Automated connectivity-based cortical mapping using registration-constrained classification

K. Eschenburg, D. Haynor, T. Grabowski

K. Eschenburg, D. Haynor, T. Grabowski, "Automated connectivity-based cortical mapping using registration-constrained classification," Proc. SPIE 10578, Medical Imaging 2018: Biomedical Applications in Molecular, Structural, and Functional Imaging, 105782T (12 March 2018); doi: 10.1117/12.2293968

SPIE.

Event: SPIE Medical Imaging, 2018, Houston, Texas, United States

Automated Connectivity-Based Cortical Mapping using Registration-Constrained Classification

K. Eschenburg*[a,b], D. Haynor[a,b], T. Grabowski[b]
[a]Dept. Of Bioengineering, University of Washington, 3720 15th Ave NE, Seattle, WA 98105;
[b]Dept. of Radiology, University of Washington, 1959 NE Pacific St., Seattle, WA 98195

1. Abstract

An important goal in neuroscience has been to map the surface of the human brain, and many researchers have developed sophisticated methods to parcellate the cortex. However, many of these methods stop short of developing a framework to apply existing cortical maps to new subjects in a consistent fashion. The computationally complex step is often the initial mapping of a large set of brains, and it is inefficient to repeat these processes for every new data sample. In this analysis, we propose the use of a library of training brains to build a statistical model of the parcellated cortical surface and to act as templates for mapping new data. We train classifiers on training data sampled from local neighborhoods on the cortical surface, using features derived from training brain connectivity information, and apply these classifiers to map the surfaces of previously unseen brains. We demonstrate the performance of 3 different classifiers, each trained on 3 different types of training features, to accurately predict the map of new brain surfaces.

2. Background

A fundamental assumption [1-4] in neuroscience is that the human neocortex can be compartmentalized into discrete subunits, based on cellular histology, activation patterns, gene expression, connectivity, etc. To identify and characterize these cortical subunits is an important, albeit difficult, goal and has been approached in a variety of ways. One of the first studies examined cortical cyto-architecture using Nissl staining, and these results are still a reference for cortical discretization for many research projects [5]. Recently, with the rise of machine learning and high-performance computing, various groups have applied data-driven approaches to discovering the hidden substructure of the cortex using unsupervised clustering. These methods attempt to find hidden structure in the cortical data by grouping cortical voxels with similar properties into higher-order latent clusters [6]. Few studies, however, have examined how to adapt an existing cortical map, or set of maps, to a new brain image in a way that conforms both to the existing maps and the new data. In this work, we propose an approach to mapping the brain surface of a new subject, using a training dataset of previously mapped brains and data regularly acquired in standard functional MRI scans, i.e. diffusion and resting-state data.

We process a set of high-quality MRI imaging data acquired by the Human Connectome Project (HCP) [7, 8] along with a set of labels that were derived for the HCP data using resting-state fMRI, task-based fMRI, and various cortical surface scalar measures and vetted by trained neuroanatomists [7, 8]. We then treat the HCP fMRI-connectivity-based labels as the dependent variables, and train classifiers to distinguish between each label, based on the characteristics of the functional and structural connectivity data associated with it. We apply these classifiers to previously unseen subjects in the HCP dataset, and characterize the accuracy of the new parcellations as a function of classifier type, input feature data, and weighting schemes. Additionally, we examine the regional homogeneity of predicted regions, defined by the aggregate similarity of the feature vectors within a region, in relation to the homogeneity of the original HCP maps ("ground truth" maps). We show that the new maps are generally consistent with the ground truth maps and demonstrate similar degrees of within-parcel homogeneity.

3. Methods

3.1 Data

We use a set of 85 subjects acquired by the HCP. For each subject, there are 4 resting state images, a diffusion weighted image (DWI), and a structural T1 image. The data were minimally preprocessed by the HCP consortium using a custom pipeline specifically designed for the HCP data [8], and converted to a CIFTI data format, mapping each grey matter voxel or surface vertex to a single index in a vector, referred to as "grayordinates" by the HCP. Importantly, each subject's brain mesh has the same number of vertices, represented by two different vertex resolutions. Additionally,

Medical Imaging 2018: Biomedical Applications in Molecular, Structural, and Functional Imaging,
edited by Barjor Gimi, Andrzej Krol, Proc. of SPIE Vol. 10578, 105782T · © 2018 SPIE
CCC code: 1605-7422/18/$18 · doi: 10.1117/12.2293968

each subject's surface has been spatially normalized to the Montreal Neurological Institute (MNI) template space, such that a given vertex is always in approximately the same area of the brain across subjects.

As described in [8], the minimal preprocessing steps performed by the HCP include running FreeSurfer [9, 10] to generate a white matter surface mesh, which is then registered to MNI space and resampled to 32K vertices. We use this mesh throughout our analysis. The resting state time series are mapped onto the mesh, as are the cortical segmentations and scalar maps generated by FreeSurfer. Additionally, the HCP has released a set of connectivity-based parcellations (CBP), with roughly 180 regions per hemisphere, computed from boundary maps derived from resting-state, task-based fMRI, and various cortical scalar metrics [7]. We use these labels, taking values in {1, 2, …, 180}, as the response variable in our three classifiers.

