE\left\{w_{i,\mathcal{T}}(q,q')\right\}}$$
$$\mathrm{ICC}_{MSE}\left\{\widetilde{w}_{i,\mathcal{T}}(q,q')\right\} = \frac{\hat{\sigma}^2_{\mu}(q,q')}{\hat{\sigma}^2_{\mu}(q,q') + MSE\left\{\widetilde{w}_{i,\mathcal{T}}(q,q')\right\}},$$

where we estimate MSE as the mean squared difference (MSD) across repeated estimates, divided by two since both sets of estimates are observed with noise. That is, letting $\mathcal{T}'$ index a second fMRI session the same length as $\mathcal{T}$,

$$MSE\left\{w_{i,\mathcal{T}}(q,q'))\right\} = \frac{1}{2n}\sum_{i=1}^{n}\left\{w_{i,\mathcal{T}}(q,q') - w_{i,\mathcal{T}'}(q,q')\right\}^2,$$
$$MSE\left\{\widetilde{w}_{i,\mathcal{T}}(q,q')\right\} = \frac{1}{2n}\sum_{i=1}^{n}\left\{\widetilde{w}_{i,\mathcal{T}}(q,q') - w_{i,\mathcal{T}'}(q,q')\right\}^2.$$

Note that for both the traditional and shrinkage estimates, in the absence of the ground truth connectivity we use the *traditional* estimate from a second fMRI session to estimate MSE. In

particular, we do not use shrinkage estimates from the second session to assess reliability of shrinkage estimates; this would artificially reduce MSE, since estimates from both sessions would be moved towards the group average, which is very stable across visits. Instead, using the traditional estimate from the second session in place of the truth allows the shrinkage and traditional estimates to be compared fairly.

If we wish to quantify the reliability of an entire seed connectivity map, the image intra-class correlation coefficient (I2C2) is often used instead of the ICC (Shou et al., 2013). Using the ICC_{MSE} framework, we can assess the accuracy of traditional and shrinkage estimates of connectivity maps and matrices using measures analogous to the I2C2. For seed q, let $\mathbf{w}_{i,\mathcal{T}}(q)$ and $\widetilde{\mathbf{w}}_{i,\mathcal{T}}(q)$ be the traditional and shrinkage estimates, respectively, of the true $Q \times 1$ seed connectivity map $\boldsymbol{\mu}_i(q)$. Then, the I2C2_{MSE} of each estimated image is computed as

$$\text{I2C2}_{MSE}\left\{\mathbf{w}_{i,\mathcal{T}}(q)\right\} = \frac{\sum_{q' \neq q} \hat{\sigma}^2_{\mu}(q,q')}{\sum_{q' \neq q} \hat{\sigma}^2_{\mu}(q,q') + \sum_{q' \neq q} MSE\left\{w_{i,\mathcal{T}}(q,q')\right\}},$$
$$\text{I2C2}_{MSE}\left\{\widetilde{\mathbf{w}}_{i,\mathcal{T}}(q)\right\} = \frac{\sum_{q' \neq q} \hat{\sigma}^2_{\mu}(q,q')}{\sum_{q' \neq q} \hat{\sigma}^2_{\mu}(q,q') + \sum_{q' \neq q} MSE\left\{\widetilde{w}_{i,\mathcal{T}}(q,q')\right\}}.$$

Similarly, letting $\mathbf{W}_{i,\mathcal{T}}$ and $\widetilde{\mathbf{W}}_{i,\mathcal{T}}$ be the traditional and shrinkage estimates, respectively, of the true $Q \times Q$ connectivity matrix $\mathbf{M}_i$, the omnibus ICC_{MSE} (oICC_{MSE}) of each estimate is computed as

$$\text{oICC}_{MSE}\left\{\mathbf{W}_{i,\mathcal{T}}\right\} = \frac{\sum_{q}\sum_{q' > q} \hat{\sigma}^2_{\mu}(q,q')}{\sum_{q}\sum_{q' > q} \hat{\sigma}^2_{\mu}(q,q') + \sum_{q}\sum_{q' > q} MSE\left\{w_{i,\mathcal{T}}(q,q')\right\}},$$
$$\text{oICC}_{MSE}\left\{\widetilde{\mathbf{W}}_{i,\mathcal{T}}\right\} = \frac{\sum_{q}\sum_{q' > q} \hat{\sigma}^2_{\mu}(q,q')}{\sum_{q}\sum_{q' > q} \hat{\sigma}^2_{\mu}(q,q') + \sum_{q}\sum_{q' > q} MSE\left\{\widetilde{w}_{i,\mathcal{T}}(q,q')\right\}}.$$

ICC_{MSE}, I2C2_{MSE} and omnibus ICC_{MSE} each range from 0 to 1, where 0 signifies no reliability and 1 indicates perfect reliability across repeated measurements from the same subject. They reduce to ICC, I2C2, and omnibus ICC, respectively, in the absence of bias, which is the case for traditional estimates.

3 Reliability study

We use test-retest data from the Human Connectome Project (Hep) to assess the reliability of traditional and shrinkage estimates of whole-brain FC between 100 regions identified using independent components analysis (ICA). Leveraging the relatively long duration of the resting-state fMRI data in the HCP, we also also examine how scan length affects the degree of shrinkage and the reliability of both traditional and shrinkage estimates of FC. We also consider and compare full Pearson correlation and partial correlation estimated using ridge regression.

3.1 Data and processing

The HCP (Van Essen et al., 2013) is a collection of neuroimaging and phenotypic information for over one thousand healthy adult subjects (http://humanconnectome.org). For the analyses described below, we used the following data from the 523 subjects included in the 2014 Human Connectome Project 500 Parcellation+Timeseries+Netmats (HCP500-PTN) release. All MRI data were acquired on a customized 3T Siemens connectome-Skyra 3T scanner, designed to achieve 100 mT/m gradient strength. For 461 subjects, a multi-band/multi-slice pulse sequence with an acceleration factor of eight (Moeller et al., 2010; Feinberg et al., 2010; Setsompop et al., 2012; Xu et al., 2012; Uğurbil et al., 2013) was used to acquire four roughly 15-minute rs-fMRI sessions, each consisting of 1200 volumes sampled every 0.72 seconds at 2 mm isotropic spatial resolution. The sessions were collected over two visits that occurred on separate days, with two runs collected at each visit. Across runs at each visit, phase encoding directions were alternated between right-to-left (RL) and left-to-right (LR) directions.

Spatial preprocessing was performed using the minimal preprocessing pipeline as described by Glasser et al. (2013), which includes correcting for spatial distortions and artifacts and projection of the data time series to the standard grayordinate space. Structured artifacts in the time series were removed using ICA + FIX (independent component analysis followed by FMRIB's ICA-based X-noiseifier; Salimi-Khorshidi et al., 2014; Griffanti et al., 2014), and each data set was temporally demeaned with variance normalization according to Beckmann and Smith (2004). As part of the HCP500-PTN release, group independent component analysis (GICA) was performed on the full rs-fMRI time series for all 461 subjects to estimate a set of 100 spatial independent components (ICs) (Beckmann and Smith, 2004). Time courses were estimated for each subject and IC by performing the first stage of dual regression (Beckmann et al., 2009). Specifically, the group IC spatial maps were used as predictors in a multivariate linear regression model against the full rs-fMRI time series. To eliminate acquisition-related variability, we only used the IC time courses associated with the runs acquired with LR phase encoding, which corresponded to observations 1–1200 (visit 1) and 2401–3600 (visit 2) within the IC text file for each subject.[3]

We assigned each IC to one of seven resting-state networks (RSNs) based on overlap with the Allen parcellation, a publicly available set of 100 ICs that have been previously classified as resting state networks (RSNs) or noise by a group of experts (Allen et al., 2014). The 50 Allen ICs classified as RSNs are organized into seven large functional groups: default mode (DMN), cognitive-control (CC), visual (V), sensorimotor (SM), auditory (A), cerebellar (CB) and sub-cortical (SC) networks. As in Choe et al. (2017), we calculated the percent variance explained by the seven sets of Allen RSNs for each of the 100 HCP ICs to determine network membership. Several ICs were reassigned based on visual inspection of their spatial distributions and temporal spectra averaged across subjects, resulting in the final

[3]Although the phase encoding direction order was reversed across visits for all data acquired after October 1, 2012 (RL/LR on visit 1 and LR/RL on visit 2), for the GICA included in the HCP500-PTN release the runs of both visits were reordered to LR/RL before concatenating all four sessions.

RSN assignments shown in Figure 1. Six HCP ICs were identified as likely representing nuisance signals; their time courses were regressed from the remaining 94 HCP signal ICs.

3.2 Reliability analysis

The quantity of interest for each subject is the $Q \times Q$ matrix of pairwise FC between each of the $Q = 94$ signal ICs. We estimate the full correlation between each pair of ICs using Pearson correlation. We also estimate the partial correlation between each pair of ICs as described in the following section. Both measures are Fisher-transformed before performing shrinkage.

3.2.1 Partial correlation estimation—We estimate partial correlation using ridge regression as proposed by Ha and Sun (2014). Specifically, given a sample correlation matrix $\mathbf{S}$ ($p \times p$), the partial correlation matrix using ridge regression is given by $\mathbf{R} = -\text{scale}\{(\mathbf{S} + \rho\mathbf{I}_p)^{-1}\}$, where $\text{scale}(\mathbf{A}) = \text{diag}(\mathbf{A})^{-1/2}\mathbf{A}\text{diag}(\mathbf{A})^{-1/2}$ (Ha and Sun, 2014). Several heuristics have been proposed to choose the regularization parameter, ρ. For instance, the discrepancy principle tries to find the point at which the residual of the regularized solution is comprised only of noise but assumes that the standard deviation of the noise is known (Morozov, 1984). The L-curve criterion tries to balance regularization error and perturbation error by finding the L-shaped corner of the log-log plot of the regularized solution versus the norm of the corresponding residual vector (Hansen, 1992). Generalized cross-validation attempts to minimize the average prediction error or maximize the out-of-sample likelihood in group studies (Golub et al., 1979; Varoquaux et al., 2010; Varoquaux and Craddock, 2013). However, many functional connectivity studies employing ridge regression choose ρ in an ad-hoc fashion, with values ranging from 0.01 (e.g., Smith et al., 2015) to 1 or larger, and currently no consensus exists on the best choice or method of choosing ρ. Therefore, we first conduct a reliability analysis to assess the impact of the choice of ρ on the reliability of partial correlation estimates for varying scan duration.

