A regularized clustering approach to brain parcellation from functional MRI data

Keith Dillon[a] and Yu-Ping Wang[a]

[a]Tulane University Department of Biomedical Engineering, New Orleans, USA

ABSTRACT

We consider a data-driven approach for the subdivision of an individual subject's functional Magnetic Resonance Imaging (fMRI) scan into regions of interest, i.e., brain parcellation. The approach is based on a computational technique for calculating resolution from inverse problem theory, which we apply to neighborhood selection for brain connectivity networks. This can be efficiently calculated even for very large images, and explicitly incorporates regularization in the form of spatial smoothing and a noise cutoff. We demonstrate the reproducibility of the method on multiple scans of the same subjects, as well as the variations between subjects.

Keywords: Functional MRI, Connectomics, Parcellation, Resolution, Clustering

1. INTRODUCTION

Understanding the functional operation of the brain and its diseases remains a daunting challenge. This is undoubtedly due to the brain's sheer complexity, as well as difficulty in imaging the function of the human brain *in vivo*, except at a relatively coarse scale. Results from neurology, however, demonstrate the value of coarse-scale structural information. For example, a clinician can often predict the location of a brain lesion from observing functional deficits in movement or sensing, and similarly predict deficits from observation of a lesion's location in neuroimaging. This perspective is further supported by differences in cytoarchitecture between these regions, as identified by Brodmann and others in cadavers.[1] As a result, neuroanatomy texts describe the brain in such a modular fashion, identifying regions of the brain with their functional role, just as is done with other organs of the body. Although clinically relevant, there exists a need to refine our resolution to capture the true complexity of the brain's inherently diverse internal structural features.

Functional Magnetic Resonance Imaging (fMRI) is a popular technique for observing the functional activity of the brain at this intermediate scale. Each voxel, on the order of millimeters, potentially describes the aggregate activity of many thousands of neurons via a complex relationship with local blood oxygenation levels. Typically, analyses of fMRI have used pre-defined neuroanatomical maps to define regions of interest (ROI);[2] the activity in each ROI being identified by the average of the temporal signals for the voxels contained in the ROI. The average temporal signals from each ROI, then, are compared to the temporal structure of various tasks and in this way the function of different ROIs are identified. As the structure of the brain is known to vary significantly between individuals, particularly in association regions of the cortex,[3] the use of pre-defined regions may result in inconsistent or erroneous results.

A variety of techniques have been developed to determine module definitions, i.e. parcellation, using the fMRI data itself, typically based on clustering techniques.[4] One of the oldest and most fundamental such techniques is k-means clustering, a greedy technique for grouping together the most similar vectors into each cluster. k-means clustering is applied to fMRI data with the assumption (common to most clustering methods) that signals in voxels within a ROI should share more correlation with each other than with voxels in other ROIs.[5]

Of course the brain is densely-interconnected, and a plethora of recent research has centered upon identifying this network structure, known as the connectome.[6] A number of researchers have used fMRI to argue that network activity, even at this scale, remains critically important for brain function and the expression of disease.[7]

Further author information: (Send correspondence to Y.P.W.)
Y.P.W.: E-mail: wyp@tulane.edu, Telephone: 1 504 865 5867
K.D.: E-mail: kdillon1@tulane.edu

Most commonly, such analyses have used the familiar ROI-based descriptions, forming network nodes from the same regions that previously were analyzed independently.[8] Traditionally, the means of determining which ROIs are networked together is essentially the same as determining which voxels make up an ROI; namely, pairwise correlations of activity. Such pairwise comparisons are clearly a crude approach to identify potentially complex networks involving several interacting ROIs. Therefore recent research has sought to use more sophisticated multi-variable approaches, such as Graphical Models,[9] to refine network estimates. In practical application, these sophisticated methods still require parcellation as the first step of fMRI data processing using a defined template.

In this paper, we present a new approach to parcellation which inherently considers network connectivity, and which can be efficiently applied on full-sized fMRI scans. The approach is based on inverse problem methods for imaging, or more specifically, the computational estimation of system resolution. A resolution cell in imaging can be defined as the smallest region within which no further detail may be discerned. Here, we will analogously consider the sets of possible edges in a network model, which cannot be differentiated using the data. It will be shown that clustering such edges produces a more consistent parcellation, compared to either clustering the voxel signals directly, or clustering their covariances.

2. METHODS

Let $\mathbf{A}$ be a matrix containing fMRI data, with time series as columns (which we will assume has been standardized and otherwise preprocessed to remove artifacts). The regression problem to estimate the functional connectivity of the kth voxel, known as neighborhood selection,[10] can be written as

$$\mathbf{A}\mathbf{x}_k = \mathbf{a}_k, \tag{1}$$

where $\mathbf{a}_k$ is the kth column of $\mathbf{A}$ and $\mathbf{x}_k$ is the kth column of the weighted adjacency matrix $\mathbf{X}$. Note that we do not restrict the diagonal to zero, hence we allow self loops in the network. If $\mathbf{A}$ was invertible, we could solve directly for the connectivity by applying the inverse as $\mathbf{x}_k = \mathbf{A}^{-1}\mathbf{A}\mathbf{x}_k = \mathbf{I}\mathbf{x}_k$, where $\mathbf{I}$ is the identity matrix. In more general cases, no inverse will exist, and a regularized result is the best that can be found. For ℓ_2-regularized approaches, we replace the inverse $\mathbf{A}^{-1}$ with a generalized or pseudoinverse, which we will denote by $\mathbf{A}^\dagger$, as in

$$\begin{aligned} \hat{\mathbf{x}}_k &= \mathbf{A}^\dagger\mathbf{A}\mathbf{x}_k \\ &= \mathbf{R}\mathbf{a}_k, \end{aligned} \tag{2}$$

defining the resolution matrix $\mathbf{R} = \mathbf{A}^\dagger\mathbf{A}$.[11] The resolution matrix is commonly interpreted as an approximation to the identity matrix;[12] it is analyzed to help understand the information loss in the reconstruction problem. For example, Fig. 1 depicts an image of an identity matrix versus a resolution matrix. For this resolution matrix