3.2 Pre-Processing

Each of the 4 resting state acquisitions, with 32k surface vertices and 1200 time-points each, was acquired independently. In order to meaningfully compare the four acquisitions, we first normalize each acquisition to have zero mean and unit variance at each vertex and then concatenate the four matrices together to generate a 4800 time-point matrix. We then compute a 32Kx32K Pearson's correlation matrix, where each vertex is described by the correlation of its aggregated time series vector with the time series of all other vertices. We also run FSL's bedpostx and probtrackx2 [11] on each DWI to model water diffusion in each voxel in the brain, and generate cortex-to-cortex and cortex-to-subcortex probabilistic tractography counts, using the parameter sets recommended by the HCP.

For both data types, we apply a simple dimensionality reduction scheme to reduce the noise in the features. For each vertex, we compute the median correlation of its time series with the time series of vertices in each of the 75 ipsilateral Destrieux regions and 7 ipsilateral subcortical regions (accumbens, amygdala, caudate, hippocampus, pallidum, putamen, thalamus). We also compute the total number of probabilistic streamlines projecting from each vertex to all ipsilateral cortical and subcortical regions. We then take the log-transform of these summed streamline counts to generate log-normal count distributions and to reduce the bias towards short-range streamline connections present in tracking algorithms [12]. This dimensionality reduction scheme produces 4 matrices: 2 of size 32Kx75, representing median correlations and log-transform streamline counts to 75 Destrieux cortical regions, and 2 of size 32Kx7, representing median correlations and log-transform streamline counts to 7 subcortical regions.

It is important to note that unique spatial normalization steps applied by the HCP preprocessing pipeline allow for direct, meaningful comparisons between the same grayordinates across subjects. For example, with regards to the full 32Kx32K correlation matrix (but without loss of generality with regards to specific data type), it would be acceptable to compare the correlation of index (k,l) in subject A with index (k,l) in subject B, because indices k and l correspond to approximately the same anatomical locations in the two subjects. However, it cannot be assumed that all future studies will apply the same preprocessing steps as the HCP -- these studies might not acquire the same images, and those studies that do might not acquire the same quality data as the HCP dataset. In cases where sophisticated spatial normalization of surfaces has not been, or cannot be, applied, it would not be reasonable to assume that indices k and l correspond to the same anatomical locations across subjects.

In order to alleviate this constraint, we make the conservative assumption that most studies interested in performing connectivity analyses will first run FreeSurfer to generate low-resolution cortical maps. With this in mind, we then aggregate the high-resolution connectivity vectors over these FreeSurfer cortical maps. In this way, we can then guarantee that the column indices in the low-dimensional connectivity vectors of each surface vertex actually represent comparable connection profiles across subjects (because they represent large-scale anatomical regions), allowing us to make legitimate and meaningful comparisons of the feature vectors.

3.3 Surface Registration

The goal of our method is to parcellate the cortical surface of any subject, given the resting state or structural connectivity fingerprints of its surface vertices. While one possible approach could train classifiers to distinguish all 180 HCP labels from each other in a one-vs.-one or one-vs.-all classification scheme, we instead constrain the label search space to a localized neighborhood using surface registration. Using a non-rigid, spectral-feature based surface

registration method [13], we match vertices on one surface to vertices on another on the basis of local similarity in surface geometry. We compute a weighted adjacency matrix of each surface, compute its graph Laplacian, L, and diagonalize L. The eigenvectors of L, along with the regionalized, low-dimensional connectivity vectors, are used as features in the point-set registration between the two surfaces. The registration produces two directed matches, where for each brain pair (s_1,s_2), every vertex in s_1 maps to a single vertex in s_2, and vice versa, though the two mappings are only approximate inverses of each other.

It is important to note that, although in this proposal we are predicting the cortical map of a new subject based on a set of template maps derived for the HCP parcellations with labels in $\{1,2 \dots 180\}$, we can apply this same methodology to any set of template cortical maps with any number of unique labels. For example, we could also predict a new cortical map based on the Destrieux parcellation scheme with 75 unique regions [14], the Desikan-Killiany parcellation scheme with 35 unique regions [15], or the Yeo et al. cortical map with 7 (17) intrinsic functional connectivity regions [16].