For $\rho = \{0.01, 0.05, 0.1, 0.5, 1, 5, 10, 50, 100\}$, we estimate the partial correlation matrix for each subject using the first $\ell = \{100, 200, \ldots, 1200\}$ volumes of both visits. For comparison, we also compute estimates of full correlation for each scan duration. We use the resulting set of repeated estimates to compute the omnibus intra-class correlation coefficient (oICC) of the partial and full correlation matrix estimates as an overall measure of their reliability. As with ICC, omnibus ICC ranges from 0 to 1, with 0 indicating that the estimates contain no subject-level information and 1 indicating that they are perfectly reliable across multiple observations of the same subject. Note that the choice of ρ changes the scale of partial correlation values, with larger ρ resulting in smaller partial correlations. Therefore, partial correlations obtained using different ρ values cannot be compared directly or with full correlations; however, ICC is scale-free and can be compared across measures on different scales. To assess the reliability of full and partial correlations for different types of connections, we also compute the average ICC across all *within-network* connections for each network shown in Figure 1.

Figure 2 displays the results of this reliability analysis. Figure 2a shows that, as expected, reliability increases with scan duration for each measure. More surprisingly, however, we

observe that the reliability of partial correlation estimates using small values of ρ are much less reliable than full correlations; reliability increases with ρ up to $\rho = 50$, with $\rho = 50$ and $\rho = 100$ resulting in partial correlations with similar overall reliability to that of full correlations. Figure 2b shows that for highly reliable connections, e.g. those within the DMN and CC networks, partial correlations are more reliable than full correlations for p greater than 0.5, and reliability of these connections is maximized at $\rho = 5$. For other networks, partial correlations are generally less reliable than full correlations for smaller values of ρ and achieve similar reliability for larger values of ρ. Based on these findings, we use $\rho = 5$ to estimate partial correlations, as this choice maximizes the reliability of the most consistent connections and achieves somewhat comparable reliability to full correlations for all other within-network connections.

3.2.2 Shrinkage estimation of functional connectivity—We use the first visit from each subject to compute traditional and shrinkage estimates of functional connectivity (full and partial correlation) based on varying scan length. We reserve the second visit to assess the accuracy of each of those estimates. Specifically, as described in Section 2.1 we compute traditional and shrinkage estimates of the FC matrix using the first $\ell = 100, 200, \ldots, 1200$ volumes of visit 1 for each subject. With a TR of 0.72 seconds, the resulting time series range in duration from 1.2 to 14.4 minutes. Using the second visit, we compute only the traditional estimate of the FC matrix using the full time series (1200 volumes) for each subject as an unbiased proxy for the ground truth. We also assess the performance of an "oracle" shrinkage estimator, which uses the oracle estimate of within-subject variance described at the beginning of Section 2.1.2. This provides an upper bound on the reliability of shrinkage estimators relative to the second visit, since it is based on the actual within-subject variance across the two visits.

For each scan duration, we assess the reliability of traditional, shrinkage and oracle estimates at multiple resolutions. First,/we compute ICC_{MSE} for each connection (q, q'), $q' \neq q = 1, \ldots, 100$. Second, we compute $I2C2_{MSE}$ for the seed connectivity map using each IC $q = 1, \ldots, 100$ as the seed. Third, we compute the omnibus reliability of the connectivity matrix using $oICC_{MSE}$. Finally, we investigate reliability of traditional and shrinkage estimates of FC for connections within each resting-state network shown in Figure 1. For each network, we compute the average ICC_{MSE} of traditional and shrinkage estimates of FC across all within-network connections to determine the overall reliability of within-network connectivity at each scan duration.

3.3 Results

Throughout this section, we present the results using *full* correlations as the measure of FC in the upper triangle of each matrix and those using *partial* correlations (using $\rho = 5$) in the lower triangle.

Figure 3 displays the group-average connectivity matrix based on all 1200 volumes (approximately 15 minutes) of visit 1, which serves as the target for shrinkage of subject-level estimates based on the same scan duration. Figure 4 displays traditional and single-session shrinkage estimates of FC for three randomly selected subjects based on all 1200

volumes of visit 1, along with the "gold-standard" for each, the traditional estimate based on visit 2, which we use to assess reliability of the visit 1 estimates. The root MSE (rMSE) of each estimate, relative to visit 2, is displayed above the matrix. For all three subjects, the shrinkage estimate is closer to the observed connectivity at visit 2 (lower rMSE) than the traditional estimate. This improvement is seen for both full and partial correlations, although note that rMSE is not comparable across the two measures due to differences in scale.

Figure 4 also illustrates how the unique features of each subject are not lost through shrinkage. The shrinkage estimates are qualitatively similar to the traditional estimates for each subject, and many similar features are seen in both. For example, subject 1 displays strong anticorrelations between some cerebellar regions and the visual, sensorimotor and auditory networks, features that are preserved after shrinkage and replicated at visit 2. Subject 3 displays unusually strong negative partial correlations between some DMN and CC regions, features that are again preserved after shrinkage and reliable across both visits. This illustrates how, rather than discarding true subject-level information, shrinkage serves to reduce noise by bringing unreliable, noisy features closer to the group average, while retaining features that are observed to be reliable at the subject level.

Figure 5 displays the degree of shrinkage towards the group mean for single-session and oracle shrinkage for three different scan durations: 400, 800 and 1200 volumes, corresponding to approximately 5, 10 and 15 minutes, respectively. For each scan duration and shrinkage type, the degree of shrinkage for each unique pair of nodes is displayed. Figure 5a shows that the degree of shrinkage tends to decrease as scan duration increases, as traditional estimates of connectivity become more reliable and less shrinkage is required. The degree of shrinkage varies dramatically across the connectivity matrix, with more reliable connections (e.g., most within-network connections) receiving relatively little shrinkage and less reliable connections (e.g., those between the visual network and other networks) receiving more shrinkage. These patterns are somewhat similar whether full or partial correlation is used as the measure of FC, with some exceptions. For example, most connections involving cerebellar or subcortical regions receive more shrinkage using partial correlation, indicating that these connections are less reliable based on partial correlation. Meanwhile, some connections within and between the DMN and cognitive control networks receive slightly less shrinkage when partial correlation is used, indicating that these connections can be more reliably measured using partial correlation, with the caveat that ρ must be chosen correctly (see Figure 2).

Below each matrix, we "zoom in" on all connections with one DMN seed region to better visualize the spatial differences in the degree of shrinkage for full correlations. The selected seed shown in black in each image. For this seed, the degree of shrinkage is highest for connections with motor, visual and auditory regions and is lowest with other default mode regions and cognitive control regions. This is consistent with previous findings that connectivity within and between the default mode and cognitive control networks tends to be highly reliable, while connectivity between the DMN and other networks is generally less reliable (Van Dijk et al., 2010; Finn et al., 2015; Laumann et al., 2015; Airan et al., 2016).

Figure 5b shows that oracle shrinkage results in substantially greater shrinkage compared with single-session shrinkage. Recall that single-session shrinkage is based on an estimate of within-subject variance from a single session, while oracle shrinkage is based on within-subject variance across multiple sessions, in this case occurring on separate days. The greater degree of shrinkage with oracle shrinkage is reflective of greater within-subject variance in FC *across* sessions than *within* a single session. This is not surprising, since inter-session differences in FC and related measures have been found to be greater than intra-session differences (Shehzad et al., 2009; Anderson et al., 2011; Birn et al., 2013; Zuo et al., 2013). Single-session shrinkage can therefore be considered a conservative form of shrinkage, since it tends to preserve more subject-level information than oracle shrinkage. However, the patterns of shrinkage within the connectivity matrix are similar for oracle and single-session shrinkage, suggesting that the proposed single-session within-subject variance estimator is an attenuated version of the true within-subject variance across multiple sessions.

Figure 6 displays the reliability of traditional, single-session shrinkage and oracle shrinkage estimates of connectivity by scan duration. The upper triangle of each matrix shows the ICC_{MSE} of full correlations; the lower triangle shows that of partial correlations. For each measure, the $I2C2_{MSE}$ of each node and the omnibus ICC_{MSE} are displayed in the margins of the upper and lower triangles. As previously observed in terms of degree of shrinkage, Figure 6a shows that reliability tends to increase with scan duration and the most reliable connections are those within and between the DMN and cognitive control networks. Connections between these networks and some visual and cerebellar nodes are also among the most reliable.

Figure 6b shows that single-session shrinkage estimates are significantly more reliable than traditional estimates of connectivity at each scan duration, with improvement observed across all parts of the connectivity matrix. Not surprisingly, the benefits of shrinkage are greatest for shorter scan duration, but improvement is still substantial for 1200-volume (approximately 15-minute) acquisitions. Specifically, based on full correlations, single-session shrinkage increases omnibus ICC_{MSE} by 17.6% for 400-volume scans, 11.3% for 800-volume scans, and 8.9% for 1200-volume scans; based on partial correlations, the improvement is slightly higher at 22.1%, 13.6% and 10.4%, respectively. Notably, single-session shrinkage estimates based on *5-minute* (400-volume) acquisitions are nearly as reliable as traditional estimates based on *15-minute* (1200-volume) acquisitions.

Figure 6c displays the reliability of oracle shrinkage estimates. Oracle shrinkage provides an upper bound on the benefit of shrinkage, since it uses the same repeat measurements used to assess reliability to determine the optimal degree of shrinkage. While oracle shrinkage estimates are more reliable than single-session shrinkage estimates, the improvement is less dramatic than that of single-session shrinkage estimates over traditional estimates. For example, based on full correlations oracle shrinkage increases omnibus ICC_{MSE} over single-session shrinkage by an additional 4.2% with 400 volumes and 6.5% with 1200 volumes. However, as seen in Figure 5, on average oracle shrinkage requires nearly twice the amount of shrinkage toward the group mean; single-session shrinkage therefore achieves much of

the possible improvements in reliability, while preserving more subject-level information than oracle shrinkage.