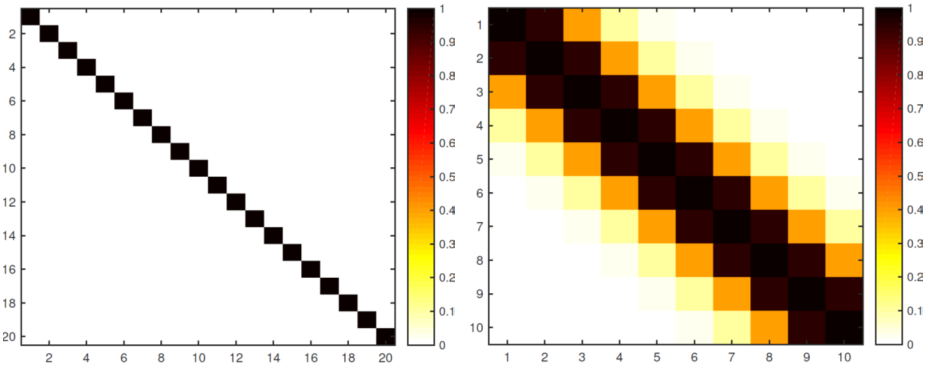

Figure 1. Identity matrix (left) versus resolution matrix (right) which depicts a blurring of each sample among the nearest three samples.

example, we see that a blurring of the information has occurred and details finer than about three samples will

not be present in the estimate. When this perspective is adapted to the network estimation problem of Eq. (1), $\mathbf{R}$ describes a "blurring" in our ability to discern edges in the adjacency matrix $\mathbf{X}$. A single point in the estimate, $(\hat{\mathbf{x}}_k)_i$, would be computed from Eq. (2) as $(\hat{\mathbf{x}}_k)_i = (\mathbf{R}\mathbf{x}_k)_i = \mathbf{r}_i^T\mathbf{x}_k$, where $\mathbf{r}_i^T$ is the ith column of the resolution matrix. Hence a spread of the resolution matrix as in Fig. 1 will mean $(\hat{\mathbf{x}}_k)_i$ contains an average over multiple samples of the true adjacency matrix column $\mathbf{x}_k$.

We can see that this is a kind of minimal choice of blurring region by solving for the best approximation to an inverse, using the following optimization program

$$\min_{\mathbf{y}_i} \|\mathbf{y}_i^T\mathbf{A} - \mathbf{e}_i^T\|_2^2, \tag{3}$$

where $\mathbf{e}_i$ is a vector of zeros with a value of one in the ith element. This optimization must be performed for every row $i = 1, ..., m$. With the ℓ_2-norm, this problem can be solved analytically. For the case of an underdetermined $\mathbf{A}$, we get

$$\mathbf{y}_i = (\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{a}_k, \tag{4}$$

which is a row of the pseudoinverse. So $\mathbf{r}_i^T = \mathbf{y}_i^T\mathbf{A}$ is an optimal solution that minimizes off-diagonal values, in effect finding the minimal blurring of information that can be resolved. This optimization problem may also be viewed from an intuitive perspective based on seeking a function of the unknown variables (in this case the unknown network edge weights) which is estimable.[13] We considered this perspective in a previous publication,[14] where we noted systems of the form $\mathbf{r}_i^T = \mathbf{y}_i^T\mathbf{A}$ provide simple examples of "unambiguous components", features which we may estimate with high confidence. In computational imaging, resolution cells may be viewed as local averages which may be estimated from the data. This averaging operator is $\mathbf{r}_i$, the row of the resolution matrix. In the network estimate problem of Eq. (1), $\mathbf{r}_i$ gives an average over edge weights which link to the ith node, which we compute by applying $\mathbf{y}_i$ to the data.

Fig. 2 depicts a small network consisting of three subgroups of nodes. During each of three time intervals,