We represent a surface mesh as a graph $G = (V, E)$, with V representing the vertices and E the edges. For two arbitrary surface meshes, s and t, we define $f_{s,t}: s \mapsto t$ as the mapping of vertices from mesh s to mesh t. Without loss of generality with regards to the template cortical map dependent variable, we define $lab(v, s, p)$ as the true label of vertex $v \in V(s)$ in an arbitrary mesh s for template map p. Given a new brain surface mesh t, a set $S = \{s_1 \dots s_n\}$ of n target surface meshes, a set $P = \{p_1 \dots p_n\}$ of n target label maps, and a set $F = \{f_1 \dots f_n\}$ of mappings $t \mapsto s_i, \forall\, i \in \{1 \dots n\}$, we define

$$st(v, S, F) = \{\, lab(f_i(t(v)), s_i, p_i) \;\forall\; i \in \{1 \dots n\} \,\}$$

as the set of labels that vertex $v \in V(t)$ maps to over the whole set of target brains in S. We can then associate a discrete probability distribution to each vertex in the new brain, describing the empirical frequency with which that vertex maps to any given label across the entire set of target brains. An example of this data is displayed in **Figure 1.a**. We pick an arbitrary vertex with a true HCP label of 52 and show that this vertex most frequently maps to other vertices whose true HCP labels are 9, 39, 51, and 52 – not coincidently, this vertex maps most frequently to vertices with the same label as itself.

It is important to note that $st(v, S, F)$ acts as a sort of discrete prior distribution on the label of vertex v, where the prior probabilities are defined by the frequency with which vertex v maps to any of the regions $m \in M$. Our prior beliefs are then updated by the probability a given classifier assigns to each label.

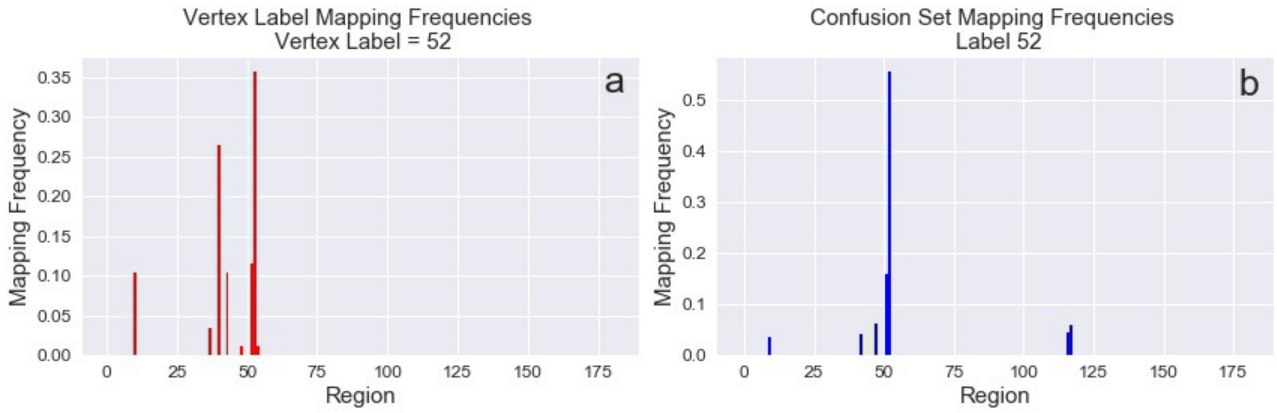

Figure 1. **a**. Vertex-to-label mapping frequencies for a single test vertex with true label = 52. This specific vertex maps predominantly to regions 9, 39, 51, and 52. Only these regions are considered in the final classification step. **b**. Confusion set mapping frequencies for label 52. Label 52 is most often confused with labels 9, 42, 47, 51, 116, and 117. The classifier for label 52 is then trained on data for these confusion labels.

In predicting the label of a new test vertex, given its connectivity data, it would not be efficient to consider all labels as equally likely candidates – for example, if we know a vertex is within the temporal lobe, it would not make sense to consider labels associated with the visual cortex as viable candidates. As such, we can constrain the candidate labels for a new test vertex to $st(v, S, F)$, as described above.

Likewise, given that in the current proposal there exist up to 180 possible regions in label set $M = \{1,2 \dots 180\}$ with which a test vertex could be labeled, it would also be inefficient to train a single classification model to distinguish all 180 regions from one another. Without loss of generality on the size of set M of possible labels, given a training set $S = \{s_1 \dots s_n\}$ of target brain meshes with labeled cortices $P = \{p_1 \dots p_n\}$, for each label value $m \in M$, we compute all pairwise surface mappings between training brains. We define $vtx(m) = \{ v \in V(s) \; \forall \; s \in S \mid lab(v, s, p) = m \}$, as the set of vertices across all training brains with the true label *m*. For each $s \in S$, and its mappings to all remaining training brains $F_s = \{f_{s,s_t} \; s \mapsto s_t \; \forall \; s_t \in S \backslash s\}$, we compute $st(v, s, F_s) \; \forall \; v \in V(s)$. Next, for each label $m \in M$, we define

$$lst(m) = \{ st(v, s, F_s) \; \forall \; v \in V(s), s \in S \mid v \in vtx(m) \}$$

as the set of all labels to which any vertex with label *m* maps to, based only on surface registration, across the full set of training brains. Because the surface registration produces a smooth mapping, the labels in $lst(m)$ will all be in the immediate neighborhood of *m*. We refer colloquially to $lst(m)$ as the "confusion" set of label *m* – those labels that are most often confused with *m* on the basis of the surface registration. An example of this data is displayed in **Figure 1.b**. In the figure, we show the confusion set mapping frequencies for the label 52, and see that the labels most often confused with label 52 are 9, 42, 47, 51, 116, and 117.