Figure 7 displays the average reliability of all within-network connections for the seven resting-state networks shown in Figure 1. For traditional and shrinkage estimates alike, connections within the default mode and cognitive control networks are by far the most reliable in terms of ICC_{MSE}, while connections within the visual, sensorimotor, auditory, cerebellar and subcortical networks tend to be less reliable. Due to the choice of $\rho = 5$ (see Figure 2), partial correlations are more reliable than full correlations for connections within the DMN and cognitive control networks and are worse than or comparable to full correlations for all other networks.

For both full and partial correlations, single-session shrinkage estimates of within-network FC are sub-stantially more reliable than traditional estimates across all scan durations and networks. The improvement is most dramatic for shorter durations, with an increase in ICC_{MSE} of 24.0 – 50.9% across the different networks using full correlation and 24.3 – 47.7% using partial correlation, based on 100-volume acquisitions. Based on 1200-volume acquisitions, the improvement is 5.1 – 9.0% using full correlation and 4.7 – 12.6% using partial correlation. Note that even the most reliable connections based on the longest scan durations become more reliable with shrinkage. For example, based on 1200-volume acquisitions, the ICC_{MSE} of partial correlations within the DMN increases from 0.649 to 0.680 (4.7%) due to shrinkage, and within the CC network from 0.621 to 0.654 (5.3%). This suggests that reliability of certain connections can be maximized by simultaneously increasing scan duration, employing partial correlations—if ρ can be chosen optimally—*and* performing shrinkage on the resulting estimates.

4 Discussion

In this paper, we propose an empirical Bayes shrinkage method for resting-state functional connectivity (FC) using single-session fMRI data. The proposed method is rooted in a novel measurement error model for FC that takes into account between-subject differences in FC, sampling error in estimation of FC, and dynamic changes in FC over time. In contrast with previously proposed Bayesian or shrinkage methods for FC, this method does not require access to multiple fMRI sessions. To compare the reliability of shrinkage and traditional estimates in a scale-free way, we also propose a novel reliability measure, ICC_{MSE}, applicable to both biased and unbiased estimators. The ICC_{MSE} reduces to the intra-class correlation coefficient (ICC) in the absence of bias. Similar to ICC, it ranges from 0 to 1, and is a measure of how much reliable subject-level information is contained in a set of estimates, relative to the variance across subjects. As with ICC, ICC_{MSE} can also be used to compare reliability of estimates on different scales, such as full and partial correlations. Finally, we illustrate a multi-resolution approach to evaluate reliability of FC at the level of individual connections, seed maps, networks, and the entire entire connectivity matrix. This approach can be used to assess the effects of shrinkage, scan duration, connectivity measure and other factors on overall reliability of FC *and* how these effects vary across different networks, regions and connections.

We test the proposed methods on 461 subjects from the Human Connectome Project, using both full and partial correlation as the measure of FC. We find that single-session shrinkage improves overall reliability of FC estimates by approximately 18 – 22% for 5-minute (400-volume) acquisitions and 9 – 10% for 15-minute (1200-volume) acquisitions. As consistently observed in the literature, we find that reliability varies substantially across connections, with the most reliable connections being those within and between the default mode and cognitive control networks, and between those networks and certain visual and cerebellar regions. Notably, we find that shrinkage substantially improves even the most reliable connections based on the longest acquisitions.

Both full and partial correlations benefit from shrinkage to similar degrees. However, we find that the choice of ρ is critical for partial correlations, and that partial correlations are generally less reliable overall than full correlations, particularly for small values of ρ. This is consistent with the findings of Abraham et al. (2017), whose analysis was based on the Autism Brain Imaging Data Exchange (ABIDE) dataset (Di Martino et al., 2014); compared to the HCP, the resting-state fMRI data of ABIDE are of shorter duration, have longer TR, and are based on less sophisticated acquisition and processing techniques. Therefore, our results serve to somewhat generalize these previous findings. However, we also find that certain connections actually become more reliable using partial correlations for specific values of ρ, as illustrated in Figure 2b. This likely reflects the property that partial correlations are used to estimate a relative few “direct” connections, while shrinking the remaining “indirect” connections towards zero. These indirect connections may be meaningfully estimated when using full correlation but are essentially equal to zero for all subjects using partial correlations, hence resulting in worse reliability as measured by ICC. This illustrates the importance of investigating reliability at multiple resolutions and focusing on connections of interest when assessing the value of various processing and analysis choices or determining optimal parameter values.

If “direct” connections are of primary interest, the use of partial correlation combined with a principled method for choosing ρ may thus be beneficial. While we find $\rho = 5$ to maximize the reliability of certain connections, the optimal value likely depends on many factors specific to the dataset being analyzed, along with the particular connections of interest. Therefore, ρ should be chosen for each analysis in a principled manner. For example, ρ can be chosen to maximize reliability using a measure (e.g., ICC) that (1) is scale-free and therefore robust to changes in scale of partial correlations induced by changes in ρ, and (2) reflects within-subject reliability *relative* to differences between subjects. We did not find that scan duration strongly influenced the choice of ρ to maximize overall reliability, so it seems reasonable to use intra-session reliability as a proxy for inter-session reliability in the absence of retest data. Overall, our results suggest that future research is needed to develop principled methods for estimating partial correlations in order to fully realize their benefits.

We also perform “oracle” shrinkage to provide an upper limit on the inter-session reliability that can be achieved with shrinkage. We find that oracle shrinkage results in approximately double the degree of shrinkage towards the group average versus single-session shrinkage, as seen in Figure 5. This is due to higher inter-session variance of FC, the basis of oracle shrinkage, compared with intra-session variance, the basis of single-session shrinkage.

Single-session shrinkage is therefore a conservative form of shrinkage for FC, as it places greater weight on the original subject-level estimates compared with oracle shrinkage; yet it captures most of the possible improvement in inter-session reliability, suggesting that the proposed single-session shrinkage methods offer significant benefits while preserving subject-level information.

While in this paper we focus on static connectivity, the proposed methods are directly applicable to shrinkage of dynamic connectivity states shared across subjects. For each state, the proposed model is simply applied to the time series for each subject corresponding to that state (these are allowed to differ in duration across subjects, with subject-specific variance estimators given in Appendix A). Performing shrinkage within each state is indeed likely beneficial, since within each state variation in true connectivity is reduced, and therefore shrinkage will be driven more by random noise than dynamic fluctuations, which are biologically meaningful and potentially of interest. In the complete absence of dynamic fluctuations, within-subject variance of FC would reflect primarily differences in SNR across the brain as well as motion and physiological confounds. These latter sources of variation, if measured, can also be removed from the data before performing shrinkage, which would serve to reduce the total noise and lead to less shrinkage of subject-level estimates of FC. Indeed, any processing steps that serve to reduce noise at the subject level are highly advisable prior to shrinkage in order to maximize the SNR in the estimates and therefore minimize shrinkage. In this way, shrinkage can also be used as a way to assess the benefit of various processing and analysis choices in the absence of retest data, with choices leading to minimal shrinkage being most beneficial.

Here we have focused on reliability across multiple visits, but it is important to consider the measure or effect of interest when choosing how to assess reliability. While reliability across multiple visits is appropriate for detecting trait-level effects, which are stable over long periods of time, reliability within a single visit is more appropriate for state-level effects. Functional connectivity is known to vary substantially across visits, so single-session estimates of FC will tend to achieve only modest inter-session reliability, limiting their utility for studying trait-level effects. Several recent studies have suggested that it may be preferable to combine data from multiple sessions occurring on different days, rather than simply increase scan duration within a single session (Shehzad et al., 2009; Laumann et al., 2015; Noble et al., 2017b).

In this study we assess the reliability of FC estimates produced from scans up to approximately 15 minutes in duration. Even within this relatively short scan duration, the results shown in Figure 7 suggest that increasing scan duration, while beneficial, offers diminishing returns in terms of improving inter-session reliability. Reliability of FC has been shown to depend on many factors, including the connection(s) of interest, temporal and spatial resolution, node definition, and the measure of functional connectivity, as well as how reliability is defined (Zuo et al., 2013; Murphy et al., 2007; Anderson et al., 2011; Smith et al., 2011; Dawson et al., 2013; Birn et al., 2013; Zuo et al., 2013; Liao et al., 2013; Airan et al., 2016). In light of this, it is not surprising that recommendations regarding how long to scan have differed widely, ranging from 7 minutes or less (Tomasi et al., 2016; Airan et al., 2016) to 90 minutes or more (Laumann et al., 2015). Our findings contribute to a

growing body of work that suggests that maximizing reliability requires not only increasing the duration or number of scans, but adopting other best practices in processing and analysis, such as borrowing strength across subjects through shrinkage.

Our study has several limitations. Our results are based on a single dataset, parcellation, and subject group; while we believe that the findings are generalizable, future work should focus on determining the effect of shrinkage in different contexts. Additionally, while we perform connection-specific shrinkage, we apply the same degree of shrinkage to each subject. It is plausible that reliability of FC differs based on subject characteristics, and therefore differential shrinkage across subjects would be more beneficial. While tailoring the degree of shrinkage for different subjects is possible, doing so increases the number of parameters that must be estimated and may therefore actually worsen the performance of shrinkage estimators (Mejia et al., 2015). This should be explored as an area of future research.

Acknowledgments

Funding

Data were provided by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University. This research was supported in part by NIH grants R01 EB016061, R01 EB012547, and P41 EB015909 from the National Institute of Biomedical Imaging and Bioengineering, R01 MH095836 and K01 MH109766 from the National Institute of Mental Health, and the Craig H. Neilsen Foundation (Project Number 338419).