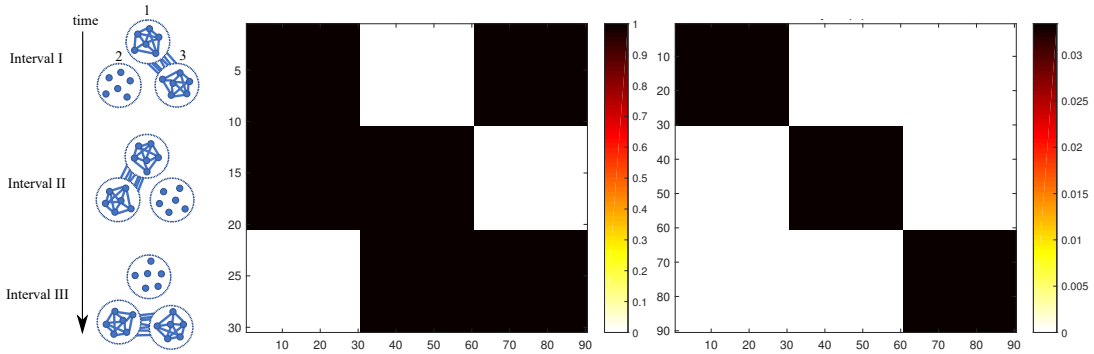

Figure 2. Simulated network consisting of three subnetworks (labeled 1, 2, and 3), which are simultaneously active in different combinations for three time intervals (labeled I, II, and III); corresponding data matrix (middle) and resolution matrix (right)

the nodes in two of the subgroups are all active (depicted as being densely interconnected in the figure). Time is arranged vertically to correspond to the time series in the data matrix. The time intervals correspond to groups of rows (in successive groups of ten), and the node activity is described by the columns (in successive groups of 30). In Fig. 2 we see that the resolution matrix, formed by applying the pseudoinverse of the data matrix to the data matrix itself, is able to separate the three groups but not resolve nodes any further.

2.1 Noise Regularization

The resolution matrix describes the limit to which diversity in the data may be exploited to determine a solution to an inverse problem. Noise, however, provides a false source of diversity which can lead to misleading results. This is similar to problems caused by noise when inverting a poorly-conditioned linear system, where noise is

known to cause a loss of resolution.[15] Note that our case here differs from usual treatments of noise, in that our noise is an error on the entire data matrix $\mathbf{A}$. The result is that noise produces an overestimate of resolution. This is demonstrated in Fig. 3, where the naive (i.e., unregularized) resolution matrix is much closer to the identity.

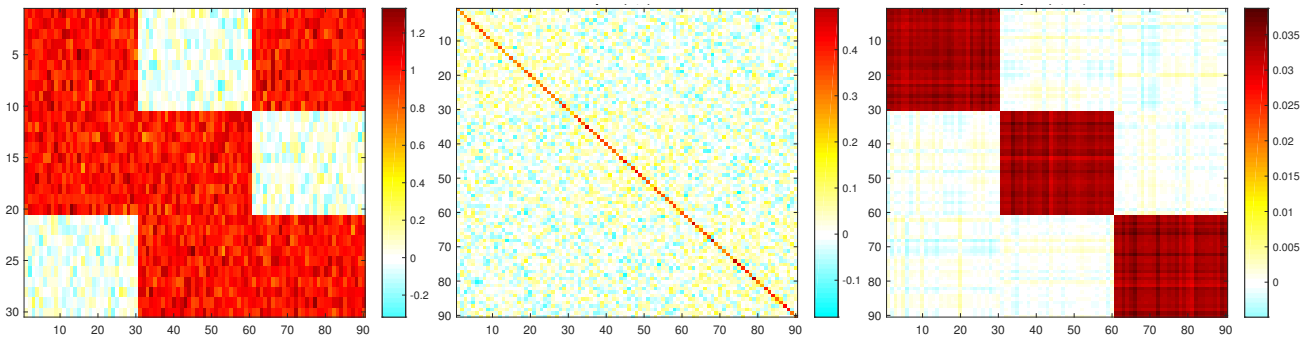

Figure 3. Noisy version of data matrix (left), resolution matrix estimate with no regularization (middle), and with regularization (right), demonstrating false gains in resolution due to noise and their correction by regularization.

Regularization, of course, is a technique to address noise by imposing knowledge about likely noise versus signal components. The most straightforward (and computationally simple) approach would be to impose a regularization penalty which eliminates weak components of the model, as in

$$\min_{\mathbf{x}_k} \|\mathbf{A}\mathbf{x}_k - \mathbf{a}_k\|_2^2 + \mu\|\mathbf{x}_k\|_2^2 \tag{5}$$

This leads to the equivalent problem for estimating the regularized resolution matrix,

$$\min_{\mathbf{y}_i} \|\mathbf{y}_i^T\mathbf{A} - \mathbf{e}_i^T\|_2^2 + \mu\|\mathbf{y}_i^T\|_2^2, \tag{6}$$

with analytical solution

$$\mathbf{y}_i = (\mathbf{A}\mathbf{A}^T + \mu\mathbf{I})^{-1}\mathbf{a}_k, \tag{7}$$

As Fig. 3 demonstrates, an accurate choice of μ can indeed reduce the resolution to a more realistic result. Of course the choice of μ is a non-trivial question, as it implies we know the cutoff between signal and noise. As our goal is parcellation, we will utilize a tactic inspired by the above example, and seek the regularization level at which a coherent parcellation emerges.

In the example of Fig. 3, the noise variation is clearly lower than the signal variation, hence setting a cutoff to eliminate noise is simple. In fMRI data, however, it is not so clear this is the case; the data generally undergoes a great deal of preprocessing, including spatial smoothing and temporal filtering, to eliminate unwanted components. Spatial smoothing is particularly interesting in parcellation as it imposes a kind of distance regularization for clustering methods.[4] Distance regularization, furthermore, has been noted to be of general value to parcellation efforts.[5] When combined with ℓ_2-norm regularization, spatial smoothing allows local variations to be eliminated completely from the resolution estimate. In order to demonstrate these effects with real data, first we must describe the resolution clustering algorithm.