Given a set of training data, we compute $lst(m) \; \forall \; m \in M$, and then train $|M|$ classifiers. The *m-th* classifier is trained only on the training data corresponding to those vertices with labels in $lst(m)$, classifying label *m* against $|lst(m) \setminus \{m\}| \ll |M|$ regions. We emphasize that $lst(m) \setminus st(v, s, F_s)$ does not always $= \emptyset$, indicating that the confusion set of a label, and the labels that a given vertex maps, to do not always perfectly overlap. For example, in **Figures 1.a** and **1.b**, we see that while there is overlap in the mapping distribution of vertex with label 52 and the confusion set distribution of label 52, the relative complement of the two sets is not empty.

For any given test brain, *t,* we register *t* to all training brains $s \in S$. For each vertex $v \in V(t)$, we compute $st(v, t, F_t)$. We define $m_{max} = argmax(\; st(v, t, F_t) \;)$ as the label to which vertex *v* maps to most frequently. After training each classifier, we then constrain the *softmax*-classification probabilities of vertex *v* to $lst(m_{max}) \cap \; st(v, t, F_t)$ -- that is, we only consider those labels in the intersection of the confusion set of m_{max} and the labels to which vertex *v* maps to.

3.4 Training Data and Classifiers

We generate 3 sets of training features for each test vertex: 1) – regionalized cortical (1 x 75) and subcortical (1 x 7) resting state correlations, 2) – regionalized log-normal cortical (1 x 75) and subcortical (1 x 7) tractography streamline counts, and 3) -- regionalized resting state correlations and regionalized streamine counts combined. All 3 sets include the sulcal depth, myelin density, and Gaussian curvature scalar maps (1 x 3) from the HCP dataset. We train 3 classifiers: Gaussian Mixture Models (GMM), Random Forests (RF), and Neural Networks (NN). For each classifier, for each of the 3 sets of input features, we perform 5 permutations of cross-validation, averaging the results over 5 sets of 75 training subjects and 10 test subjects to ensure that our classifier results are independent of the training and test sets.

Gaussian Mixture Model Training: For each region $m \in M$, we aggregate all the training data for vertices assigned to label m and fit a 2-component GMM, with diagonal covariance matrices (hyper-elliptical) for each component. Empirical tests showed that fitting more than 2 components, or using full covariance matrices, did not improve the classification accuracy of our model. We assign a test vertex to the region, among the set of candidate regions for that vertex, whose model-based likelihood is maximal.

Random Forest Training: For each of the training subjects $s \in S$, for each region $m \in M$, we aggregate the training data for vertices assigned to labels in *lst(m)* and fit a random forest to the aggregated data, using 60 tree estimators per forest and a maximum tree depth of 5 nodes. Empirical tests showed that larger, more-shallow forests outperformed forests with fewer, deeper trees. We assign a vertex to a region based on the output of the random forest.

Neural Network Training: For each $m \in M$, we aggregate the training data for all $v \in vtx(m)$ for all training subjects. We down-sample the number of samples for each label so that each label has the same number of data points as the minimally-represented class in that model. We then aggregate the down-sampled training data for each $m \in M$. Using the Keras software package, which can be downloaded from (https://keras.io/) [17], we construct a dense, feed-forward neural network with one input layer, three hidden layers with 150 nodes each, and one output layer with 180 nodes. We use the rectified linear unit activation function, apply batch normalization, and train the model for 40 epochs. We assign a vertex to a region based on the output layer node that maximizes that classification probability.

3.5 Classifier Prediction

We perform 3 types of cortical map predictions on the test data. In the first prediction scheme, operating as a type of control, we use only the results of the surface registration to define the final prediction – feature vectors are not included in this prediction step. For a test vertex $v \in V(t)$, we compute $st(v, t, F_t)$, which we can represent as vector of frequencies that we call $\boldsymbol{f}(v) \in \mathbb{R}_{\geq \mathbf{0}}^{|\boldsymbol{M}|}$. For each $i \in \{1 \ldots |M|\}$, $f[i] > 0$ if vertex v maps to region i with frequency f_i and 0 otherwise, and where $\sum_{i=1}^{|M|} f_i = 1$. We define the predicted label of vertex v as