A Differing scan duration

In some studies, subjects may have varying scan duration, due to differences in protocol, varying ability of subjects to undergo scanning, or scrubbing (removal of volumes contaminated with artifacts) during preprocessing. Let $\mathcal{T}_i = \{1, \ldots, T_i\}$ index the fMRI time series for subject *i*. Then, the model in equation (1) becomes

$$w_{i,\mathcal{T}}(q,q') = \mu_i(q,q') + \delta_{i,\mathcal{T}_i}(q,q') + \varepsilon_{i,\mathcal{T}_i}(q,q'),$$

and $\delta_{i,\mathcal{T}_i}(q,q')$ and $\varepsilon_{i,\mathcal{T}_i}(q,q')$ are approximately distributed

$$\delta_{i,\mathcal{T}_i}(q,q') = \frac{1}{T_i}\sum_{t=1}^{T_i}\delta_{i,t}(q,q') \sim N\left\{0, T_{\delta,i}^{-1}\sigma_\delta^2(q,q')\right\} \text{and}$$
$$\varepsilon_{i,\mathcal{T}_i}(q,q') = \frac{1}{T_i}\sum_{t=1}^{T_i}\varepsilon_{i,t}(q,q') \sim N\left\{0, T_{\varepsilon,i}^{-1}\sigma_\varepsilon^2(q,q')\right\},$$

where $T_{\delta,i} = T_i/\tau_\delta$ and $T_{\varepsilon,i} = T_i/\tau_\varepsilon$ are the the effective sample size (ESS) of $\{\delta_{i,t}(q,q')\}_{t\in\mathcal{T}_i}$ and $\{\varepsilon_{i,t}(q,q')\}_{t\in\mathcal{T}_i}$, respectively.

Then, the variance estimation procedures described in Section 2.1.2 can be modified as follows. First, note that the between-subject variance of the long-term average FC for each subject, $\sigma^2_{\mu}(q,q')$, is unchanged, but the total variance of the estimates $w_{i,\mathcal{T}_i}(q,q')$ will vary across subjects due to differences in within-subject variance. Specifically, the within-subject variance for each subject is

$$\begin{aligned}\sigma^2_{within,i} &= T^{-1}_{\delta,i}\sigma^2_{\delta}(q,q') + T^{-1}_{\varepsilon,i}\sigma^2_{\varepsilon}(q,q') \\ &= \tau_{\delta}T_i^{-1}\sigma^2_{\delta}(q,q') + \tau_{\varepsilon}T_i^{-1}\sigma^2_{\varepsilon}(q,q') \\ &= T_i^{-1}\left\{\tau_{\delta}\sigma^2_{\delta}(q,q') + \tau_{\varepsilon}\sigma^2_{\varepsilon}(q,q')\right\} \\ &= :T_i^{-1}c(q,q')\end{aligned}$$

and the common term $c(q, q)$ can be estimated as follows. Assuming without loss of generality that T is even, define the two sub-time series $\mathcal{T}_{i,1} = \left\{1,\ldots,\frac{T_i}{2}\right\}$ and $\mathcal{T}_{i,2} = \left\{\frac{T_i}{2}+1, T_i\right\}$. For $j = 1, 2$, let $w_{i,\mathcal{T}_{i,j}}(q,q')$ be the estimate of static connectivity within sub-time series $\mathcal{T}_j$ for subject i. As before, we can write $w_{i,\mathcal{T}_{i,j}}(q,q') = \mu_i(q,q') + \delta_{i,\mathcal{T}_{i,j}}(q,q') + \varepsilon_{i,\mathcal{T}_{i,j}}(q,q')$ for $j = 1, 2$, where the ESS of $\left\{\delta_{i,t}(q,q')\right\}_{t\in\mathcal{T}_{i,j}}$ is $(T_i/2)/\tau_{\delta}$ and the ESS of $\varepsilon_{i,\mathcal{T}_{i,j}}(q,q')$ is $(T_i/2)/\tau_{\varepsilon}$. Therefore, $\delta_{i,\mathcal{T}_{i,j}}(q,q') \sim N\left\{0, 2\tau_{\delta}T_i^{-1}\sigma^2_{\delta}(q,q')\right\}$ and $\varepsilon_{i,\mathcal{T}_{i,j}}(q,q') \sim N\left\{0, 2\tau_{\varepsilon}T_i^{-1}\sigma^2_{\varepsilon}(q,q')\right\}$ for $j = 1, 2$.

Taking the difference $w_{i,\mathcal{T}_{i,1}}(q,q') - w_{i,\mathcal{T}_{i,2}}(q,q')$ for each subject, we can compute

$$\begin{aligned}Var_i\left\{w_{i,\mathcal{T}_{i,1}}(q,q') - w_{i,\mathcal{T}_{i,2}}(q,q')\right\} &= Var_i\left\{\delta_{i,\mathcal{T}_{i,1}}(q,q') + \varepsilon_{i,\mathcal{T}_{i,1}}(q,q') - \delta_{i,\mathcal{T}_{i,2}}(q,q') - \varepsilon_{i,\mathcal{T}_{i,2}}(q,q')\right\} \\ &= 2Var_i\left\{\delta_{i,\mathcal{T}_{i,1}}(q,q')\right\} + 2Var_i\left\{\varepsilon_{i,\mathcal{T}_{i,1}}(q,q')\right\} \\ &= 2E\left[\delta^2_{i,\mathcal{T}_{i,1}}(q,q')\right] + 2E\left[\varepsilon^2_{i,\mathcal{T}_{i,1}}(q,q')\right] \\ &= 2E_{T_i}E\left[\delta^2_{i,\mathcal{T}_{i,1}}(q,q')\middle|T_i\right] + 2E_{T_i}E\left[\varepsilon^2_{i,\mathcal{T}_{i,1}}(q,q')\middle|T_i\right] \\ &= 2E\left[2\tau_{\delta}T_i^{-1}\sigma^2_{\delta}(q,q')\right] + 2E\left[2\tau_{\varepsilon}T_i^{-1}\sigma^2_{\varepsilon}(q,q')\right] \\ &= 4\left(\tau_{\delta}\sigma^2_{\delta}(q,q') + \tau_{\varepsilon}\sigma^2_{\varepsilon}(q,q')\right)E\left[T_i^{-1}\right] \\ &= 4c(q,q')E\left[T_i^{-1}\right]\end{aligned}$$

Therefore, $c(q, q')$ can be estimated as $\hat{c}(q,q') = \frac{1}{4}\left(\frac{1}{n}\sum_{i=1}^{n}T_i^{-1}\right)^{-1}Var_i\left\{w_{i,\mathcal{T}_{i,1}}(q,q') - w_{i,\mathcal{T}_{i,2}}(q,q')\right\}$, and the within-subject variance for subject i is estimated as $\hat{\sigma}^2_{within,i} = T_i^{-1}\hat{c}(q,q')$.

The between-subject variance $\sigma^2_\mu(q,q')$ can be easily estimated as follows. Writing the total variance, which can be estimated directly from the data, as $Var_i\left\{w_{i,\mathcal{T}_i}(q,q')\right\} = \sigma^2_\mu(q,q') + Var_i\left\{\delta_{i,\mathcal{T}_i}(q,q')\right\} + Var_i\left\{\varepsilon_{i,\mathcal{T}_i}(q,q')\right\}$, the between-subject variance estimate is $\hat{\sigma}^2_\mu(q,q') = Var_i\left\{w_{i,\mathcal{T}_i}(q,q')\right\} - \frac{1}{2}Var_i\left\{w_{i,\mathcal{T}_{i,1}}(q,q') - w_{i,\mathcal{T}_{i,2}}(q,q')\right\}$.

The degree of shrinkage for subject i is given by $\lambda_i(q,q') = \hat{\sigma}^2_{within,i}(q,q')/\left(\hat{\sigma}^2_{within,i}(q,q') + \hat{\sigma}^2_\mu(q,q')\right)$. Subjects with longer scans will receive less shrinkage for a given connection than subjects with shorter scans, since their estimates are based on more data and are therefore more reliable.

References

Abraham A, Milham MP, Di Martino A, Craddock RC, Samaras D, Thirion B, Varoquaux G. Deriving reproducible biomarkers from multi-site resting-state data: An autism-based example. NeuroImage. 2017; 147:736–745. [PubMed: 27865923]

Airan RD, Vogelstein JT, Pillai JJ, Caffo B, Pekar JJ, Sair HI. Factors affecting characterization and localization of interindividual differences in functional connectivity using MRI. Human brain mapping. 2016; 37(5):1986–1997. [PubMed: 27012314]

Allen EA, Damaraju E, Plis SM, Erhardt EB, Eichele T, Calhoun VD. Tracking whole-brain connectivity dynamics in the resting state. Cerebral cortex. 2014; 24(3):663–676. [PubMed: 23146964]

Anderson JS, Ferguson MA, Lopez-Larson M, Yurgelun-Todd D. Reproducibility of single-subject functional connectivity measurements. American journal of neuroradiology. 2011; 32(3):548–555. [PubMed: 21273356]

Beckmann CF, Mackay CE, Filippini N, Smith SM. Group comparison of resting-state fMRI data using multi-subject ICA and dual regression. NeuroImage. 2009; 47(Suppl 1):S148.

Beckmann CF, Smith SM. Probabilistic independent component analysis for functional magnetic resonance imaging. Medical Imaging, IEEE Transactions on. 2004; 23(2):137–152.