2.2 Clustering Resolution

Clustering of $\mathbf{R}$ can be performed nearly as efficiently as conventional k-means clustering, and hence be applied to realistic data sizes. If we reconsider the resolution matrix of Fig. 1, the rationale for a clustering stage is obvious: first the resolution calculation separates resolvable regions from one another, then clustering can tell us

the remaining groups of unresolvable nodes. A basic parcellation approach based on k-means clustering of the columns of $\mathbf{A}$, can be implemented as follows:

1. Choose number of clusters K and initialize cluster centers $\mathbf{c}_k,\ k = 1, ..., K$;
while *Convergence criterion not met* **do**
   2. Calculate distances D_{ik} between every column $\mathbf{a}_k$ and every cluster center $\mathbf{c}_i$;
   3. Label each column as belonging to nearest cluster center: $l_k = \arg\min_i D_{ik}$;
   4. Recalculate cluster centers as mean over data columns with same label: $\mathbf{c}_i = \frac{1}{|S_i|}\sum_{j\in S_i}\mathbf{a}_j$, where $S_i = \{k|l_k = i\}$;
end

Algorithm 1: Basic k-means applied to data columns of $\mathbf{A}$.

A typical stopping criterion would be to run the algorithm until the cluster labels stop changing. This proved sufficient for our data analyses. Starting clusters were chosen randomly, using the same random choices for each of the methods to be compared.

To apply clustering to the resolution matrix, we replace $\mathbf{A}$ with $\mathbf{R} = \mathbf{A}^\dagger\mathbf{A}$. However, for a typical dataset, the matrix $\mathbf{A}$, of size $n \times m$, may have hundreds of rows (m is small) but hundreds of thousands of columns (n is large). Hence, the resolution matrix, which is $n \times n$, will be extremely large, requiring many gigabytes of storage. We can avoid storing $\mathbf{R}$ by maintaining the pseudoinverse $\mathbf{A}^\dagger$ instead, which is the same size as $\mathbf{A}^T$. Then each calculation in the algorithm, e.g. each matrix-vector product of the form $\mathbf{Rv}$, can be implemented in two steps as in $\mathbf{A}^\dagger(\mathbf{Av})$. In particular, the squared distances between a given center $\mathbf{c}_i$ and a column $\mathbf{r}_k$ of $\mathbf{R}$ can be calculated as

$$\begin{aligned} D_{ik}^2 &= \|\mathbf{c}_i - \mathbf{r}_k\|_2^2 \\ &= \mathbf{c}_i^T\mathbf{c}_i + \mathbf{r}_k^T\mathbf{r}_k - 2\mathbf{c}_i^T\mathbf{r}_k \\ &= \mathbf{c}_i^T\mathbf{c}_i + \mathbf{r}_k^T\mathbf{r}_k - 2\mathbf{c}_i^T\mathbf{A}^\dagger\mathbf{a}_k. \end{aligned} \tag{8}$$

Since we are only concerned with the class index i of the cluster with the minimum distance to each column, we do not need to compute the $\mathbf{r}_k^T\mathbf{r}_k$ term, so we can compute

$$\begin{aligned} l_k &= \arg\min_i D_{ik}^2 \\ &= \arg\min_i \left\{\mathbf{c}_i^T\mathbf{c}_i - 2\mathbf{c}_i^T\mathbf{A}^\dagger\mathbf{a}_k\right\}. \end{aligned} \tag{9}$$

By forming a matrix $\mathbf{C}$ with cluster centers $\mathbf{c}_i$ as columns, we can efficiently compute the cross term in brackets for all i and k as $(\mathbf{C}^T\mathbf{A}^\dagger)\mathbf{A}$, a K by n matrix.

Similar tactics can be used to efficiently compute the mean over columns in each cluster, by noting that the mean over a set S of column indices can be written as

$$\begin{aligned} \mathbf{c}_i &= \frac{1}{|S|}\sum_{j\in S}\mathbf{r}_j \\ &= \frac{1}{|S|}\sum_{j\in S}\mathbf{A}^\dagger\mathbf{a}_j \\ &= \frac{1}{|S|}\mathbf{A}^\dagger\sum_{j\in S}\mathbf{a}_j. \end{aligned} \tag{10}$$

In general, we see that clustering of $\mathbf{R}$ requires additional storage for a matrix the same size as $\mathbf{A}$ (i.e., $\mathbf{A}^\dagger$), and roughly double the number of calculations as conventional clustering, with two matrix-vector multiplies replacing each single one in the conventional algorithm.

Finally, we note that very similar code can be used to efficiently cluster the sample covariance matrix, as an alternative approach to network-based parcellation. Again, this matrix is often too large to fit in memory (in fact it is the same size as $\mathbf{R}$). If $\mathbf{A}$ is standardized, we can estimate the sample covariance as $\mathbf{A}^T\mathbf{A}$, and use an analogous approach to the above, replacing $\mathbf{A}^\dagger$ with $\mathbf{A}^T$. We will also use this approach for comparison, referring to it as clustering of $\mathbf{A}^T\mathbf{A}$.