$$class_{surfreg}(v) = m_{max} = argmax(\boldsymbol{f}(v))$$

For the remaining two prediction schemes, for a given test vertex v, we compute the prediction probabilities only for those labels in $\{st(v, t, F_t) \cap lst(m_{\max})\}$ – that is, we consider only those predictions made for classes within the vertex-to-label mapping set and the confusion set of m_{max}. For a given classification model $\boldsymbol{Q}$, we represent the raw classification probabilities of a test vertex $v \in V(t)$, as a vector of probabilities that we call $\boldsymbol{p}(v, \boldsymbol{Q}) \in \mathbb{R}_{\geq \mathbf{0}}^{|\boldsymbol{M}|}$. For each $i \in \{1 \ldots |M|\}$, $p[i] \geq 0$ is the probability assigned to label i for vertex v by model $\boldsymbol{Q}$ if $i \in \{st(v, t, F_t) \cap lst(m_{\max})\}$, and 0 otherwise. In the second prediction scheme, we define the predicated label of vertex v given model $\boldsymbol{Q}$ as the following:

$$class_{unweighted}(v) = argmax(\boldsymbol{p}(v, \boldsymbol{Q}))$$

In the third prediction scheme, for a given test vertex $v \in V(t)$, we weight the probabilities $\boldsymbol{p}(v, \boldsymbol{Q})$ for the candidate labels by the frequencies with which a vertex maps to each candidate, $\boldsymbol{f}(v)$. We assign the vertex to the label with the highest weighted classification probability. For a given classification model $\boldsymbol{Q}$ and test vertex v, we define the predicted label of v as the label that maximizes the following expression:

$$class_{weighted}(v) = argmax(\boldsymbol{f}(v) * \boldsymbol{p}(v, \boldsymbol{Q}))$$

Taken together, we represent the predicted label of a test vertex v for prediction schemes 2 and 3, given a classifier model, as

$$class(v) = argmax(\boldsymbol{f}(v)^{\boldsymbol{w}} * \boldsymbol{p}(v, \boldsymbol{Q}))$$

where the scalar $w \in \{0,1\}$ is the exponent applied to the frequencies. If $w = 0$, then $class(v) = class_{unweighted}(v)$ is based only on the unweighted classification probabilities, and if $w = 1$, then $class(v) = class_{weighted}(v)$ is based on the frequency-weighted classification probabilities.

3.6 Regional Homogeneity

As an additional method of comparing the predicted cortical maps to the ground truth cortical maps, we characterize the regional homogeneity of all of the cortical regions for each predicted map, and compare the computed metrics to the regional homogeneity of the corresponding ground truth regions. For a given cortical map, we define regional homogeneity to be the mean of all pairwise Pearson correlations of feature vectors for vertices within a given region – we compute the homogeneity for all regions in a cortical map, and then average across all regions to generate a mean homogeneity metric for that cortical map. Based on the cortical field hypothesis [1-3], we expect the connectivity

profiles of vertices within a region to be more similar to each other than to vertices in other regions. In total, we generate predicted cortical maps for 18 different parameter combinations for 3 models (Gaussian Mixture Model, Random Forest, and Neural Network), 3 types of training data (regionalized resting state correlations, regionalized structural connectivity counts, and both combined), and 2 prediction schemes (unweighted with $w = 0$, and weighted with $w = 1$). For each set of predicted cortical maps generated with a single combination of parameters, we compute the regional homogeneity of those maps, using only the regionalized resting state correlations or regionalized structural connectivity feature data to compute homogeneity estimates for each map. In total, we generate 36 sets of homogeneity metrics, for all possible combinations of model, training data, weighting scheme, and similarity data.

We also compute the regional homogeneity for all regions in the ground truth HCP maps, as well as all regions in maps computed only from the surface registration, using the same sets of similarity data. As controls, we compute the regional homogeneities for the Desikan-Killiany and Destrieux atlas cortical maps generated by FreeSurfer. We expect the variability of features to increase, and therefore similarity to decrease, as a function of distance between vertices (**Figure 2**). Because the Destrieux and Desikan-Killiany cortical regions are larger than the HCP regions, we expect the mean similarity of the Desikan-Killiany and Destrieux regions to be lower than regions in the ground truth HCP maps, and possibly also lower than the predicted maps.

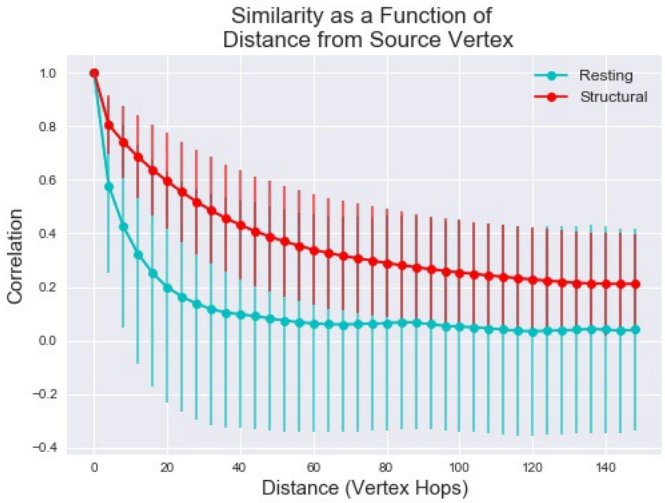

Figure 2. Similarity of feature vectors as a function of geodesic distance between vertices on surface mesh. Structural connectivity features (red) display higher similarity at a given geodesic distance than resting state connectivity features.