Birn RM, Molloy EK, Patriat R, Parker T, Meier TB, Kirk GR, Nair VA, Meyerand ME, Prabhakaran V. The effect of scan length on the reliability of resting-state fMRI connectivity estimates. Neuroimage. 2013; 83:550–558. [PubMed: 23747458]

Button KS, Ioannidis JP, Mokrysz C, Nosek BA, Flint J, Robinson ES, Munafò MR. Power failure: why small sample size undermines the reliability of neuroscience. Nature Reviews Neuroscience. 2013; 14(5):365–376. [PubMed: 23571845]

Choe AS, Nebel MB, Barber AD, Cohen JR, Xu Y, Pekar JJ, Caffo B, Lindquist MA. Comparing test-retest reliability of dynamic functional connectivity methods. NeuroImage. 2017; 158:155–175. [PubMed: 28687517]

Chong M, Bhushan C, Joshi A, Choi S, Haldar J, Shattuck D, Spreng R, Leahy R. Individual parcellation of resting fMRI with a group functional connectivity prior. NeuroImage. 2017; 156:87–100. [PubMed: 28478226]

Collaboration OS. Estimating the reproducibility of psychological science. Science. 2015; 349(6251)

Dai T, Guo Y, Initiative ADN, et al. Predicting individual brain functional connectivity using a Bayesian hierarchical model. NeuroImage. 2016

Dawson DA, Cha K, Lewis LB, Mendola JD, Shmuel A. Evaluation and calibration of functional network modeling methods based on known anatomical connections. Neuroimage. 2013; 67:331–343. [PubMed: 23153969]

Di Martino A, Yan CG, Li Q, Denio E, Castellanos FX, Alaerts K, Anderson JS, Assaf M, Bookheimer SY, Dapretto M, et al. The autism brain imaging data exchange: towards large-scale evaluation of

the intrinsic brain architecture in autism. Molecular psychiatry. 2014; 19(6):659. [PubMed: 23774715]

Efron B, Morris C. Data analysis using Stein's estimator and its generalizations. Journal of the American, Statistical Association. 1975; 70(350):311–319.

Feinberg DA, Moeller S, Smith SM, Auerbach E, Ramanna S, Gunther M, Glasser MF, Miller KL, Ugurbil K, Yacoub E. Multiplexed echo planar imaging for sub-second whole brain fMRI and fast diffusion imaging. PloS one. 2010; 5(12):e15710. [PubMed: 21187930]

Finn ES, Shen X, Scheinost D, Rosenberg MD, Huang J, Chun MM, Papademetris X, Constable RT. Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. Nature neuroscience. 2015

Glasser MF, Sotiropoulos SN, Wilson JA, Coalson TS, Fischl B, Andersson JL, Xu J, Jbabdi S, Webster M, Polimeni JR, et al. The minimal preprocessing pipelines for the human connectome project. Neuroimage. 2013; 80:105–124. [PubMed: 23668970]

Golub GH, Heath M, Wahba G. Generalized cross-validation as a method for choosing a good ridge parameter. Technometrics. 1979; 21(2):215–223.

Gong L, Flegal JM. A practical sequential stopping rule for high-dimensional Markov chain Monte Carlo. Journal of Computational and Graphical Statistics. 2016; 25(3):684–700.

Gordon EM, Laumann TO, Gilmore AW, Newbold DJ, Greene DJ, Berg JJ, Ortega M, Hoyt-Drazen C, Gratton C, Sun H, et al. Precision functional mapping of individual human brains. Neuron. 2017; 95(4):791–807. [PubMed: 28757305]

Griffanti L, Salimi-Khorshidi G, Beckmann CF, Auerbach EJ, Douaud C, Sexton CE, Zsoldos E, Ebmeier KP, Filippini N, Mackay CE, et al. ICA-based artefact removal and accelerated fMRI acquisition for improved resting state network imaging. Neuroimage. 2014; 95:232–247. [PubMed: 24657355]

Ha MJ, Sun W. Partial correlation matrix estimation using ridge penalty followed by thresholding and re-estimation. Biometrics. 2014; 70(3):762–770.

Hansen PC. Analysis of discrete ill-posed problems by means of the L-curve. SIAM review. 1992; 34(4):561–580.

James W, Stein C. Estimation with quadratic loss. Proceedings of the fourth Berkeley symposium on mathematical statistics and probability. 1961; 1:361–379.

Kass RE, Carlin BP, Gelman A, Neal RM. Markov chain Monte Carlo in practice: a roundtable discussion. The American Statistician. 1998; 52(2):93–100.

Laumann TO, Cordon EM, Adeyemo B, Snyder AZ, Joo SJ, Chen MY, Cilmore AW, McDermott KB, Nelson SM, Dosenbach NU, et al. Functional system and areal organization of a highly sampled individual human brain. Neuron. 2015; 87(3):657–670. [PubMed: 26212711]

Liao XH, Xia MR, Xu T, Dai ZJ, Cao XY, Niu HJ, Zuo XN, Zang YF, He Y. Functional brain hubs and their test-retest reliability: a multiband resting-state functional MRI study. Neuroimage. 2013; 83:969–982. [PubMed: 23899725]

Mejia AF, Nebel MB, Shou H, Crainiceanu CM, Pekar JJ, Mostofsky S, Caffo B, Lindquist MA. Improving reliability of subject-level resting-state fMRI parcellation with shrinkage estimators. NeuroImage. 2015; 112:14–29. [PubMed: 25731998]

Moeller S, Yacoub E, Olman CA, Auerbach E, Strupp J, Harel N, Ugurbil K. Multiband multislice ge-epi at 7 tesla, with 16-fold acceleration using partial parallel imaging with application to high spatial and temporal whole-brain fMRI. Magnetic Resonance in Medicine. 2010; 63(5):1144–1153. [PubMed: 20432285]

Morozov, V. Methods for solving incorrectly posed problems. Springer; 1984. Regular methods for solving linear and nonlinear ill-posed problems; p. 65-122.

Morris CN. Parametric empirical Bayes inference: theory and applications. Journal of the American Statistical Association. 1983; 78(381):47–55.

Mueller S, Wang D, Fox MD, Pan R, Lu J, Li K, Sun W, Buckner RL, Liu H. Reliability correction for functional connectivity: Theory and implementation. Human Brain Mapping. 2015

Munafò M, Noble S, Browne WJ, Brunner D, Button K, Ferreira J, Holmans P, Langbehn D, Lewis G, Lindquist M, et al. Scientific rigor and the art of motorcycle maintenance. Nature biotechnology. 2014; 32(9):871–873.

Murphy K, Bodurka J, Bandettini PA. How long to scan? The relationship between fMRI temporal signal to noise ratio and necessary scan duration. Neuroimage. 2007; 34(2):565–574. [PubMed: 17126038]

Noble S, Scheinost D, Finn ES, Shen X, Papademetris X, McEwen SC, Bearden CE, Addington J, Goodyear B, Cadenhead KS, et al. Multisite reliability of mr-based functional connectivity. Neuroimage. 2017a; 146:959–970. [PubMed: 27746386]

Noble S, Spann MN, Tokoglu F, Shen X, Constable RT, Scheinost D. Influences on the test-retest reliability of functional connectivity MRI and its relationship with behavioral utility. Cerebral Cortex. 2017b:1–15.

Rahim, M., Thirion, B., Varoquaux, G. International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer; 2017. Population-shrinkage of covariance to estimate better brain functional connectivity; p. 460-468.

Salimi-Khorshidi G, Douaud G, Beckmann CF, Glasser MF, Griffanti L, Smith SM. Automatic denoising of functional MRI data: combining independent component analysis and hierarchical fusion of classifiers. Neuroimage. 2014; 90:449–468. [PubMed: 24389422]

Setsompop K, Gagoski BA, Polimeni JR, Witzel T, Wedeen VJ, Wald LL. Blipped-controlled aliasing in parallel imaging for simultaneous multislice echo planar imaging with reduced g-factor penalty. Magnetic Resonance in Medicine. 2012; 67(5):1210–1224. [PubMed: 21858868]

Shehzad Z, Kelly AC, Reiss PT, Gee DG, Gotimer K, Uddin LQ, Lee SH, Margulies DS, Roy AK, Biswal BB, et al. The resting brain: unconstrained yet reliable. Cerebral cortex. 2009; 19(10): 2209–2229. [PubMed: 19221144]

Shou H, Eloyan A, Lee S, Zipunnikov V, Crainiceanu A, Nebel M, Caffo B, Lindquist M, Crainiceanu C. Quantifying the reliability of image replication studies: the image intraclass correlation coefficient (I2C2). Cognitive, Affective, & Behavioral Neuroscience. 2013; 13(4):714–724.

Shou H, Eloyan A, Nebel MB, Mejia A, Pekar JJ, Mostofsky S, Caffo B, Lindquist MA, Crainiceanu CM. Shrinkage prediction of seed-voxel brain connectivity using resting state fMRI. NeuroImage. 2014; 102:938–944. [PubMed: 24879924]

Smith SM, Miller KL, Salimi-Khorshidi G, Webster M, Beckmann CF, Nichols TE, Ramsey JD, Woolrich MW. Network modelling methods for fMRI. Neuroimage. 2011; 54(2):875–891. [PubMed: 20817103]

Smith SM, Nichols TE, Vidaurre D, Winkler AM, Behrens TE, Glasser MF, Ugurbil K, Barch DM, Van Essen DC, Miller KL. A positive-negative mode of population covariation links brain connectivity, demographics and behavior. Nature neuroscience. 2015; 18(11):1565–1567. [PubMed: 26414616]

Su SC, Caffo B, Garrett-Mayer E, Bassett SS. Modified test statistics by inter-voxel variance shrinkage with an application to fMRI. Biostatistics. 2008; 10(2):219–227. [PubMed: 18723853]

Thompson, MB. PhD thesis. 2011. Slice sampling with multivariate steps.

Tie Y, Rigolo L, Norton IH, Huang RY, Wu W, Orringer D, Mukundan S, Golby AJ. Defining language networks from resting-state fMRI for surgical planning—a feasibility study. Human brain mapping. 2014; 35(3):1018–1030. [PubMed: 23288627]

Tomasi DG, Shokri-Kojori E, Volkow ND. Temporal evolution of brain functional connectivity metrics: Could 7 min of rest be enough? Cerebral Cortex. 2016:1–13. [PubMed: 25139941]

Uğurbil K, Xu J, Auerbach EJ, Moeller S, Vu AT, Duarte-Carvajalino JM, Lenglet C, Wu X, Schmitter S, Van de Moortele PF, et al. Pushing spatial and temporal resolution for functional and diffusion MRI in the human connectome project. Neuroimage. 2013; 80:80–104. [PubMed: 23702417]

Van Dijk KR, Hedden T, Venkataraman A, Evans KC, Lazar SW, Buckner RL. Intrinsic functional connectivity as a tool for human connectomics: theory, properties, and optimization. Journal of neurophysiology. 2010; 103(1):297–321. [PubMed: 19889849]

Van Essen DC, Smith SM, Barch DM, Behrens TE, Yacoub E, Ugurbil K, WU-Minn HCP Consortium. et al. The WU-Minn human connectome project: an overview. Neuroimage. 2013; 80:62–79. [PubMed: 23684880]

Varoquaux G, Craddock RC. Learning and comparing functional connectomes across subjects. Neuroimage. 2013; 80:405–415. [PubMed: 23583357]

Varoquaux G, Gramfort A, Poline JB, Thirion B. Brain covariance selection: better individual functional connectivity models using population prior. Advances in neural information processing systems. 2010:2334–2342.