3. FMRI DATA ANALYSIS

The effect of regularization and spatial smoothing is as follows. Fig. 4 shows the effect of clustering the resolution matrix for a single fMRI dataset with a range of different (spatial) filter sizes. The data is randomly-chosen from the Philadelphia Neurodelphia Cohort (PNC), a publicly-available set containing multiple fMRI scans per individual, including both resting-state and tasks.[16] The data is not yet preprocessed, hence not yet aligned to normalized coordinates nor is the background removed.

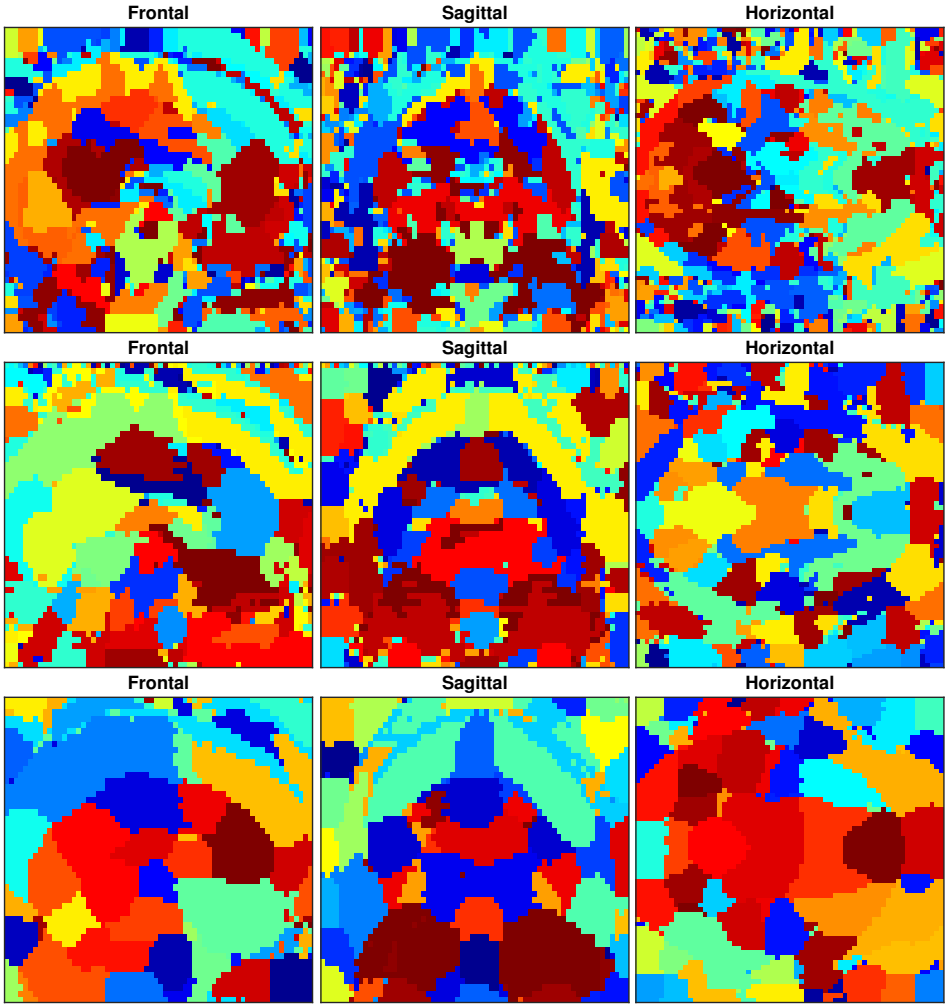

Figure 4. Slices of resolution clustering result after spatial filtering with kernels of size 1mm (top row), 3mm (middle row), and 5mm (bottom row); μ is set to a very small number. Each row depicts three slices of a final clusters. Colors are chosen randomly for each cluster. $K = 250$ (number of clusters).

From this result we generally see that for lower amounts of spatial smoothing, we get many fragmented details, but at 5mm a fairly smooth parcellation emerges. As 5 mm is a common choice for spatial smoothing in preprocesssing fMRI generally, we used this value to proceed. This choice also enabled additional preprocessing steps such as segmenting the brain and normalizing coordinates.

Next we considered the choice of number of clusters. As depicted in Fig. 5, we tested a range of cluster numbers and calculated the fraction of explained variance for each, using multiple clustering techniques. We compared clustering of the raw data $\mathbf{A}$, the resolution matrix $\mathbf{R}$, and the sample covariance matrix $\mathbf{A}^T\mathbf{A}$. We also compared the results to clustering the positions themselves, to provide a baseline which describes the extreme of distance regularization which exploits no information from the fMRI data. The plot depicts a test of

Table 1. Average dice coefficient with best matching ROI

	Same-subject-Cross-scan				Cross-subject-Same-scan			
Method	1-2	1-3	2-3	All	1	2	3	All
$\mathbf{A}$	0.3973	0.2319	0.2354	0.2882	0.2868	0.3030	0.3271	0.3056
$\mathbf{R}$	0.4954	0.5005	0.5358	0.5106	0.3654	0.3925	0.4039	0.3873
$\mathbf{A}^T\mathbf{A}$	0.2015	0.2072	0.2172	0.2086	0.1510	0.1576	0.1599	0.1562

K ranging from 5 to 200 in steps of 5, for a single subject scan which has been preprocessed, including a 5mm spatial smoothing step. Note that the explained variance was computed using the filtered data.