3.7 Spatial Overlap of Predicted Parcels with Ground Truth Maps

As a final method of comparing the predicted cortical parcellations with the ground truth parcellations, we compute the Dice coefficient between each ground truth map and predicted maps. The Dice coefficient is defined as

$$Dice(X, Y) = \frac{2|X \cap Y|}{|X| + |Y|}$$

where X in our case is the true cortical map (given by the HCP labels), and Y is the predicted cortical map.

4. Results

We compute classification accuracy relative to the "ground truth" HCP maps as a function of model, weighting scheme, and feature data type. We also report the classification accuracy for the parcels produced using only the surface registration data. We average the classification accuracies across the 5 rounds of cross-validation. We then report regional homogeneity of the features of each cortical map, and characterize the misclassified vertices.

4.1 Classification Accuracy

In **Figure 3 [a-d]** we examine the classification accuracy of each classifier, stratified by training data type and the weighting exponent, w, as well as the accuracy of the cortical maps computed by using surface registration alone. The Gaussian Mixture models **(a)** and Random Forest models **(b)** both outperform the Neural Network models **(c)** when applying the vertex-to-label frequencies as weights (exponent $w = 1$) to the classifier probabilities. However, the Neural Network models outperform both the Gaussian Mixture models and Random Forests models when using unweighted classifier probabilities. We see a dramatic improvement in classification accuracy for both the Gaussian Mixture models and Random Forest models when incorporating the frequencies as weights, and only a slight improvement for the Neural Network models. All three models, irrespective of training data type and exponent weight, are outperformed by surface registration prediction scheme **(d)**.

In **Figure 4**, we show an example of a ground truth HCP map (top row), and a predicted cortical map. The predicted map was generated using a Random Forest model, using the combined resting state and structural connectivity features as training and testing data. We see that the resulting predicted map strongly resembles the true map in terms of label topology. Additionally, from visual inspection, the predicted regions are spatially homogenous and noise-free.

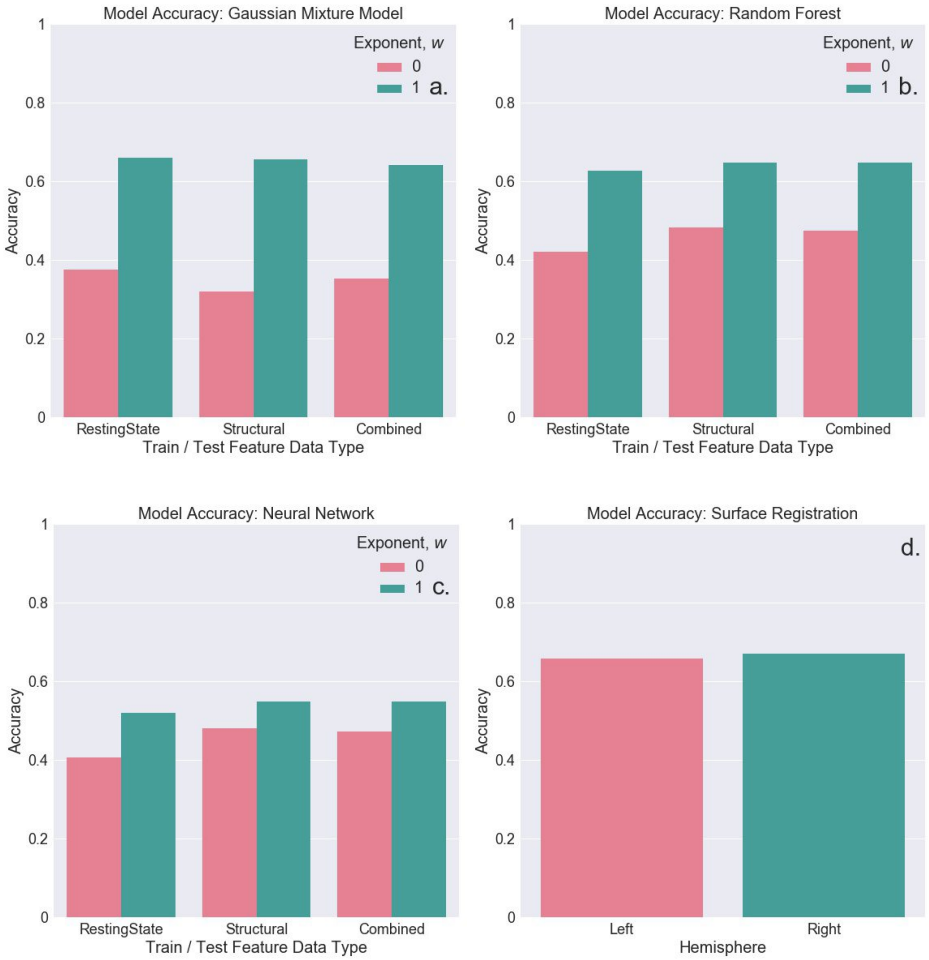

Figure 3. Classification accuracies for maps predicted by each model, and by surface registration. **a.** Gaussian Mixture Model. **b.** Random Forest. **c.** Neural networks. **d.** Surface Registration. **[a-c]** are averaged across hemisphere.