Wang Y, Kang J, Kemmer PB, Guo Y. An efficient and reliable statistical method for estimating functional connectivity in large scale brain networks using partial correlation. Frontiers in neuroscience. 2016; 10

Warnick R, Guindani M, Erhardt E, Allen E, Calhoun V, Vannucci M. A bayesian approach for estimating dynamic functional network connectivity in fmri data. Journal of the American Statistical Association. 2017 just-accepted.

Xu J, Moeller S, Strupp J, Auerbach E, Chen L, Feinberg D, Ugurbil K, Yacoub E. Highly accelerated whole brain imaging using aligned-blipped-controlled-aliasing multiband epi. Proceedings of the 20th Annual Meeting of ISMRM. 2012; 2306

Zuo XN, Xu T, Jiang L, Yang Z, Cao XY, He Y, Zang YF, Castellanos FX, Milham MP. Toward reliable characterization of functional homogeneity in the human brain: preprocessing, scan duration, imaging resolution and computational space. Neuroimage. 2013; 65:374–386. [PubMed: 23085497]

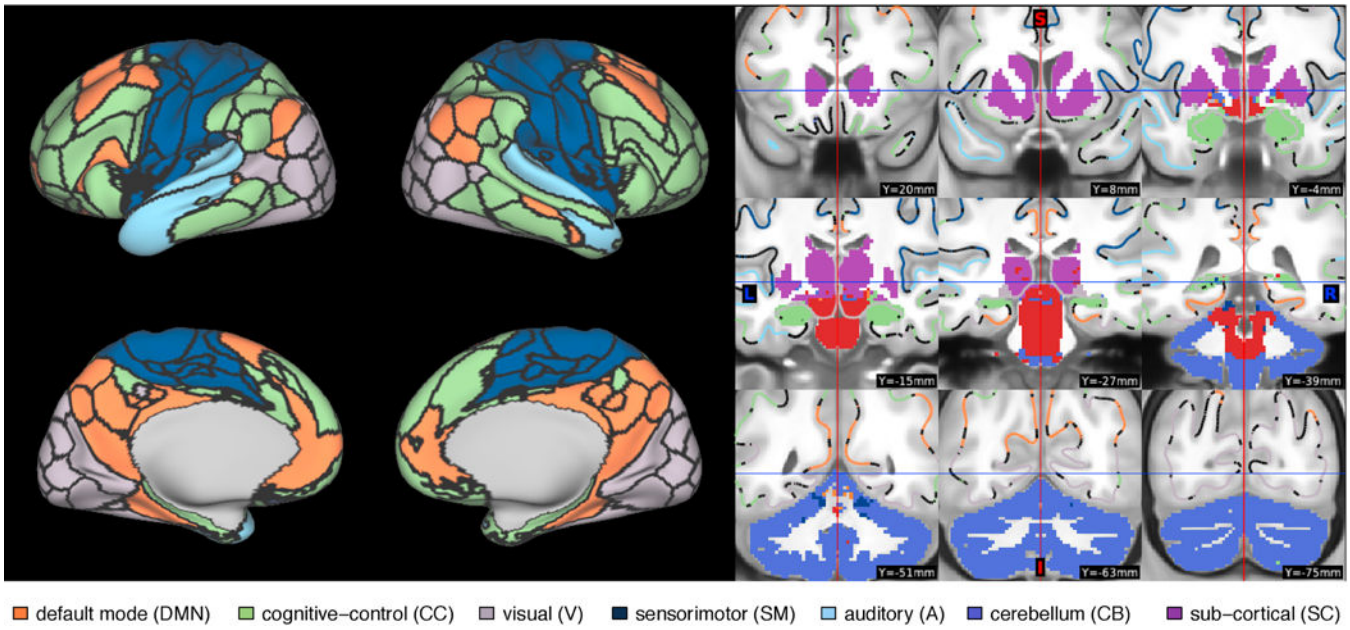
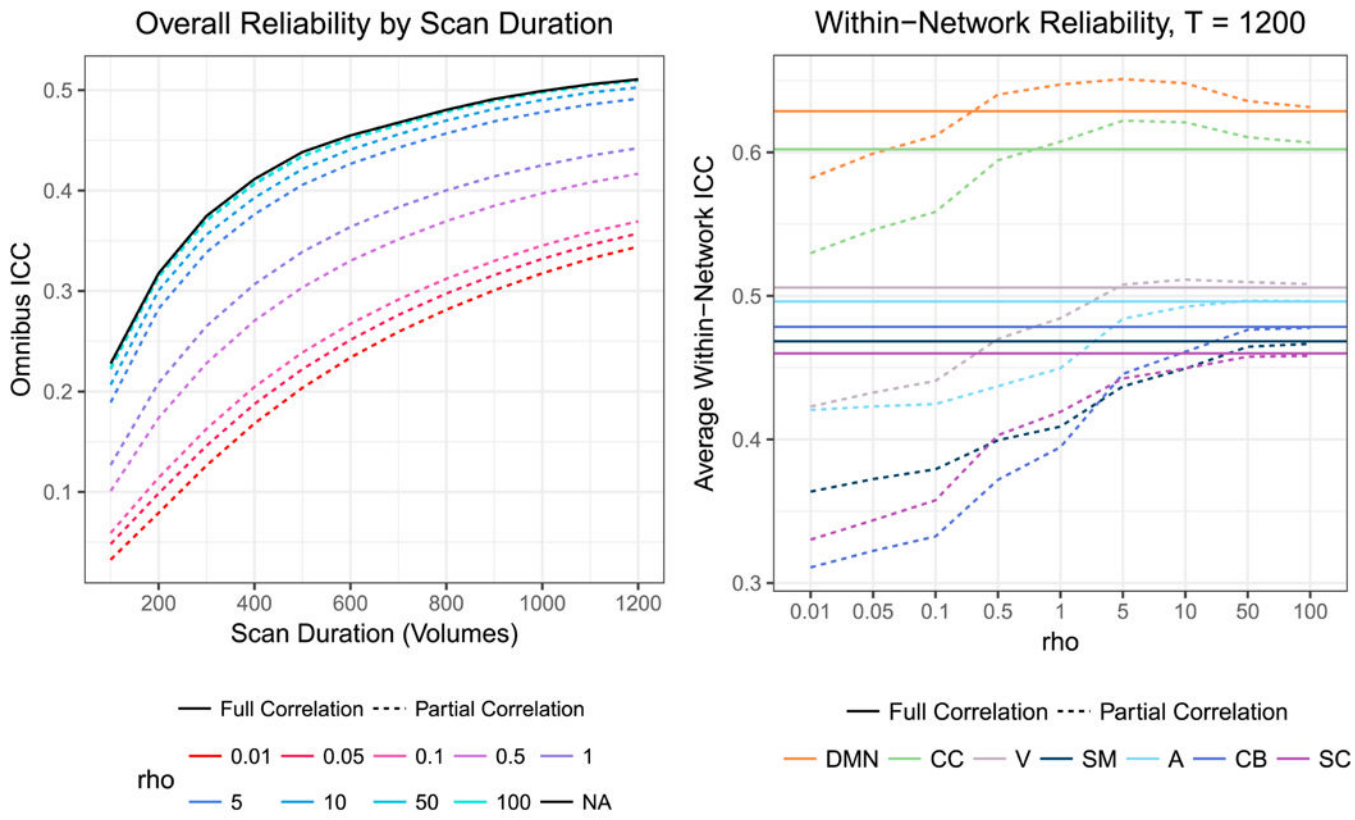


Figure 1.

Resting-state network assignment of each IC based on overlap with the Allen parcellation. Cortical boundaries delineate the borders between ICs, based on a discretization of the continuous spatial ICA maps, used for visualization purposes. ICs labeled as noise are displayed in red.



(a) Omnibus ICC of partial and full correlations.

(b) ICC of within-network connections.

Figure 2.

(a) As expected, reliability increases with scan duration for each measure. More surprisingly, we observe that the overall reliability of partial correlation estimates is much lower than that of full correlation estimates for small values of ρ, while large values of ρ (50 and 100, which are nearly indistinguishable in the plot) result in partial correlation estimates with comparable overall reliability to full correlation. Reliability begins to worsen slightly when ρ is increased from 50 to 100, suggesting that further increases in ρ would not result in improved reliability of partial correlations. (b) For connections within the DMN and CC networks, partial correlations are more reliable than full correlations for ρ greater than 0.5, and the maximal reliability is seen for partial correlations with $\rho = 5$. For all other networks, partial correlations are generally less reliable than full correlations for smaller values of ρ and achieve similar reliability for larger values of ρ.