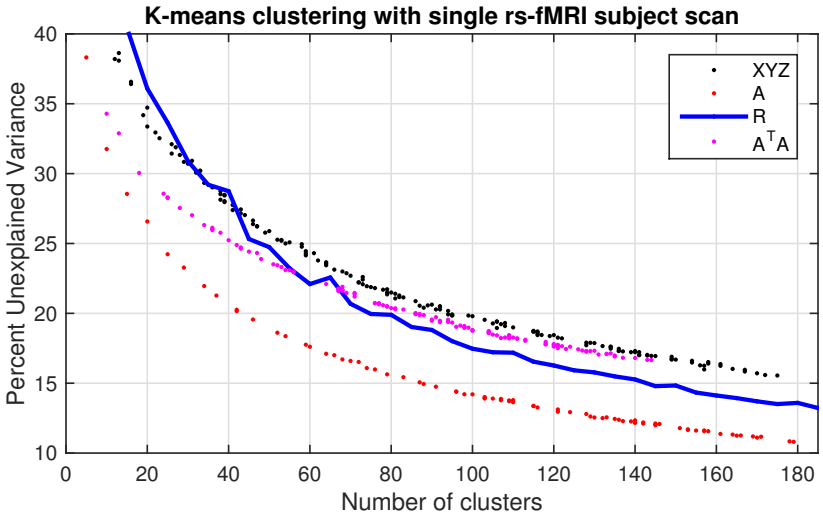

Figure 5. Cluster unexplained variance for different choices of matrices and different numbers of clusters K; XYZ refers to clustering of the $3 \times n$ matrix of voxel locations.

Explained variance was tested by taking the clusters found from clustering the different matrices, then applying them to the $\mathbf{A}$ matrix. Hence, we would expect that directly clustering $\mathbf{A}$ would have an advantage in this metric, and indeed Fig. 5 shows that clustering of $\mathbf{A}$ has the lowest unexplained variance. We also find that directly clustering $\mathbf{R}$ results in clusters that perform second-best, and notably better than clustering of position alone for $K > 70$. This demonstrates that clustering of $\mathbf{R}$ imposes information beyond a simple distance-regularized tiling of the space.

Given the results of Fig. 5, we chose a cluster number of $K = 100$, providing an elbow criterion where every method explains at least 80 percent of the variance. Fig. 6 gives cross-sections of the clustering results for our methods. Interesting and realistic-looking structures emerge in the results, although it should be noted that clusters of $\mathbf{A}^T\mathbf{A}$ look somewhat overly-complex. Fig. 7 gives projections of an example cluster in the frontal lobe for $\mathbf{R}$-clustering. Again, promising results are seen in terms of left-right symmetry and complex shape, which suggests more information is incorporated beyond simple tiling of positions.

Next, repeatability of our method was compared with different scans. We randomly selected ten subjects from the PNC dataset which had three available fMRI scans each, and which did not appear to have problematic artifacts (such as cropping of part of the head due to off-center acquisition). All three methods ($\mathbf{A}$, $\mathbf{R}$, and $\mathbf{A}^T\mathbf{A}$) were applied to each scan, then different pairs of scans for each method were compared by computing average dice coefficients between the clusters in the two scans. Fig. 8 (a) shows the results for different pairs of scans for each subject, using each method. As there were three scans per subject (one resting-state and two different tasks, which we label as scans 1, 2, and 3), the comparisons are for pairs 1-2, 2-3, and 1-3. Fig. 8 (b) further gives the average results for different pairs of scans for each scan type, computed between all pairwise combinations of all ten subjects. These results are also summarized in Table 1. We see that clustering of $\mathbf{R}$ consistently yields the most repeatable clusters in almost all cases.