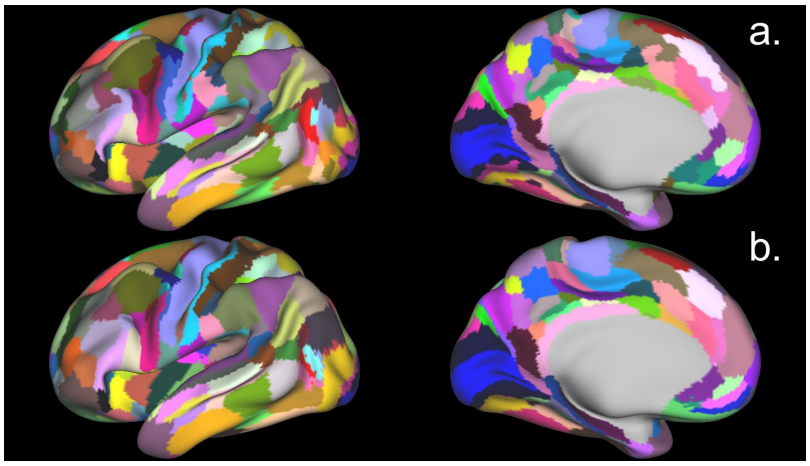

Figure 4. Comparison of true map (**a**) to predicted cortical map (**b**).

When we split the model accuracies by hemisphere, we see that there is in fact a trend towards higher prediction accuracies for the right hemisphere, when compared to the accuracies for the left hemisphere (**Table 5**). Interestingly, maps computed using the combined resting state and structural connectivity information showed no discernable differences in accuracy with respects to maps derived from each of the features individually.

Model Accuracy					
w = 0		Random Forest	GMM	Neural Network	
L	Struct.	0.475	0.309	0.474	
	Rest.	0.415	0.364	0.399	
	Comb.	0.470	0.340	0.462	
R	Struct.	0.488	0.332	0.488	
	Rest.	0.425	0.386	0.412	
	Comb.	0.477	0.364	0.481	
w = 1		Random Forest	GMM	Neural Network	Surface Registration
L	Struct.	0.641	0.654	0.541	0.659
	Rest.	0.623	0.656	0.517	
	Comb.	0.643	0.638	0.540	
R	Struct.	0.653	0.657	0.557	0.669
	Rest.	0.630	0.665	0.523	
	Comb.	0.653	0.645	0.557	

Figure 5. Classification accuracies for maps predicted by each model and by surface registration, now split by hemisphere.

4.2 Characterizing Misclassified Vertices

In addition to computing the classification accuracy of each predicted cortical map, we also characterize the labels of the misclassified vertices for each model. Specifically, we are interested in how "close" the label of a misclassified vertex is to the true label of that vertex. To this end, for each HCP cortical map p, with a corresponding surface mesh s represented as a graph $G = (V, E)$, we compute a label graph $L = (M, N)$, with M the set of labels in the cortical map, and N the set of edges between labels. For two labels $m_{\{i\}}, m_{\{j\}} \in M$, if $[\ (\exists\ v \in V: lab(v, s, p) = m_{\{i\}}) \wedge\ (\exists\ u \in V \backslash v: lab(u, s, p) = m_{\{j\}}) \wedge\ (\ (v, u) \in E\ )\ ] \rightarrow (\ (m_{\{i\}}, m_{\{j\}}) \in N)$. That is, if there exist two adjacent vertices (v, u), one with label $m_{\{i\}}$ and the other with label $m_{\{j\}}$, we consider those two labels to be adjacent in the label adjacency graph, L (**Figure 6**). Then, using Dijkstra's algorithm, we compute the shortest path length between all pairs of nodes in L. For a given predicted cortical map, we compute the following:

$$\frac{1}{N}\sum_{i=1}^{N} Dijkstra\big(class(v), lab(v,s,p)\big)\ \forall\ v \in V\ if\ class(v)\ !=\ lab(v,s,p)$$

representing the mean shortest path length between the predicted label and the true label of all misclassified vertices (**Figure 7**). We see that, on average, the mean shortest path length between the true label and the classified label of misclassified vertices is close to 1, indicating that generally, if a vertex is misclassified, it is typically misclassified as a label directly adjacent to the true label. This is expected given how the candidate labels at a particular vertex are constrained.

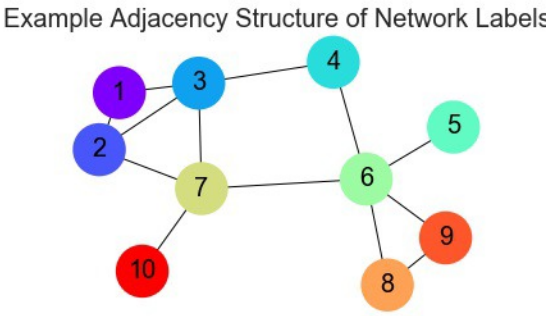

Figure 6. Example label adjacency structure of cortical parcellation, with label values in {1…10}. We can then compute the shortest paths and lengths of these paths between all pairs of nodes (labels) using Dijkstra's algorithm.