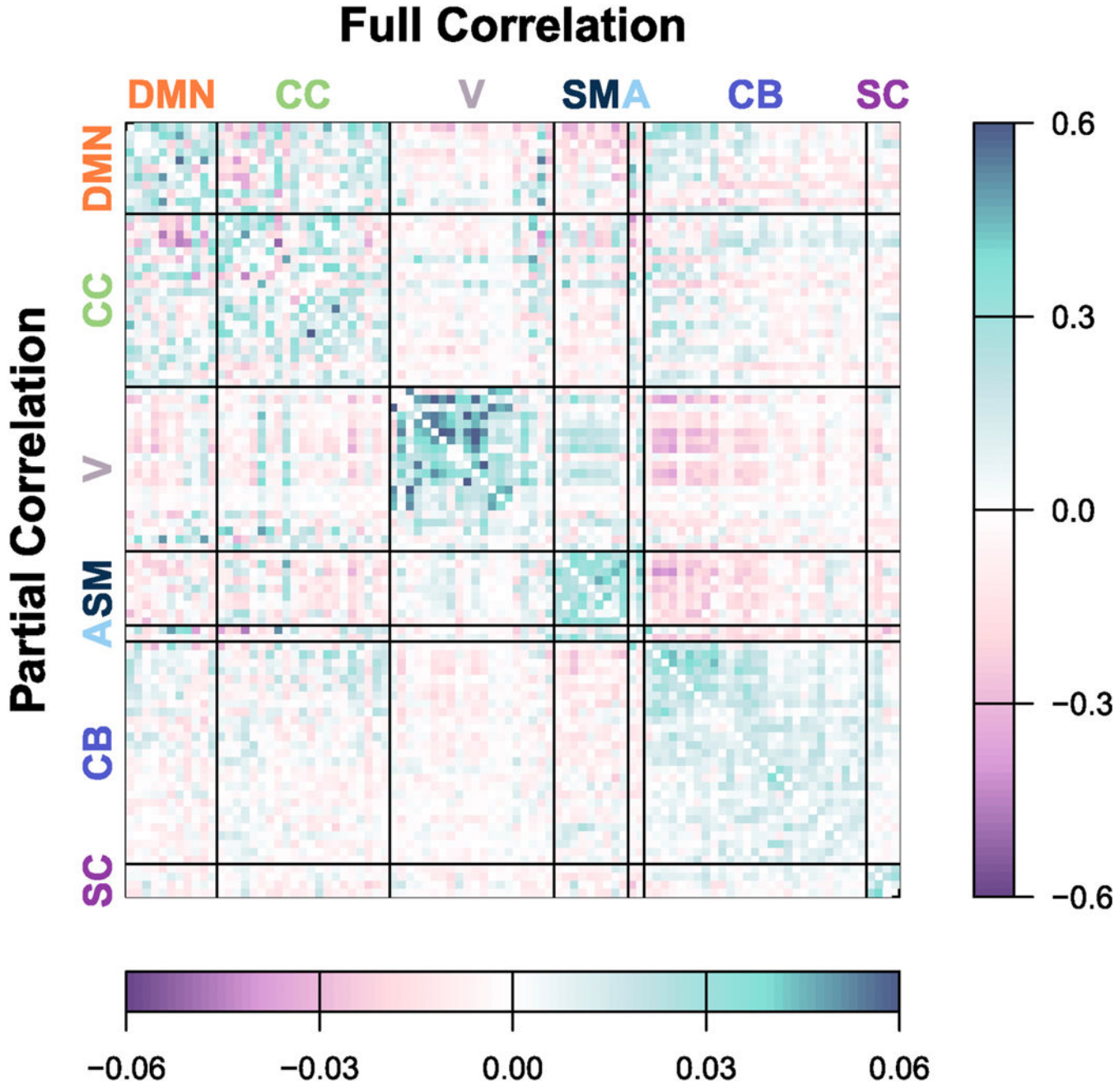


Figure 3. Group-average FC

Average FC across all subjects, based on traditional estimates using all 1200 volumes of visit 1. Full correlations are displayed in the upper triangle; partial correlations are displayed in the lower triangle. Subject-level estimates based on the same scan duration are shrunk towards this group-level average to obtain more reliable estimates, where the degree of shrinkage for each connection depends on the within-subject and between-subject variance.

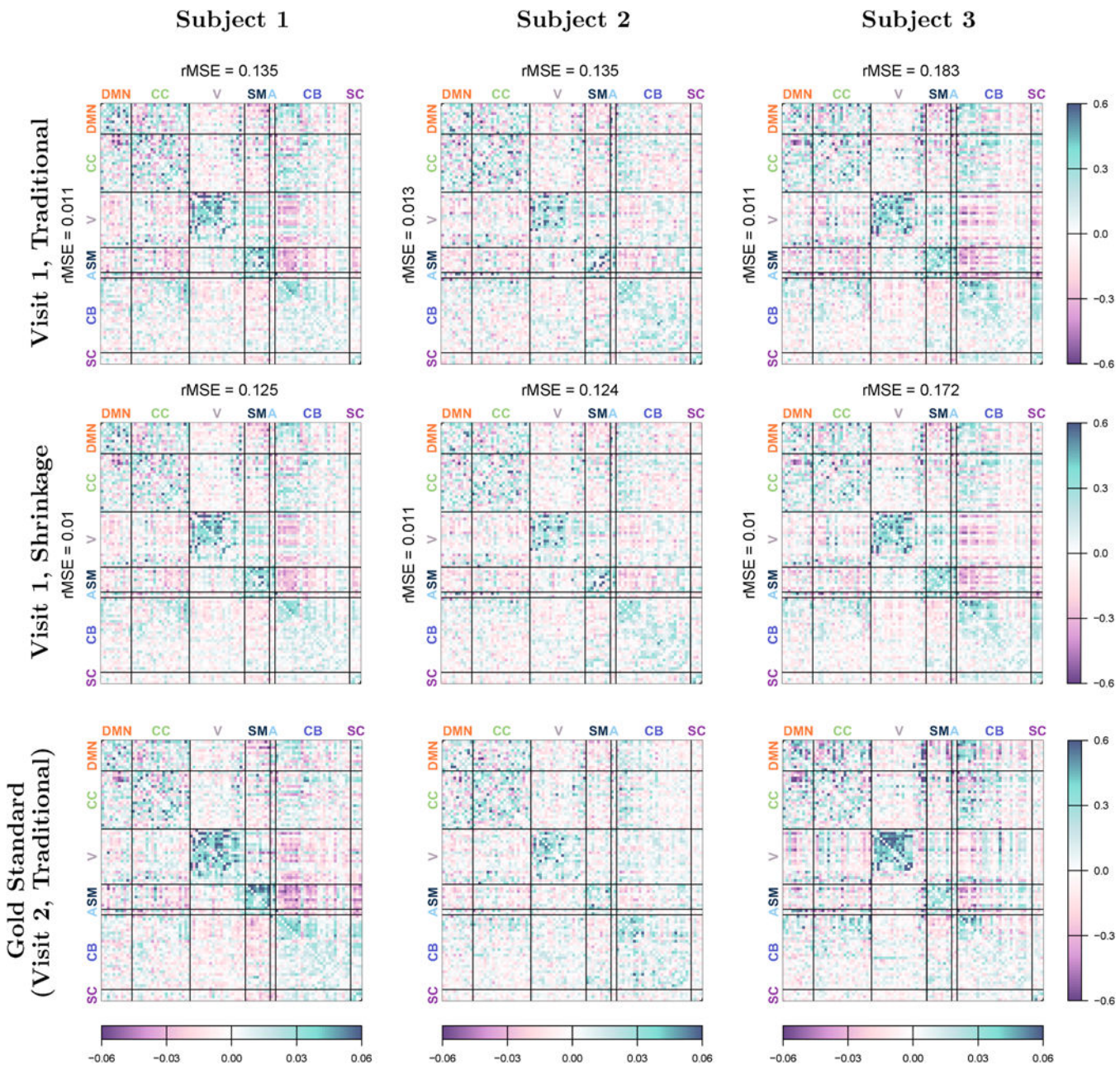


Figure 4. Subject-level estimates of FC

Traditional estimates (top row) and single-session shrinkage estimates (middle row) of subject-level FC based on all 1200 volumes of visit 1. Full correlations are displayed in the upper triangle of each matrix; partial correlations are displayed in the lower triangle. The bottom row shows FC for each subject at visit 2, based on the traditional estimate using all 1200 volumes. This estimate is used as a proxy for the truth to assess the accuracy of the visit 1 estimates. The root mean squared error (rMSE) of full and partial correlations, relative to visit 2, are displayed in the margins of each matrix.