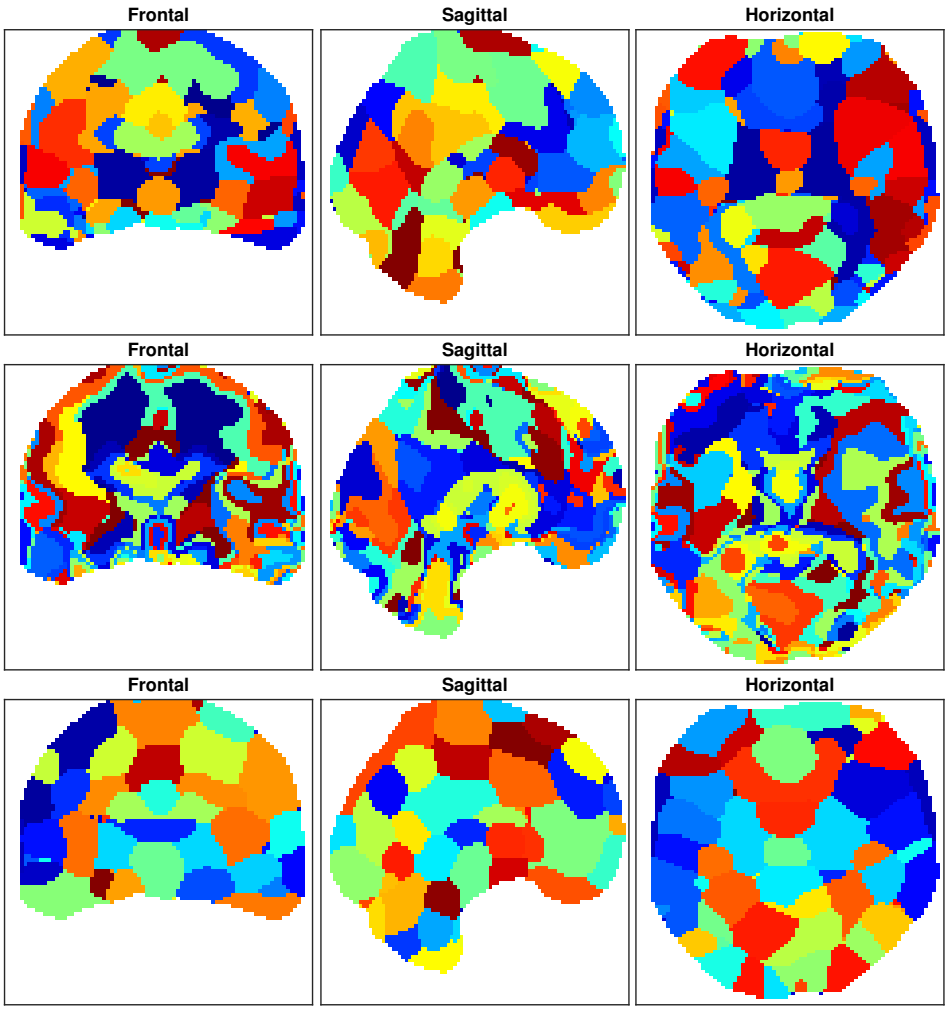

Figure 6. Cross sections of clusters based on clustering columns of $\mathbf{A}$ (top row), columns of $\mathbf{A}^T\mathbf{A}$ (middle row), and columns of $\mathbf{R}$ (bottom row); 100 clusters.

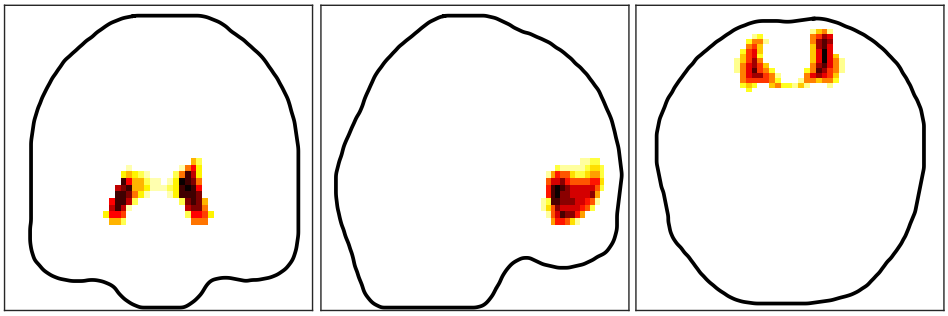

Figure 7. Projections of single class from clustering of resolution matrix.

4. DISCUSSION

In this paper, we introduced an approach to parcellation of fMRI data which incorporates network connectivity information. We also outlined an efficient algorithm for implementation, which can be applied to any case where the popular k-means approach was previously used. This proved adequate for clustering the largest available data sizes available with standard desktop computing hardware, taking only a few minutes. We also found that the result of clustering provided superior repeatability to other methods, with comparable efficiency. In particular,

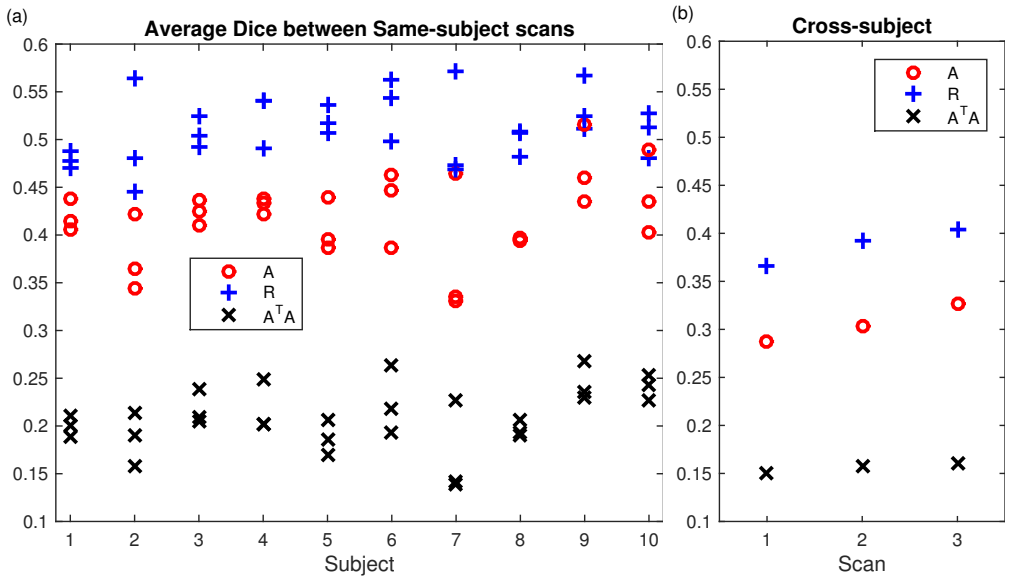

Figure 8. Dice coefficients between clusters for each method, for different pairs of scans of the same subject (a), and for different subjects for the same scan type (b); $K = 100$.

we compared our methods to the clustering of the time-series data directly, essentially the standard for such data-driven parcellation experiments, as well as an intuitive alternative in terms of clustering the covariance matrix directly. We also found that the proposed approach was second only to clustering of $\mathbf{A}$ by itself, when computing the explained variance on the original data with the clusters. As the clustering of $\mathbf{A}$ is a greedy optimization of precisely this metric, it makes sense that a method which incorporates additional information into its clustering goals would be less optimal in the same metric. Additionally, our proposed method's clustering performed better at this metric than a naive clustering of position, which demonstrates the incorporation of data-specific information.

Key to our method is the filtering and regularization step, which we note is also key to successful performance for many (if not most) other methods, though this fact often goes unnoticed. We considered this somewhat subjectively by testing alternative filter sizes and selecting a reasonable-looking result. This was primarily because this reasonable-looking result also agreed with a popular and commonly-used filtering level, largely because it enables other important pre-processing steps such as registration and normalization of coordinates, which made the subsequent comparison of dice coefficients possible. Although our dice coefficients were rather low compared to some other findings in the literature, this is likely due to mismatches in registration, especially between different subjects. In any case, the handicap is suffered by all of the methods intractably, and do not diminish their utility in providing a regularized clustering approach to brain parcellation.

ACKNOWLEDGMENTS

The authors wish to thank the NIH (R01 GM109068, R01 MH104680, R01 MH107354) and NSF (1539067) for their partial support.

REFERENCES

[1] Zilles, K. and Amunts, K., "Centenary of Brodmann's map conception and fate," *Nature Reviews Neuroscience* **11**, 139–145 (Feb. 2010).

[2] Evans, A. C., Janke, A. L., Collins, D. L., and Baillet, S., "Brain templates and atlases," *NeuroImage* **62**, 911–922 (Aug. 2012).

[3] Wang, D., Buckner, R. L., Fox, M. D., Holt, D. J., Holmes, A. J., Stoecklein, S., Langs, G., Pan, R., Qian, T., Li, K., Baker, J. T., Stufflebeam, S. M., Wang, K., Wang, X., Hong, B., and Liu, H., "Parcellating cortical functional networks in individuals," *Nature Neuroscience* **18**, 1853–1860 (Dec. 2015).

[4] Thirion, B., Varoquaux, G., Dohmatob, E., and Poline, J.-B., "Which fMRI clustering gives good brain parcellations?," *Frontiers in Neuroscience* **8** (July 2014).

[5] Blumensath, T., Jbabdi, S., Glasser, M. F., Van Essen, D. C., Ugurbil, K., Behrens, T. E. J., and Smith, S. M., "Spatially constrained hierarchical parcellation of the brain with resting-state fMRI," *NeuroImage* **76**, 313–324 (Aug. 2013).

[6] Sporns, O., Tononi, G., and Ktter, R., "The Human Connectome: A Structural Description of the Human Brain," *PLOS Computational Biology* **1**, e42 (Sept. 2005).

[7] Craddock, R. C., Jbabdi, S., Yan, C.-G., Vogelstein, J. T., Castellanos, F. X., Di Martino, A., Kelly, C., Heberlein, K., Colcombe, S., and Milham, M. P., "Imaging human connectomes at the macroscale," *Nature Methods* **10**, 524–539 (June 2013).

[8] Stanley, M. L., Moussa, M. N., Paolini, B. M., Lyday, R. G., Burdette, J. H., and Laurienti, P. J., "Defining nodes in complex brain networks," *Frontiers in Computational Neuroscience* **7** (Nov. 2013).

[9] Sporns, O., "From simple graphs to the connectome: Networks in neuroimaging," *NeuroImage* **62**, 881–886 (Aug. 2012).

[10] Meinshausen, N. and Bhlmann, P., "High-dimensional graphs and variable selection with the Lasso," *The Annals of Statistics* **34**, 1436–1462 (June 2006).

[11] Jackson, D. D., "Interpretation of Inaccurate, Insufficient and Inconsistent Data," *Geophysical Journal International* **28**, 97–109 (June 1972).

[12] An, M., "A simple method for determining the spatial resolution of a general inverse problem," *Geophysical Journal International* **191**, 849–864 (Nov. 2012).

[13] Rodgers, W. L., "Estimable Functions of Age, Period, and Cohort Effects," *American Sociological Review* **47**(6), 774–787 (1982).

[14] Dillon, K., Calhoun, V., and Wang, Y.-P., "A robust sparse-modeling framework for estimating schizophrenia biomarkers from fMRI," *Journal of Neuroscience Methods* **276**, 46–55 (Jan. 2017).

[15] Dillon, K., Fainman, Y., and Wang, Y.-P., "Computational estimation of resolution in reconstruction techniques utilizing sparsity, total variation, and nonnegativity," *Journal of Electronic Imaging* **25**(5), 053016–053016 (2016).

[16] Satterthwaite, T. D., Connolly, J. J., Ruparel, K., Calkins, M. E., Jackson, C., Elliott, M. A., Roalf, D. R., Hopson, R., Prabhakaran, K., Behr, M., Qiu, H., Mentch, F. D., Chiavacci, R., Sleiman, P. M. A., Gur, R. C., Hakonarson, H., and Gur, R. E., "The Philadelphia Neurodevelopmental Cohort: A publicly available resource for the study of normal and abnormal brain development in youth," *NeuroImage* **124**, 1115–1119 (Jan. 2016).