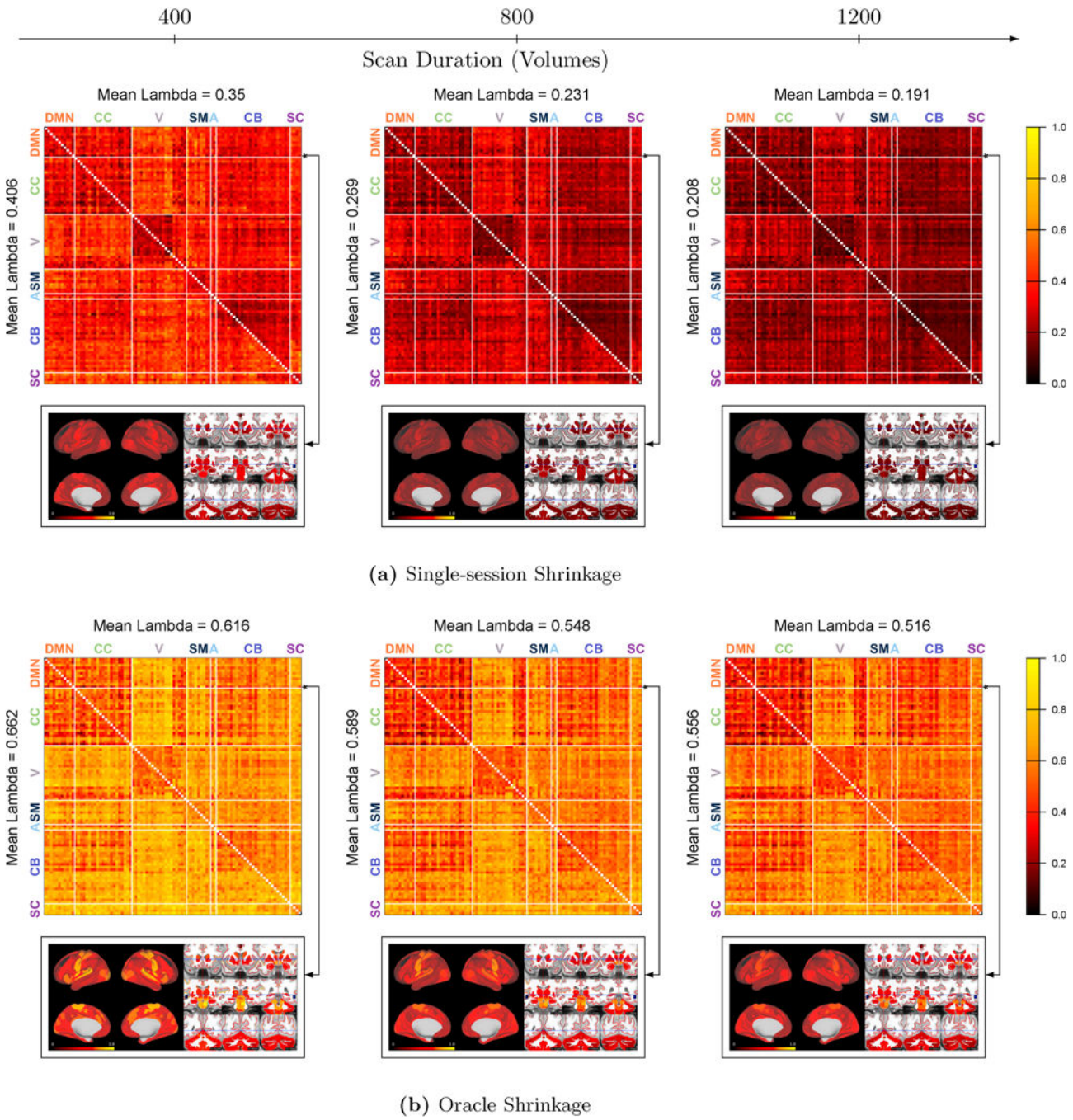


Figure 5. Degree of shrinkage

Each matrix shows the degree of shrinkage $\lambda(q, q') \in [0, 1]$ for the FC between nodes q and q', based on single-session shrinkage(a) and oracle shrinkage (b). The upper triangle corresponds to full correlations and the lower triangle to partial correlations using $\rho = 5$. For one selected seed in the DMN (shown in black in the images), the degree of shrinkage for all connections with this seed is displayed as an image below the matrix, based on full correlations. Overall, the degree of shrinkage decreases as scan length increases; some connections receive more shrinkage than others due to differences in reliability; full and

partial correlations show somewhat similar patterns of shrinkage with some notable differences; and single-session shrinkage tends to be more conservative (less shrinkage) than oracle shrinkage.

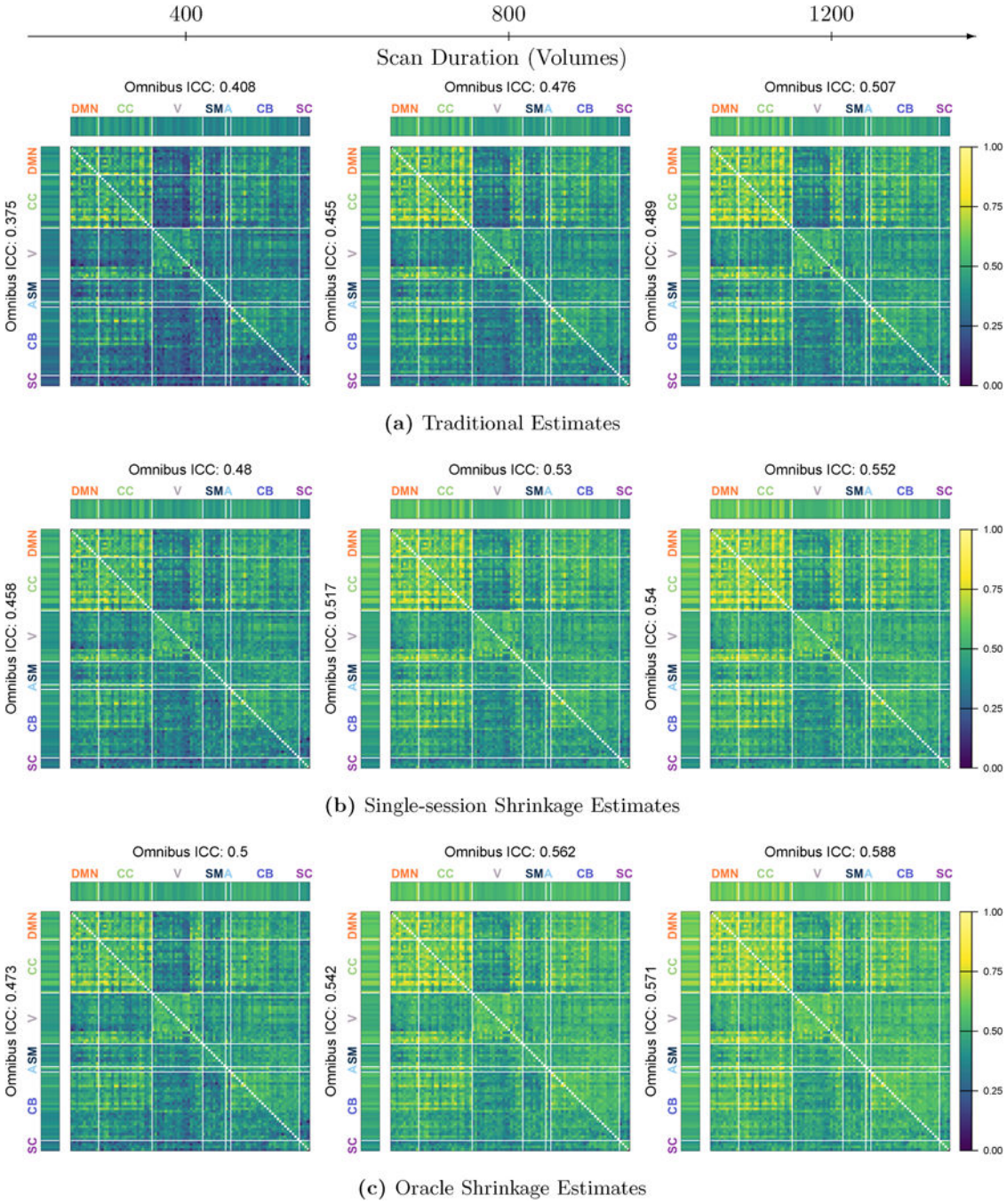


Figure 6. Reliability of FC estimates

For each estimation method and scan duration, the ICC_{MSE} of all unique region-pairs is displayed for full correlations in the upper triangle and for partial correlations in the lower triangle; $I2C2_{MSE}$ for each region and omnibus ICC_{MSE} are displayed in the margins for each measure.

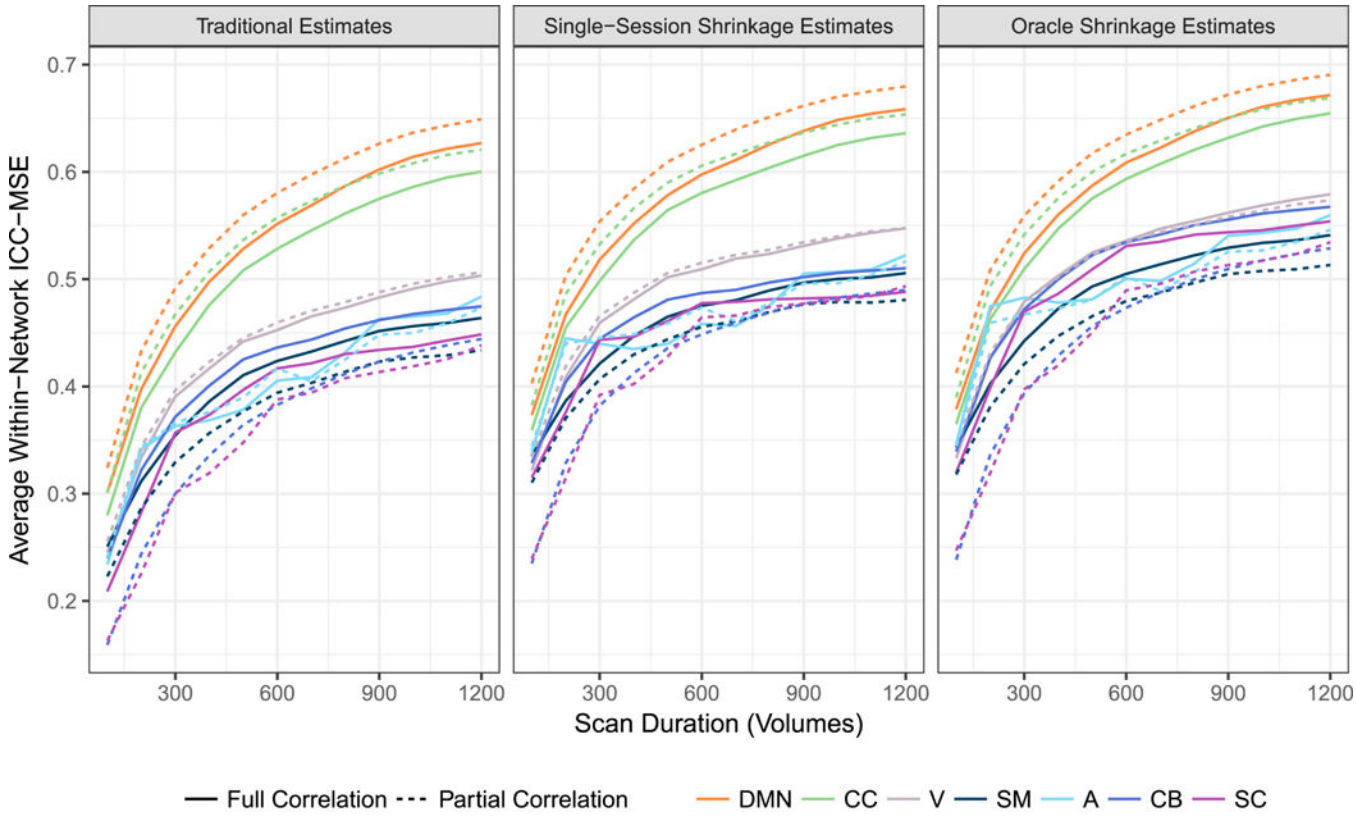


Figure 7. Accuracy of within-network FC estimates

Average ICC_{MSE} across all within-network connections within each network for traditional and single-session shrinkage estimates of FC. For both types of estimates, accuracy tends to taper off with increasing scan duration, but shrinkage estimates are substantially more accurate than traditional estimates across all durations.