Shortest Path Lengths Between Misclassified and True Labels							
w = 0	Training Data			w = 1	Training Data		
	Struct.	Rest.	Comb.		Struct.	Rest.	Comb.
Random Forest	1.076	1.100	1.081	Random Forest	1.044	1.058	1.0436
GMM	1.104	1.090	1.097	GMM	1.044	1.038	1.0509
Neural Network	1.067	1.092	1.068	Neural Network	1.059	1.078	1.0591

Figure 7. Mean shortest path length between classified label and true label for unweighted (left) and weighted (right) predicted cortical maps.

4.3 Regional Homogeneity of Parcels

For each classifier, we generate cortical maps derived from 3 types of feature data (resting state connectivity only, structural connectivity only, and both combined), and compute the homogeneity using only the resting state connectivity or the structural connectivity. For each of these 6 permutations, we average over the left and right hemispheres, and report the results after splitting on the exponent, *w*.

Homogeneity Metrics		
HCP	Similarity Features	
	Struct.	Rest.
	0.763	0.623
Surface Registration	Similarity Features	
	Struct.	Rest.
	0.769	0.622

Figure 8. Mean homogeneity estimates for ground truth HCP maps (top) and surface-registration cortical maps (bottom). Homogeneity is displayed as a function of features used to compute regional homogeneity.

Homogeneity Metrics				
w = 0			Similarity Features	
			Struct.	Rest.
Random Forest	Training Data	Struct.	0.755	0.592
		Rest.	0.739	0.596
		Comb.	0.754	0.591
GMM		Struct.	0.717	0.562
		Rest.	0.740	0.592
		Comb.	0.727	0.565
Neural Network		Struct.	0.758	0.580
		Rest.	0.748	0.596
		Comb.	0.758	0.582
w = 1			Similarity Features	
			Struct.	Rest.
Random Forest	Training Data	Struct.	0.777	0.625
		Rest.	0.768	0.622
		Comb.	0.774	0.624
GMM		Struct.	0.766	0.617
		Rest.	0.769	0.618
		Comb.	0.765	0.615
Neural Network		Struct.	0.767	0.596
		Rest.	0.762	0.616
		Comb.	0.767	0.599

Figure 9. Mean homogeneity estimates for predicted cortical maps (w = 0, top, and w=1, bottom). Homogeneity is displayed as a function of model, training data, and features used to compute regional homogeneity.

In **Figure 8**, we see that the homogeneity estimates of the surface registration very closely resemble those for the ground truth HCP maps. In **Figure 9**, we see that the mean homogeneity estimates computed from the resting state feature data are consistently lower than the metrics computed using the structural connectivity data. However, all 3 classifier models produce similar estimates of homogeneity, regardless of the feature data used to compute the similarity metrics. When using an exponent parameter of $w = 1$, the homogeneity estimates are greater than when using an exponent parameter of $w = 0$. In **Figure 10**, we show the homogeneity estimates for the Destrieux atlas (with 75 regions per hemisphere) and the Desikan-Killiany atlas (with 35 regions per hemisphere). We see that the homogeneity estimates for both the Destrieux atlas and the Desikan-Killiany atlas are less than the estimates for the ground truth maps, surface registration maps, and predicted cortical maps; this is expected since the 75 Destrieux atlas regions and 35 Desikan-Killiany atlas regions are larger than the 180 parcels.

Homogeneity Metrics		
Destrieux Atlas	Similarity Features	
	Struct.	Rest.
	0.719	0.522

Homogeneity Metrics		
Desikan-Killiany Atlas	Similarity Features	
	Struct.	Rest.
	0.645	0.431

Figure 10. Mean homogeneity estimates for the Destrieux atlas cortical maps (top) and the Desikan-Killiany atlas cortical maps (bottom). Homogeneity is displayed as a function of features used to compute regional homogeneity.

In addition to mean homogeneity estimates, we examine the empirical distributions of the homogeneity for each cortical map for all predicted regions, aggregated across all 5 folds of cross-validation. We show the results for the Random Forest model as an example, and compare the Random Forest empirical distributions to the homogeneity of the HCP

cortical maps and the surface registration maps in **Figure 11**. We see that the homogeneity metrics across the three map types produce similarly distributed estimates – the structural connectivity homogeneity is much less variable than the resting state connectivity homogeneity. Additionally, homogeneity metrics of maps generated using only surface registration closely mirror the HCP cortical maps. Along with the results in **Figures 5**, **8**, and **9,** this provides evidence against the use of classification models and in support of simply applying a surface-based registration and computing which region a given vertex maps to most frequently.

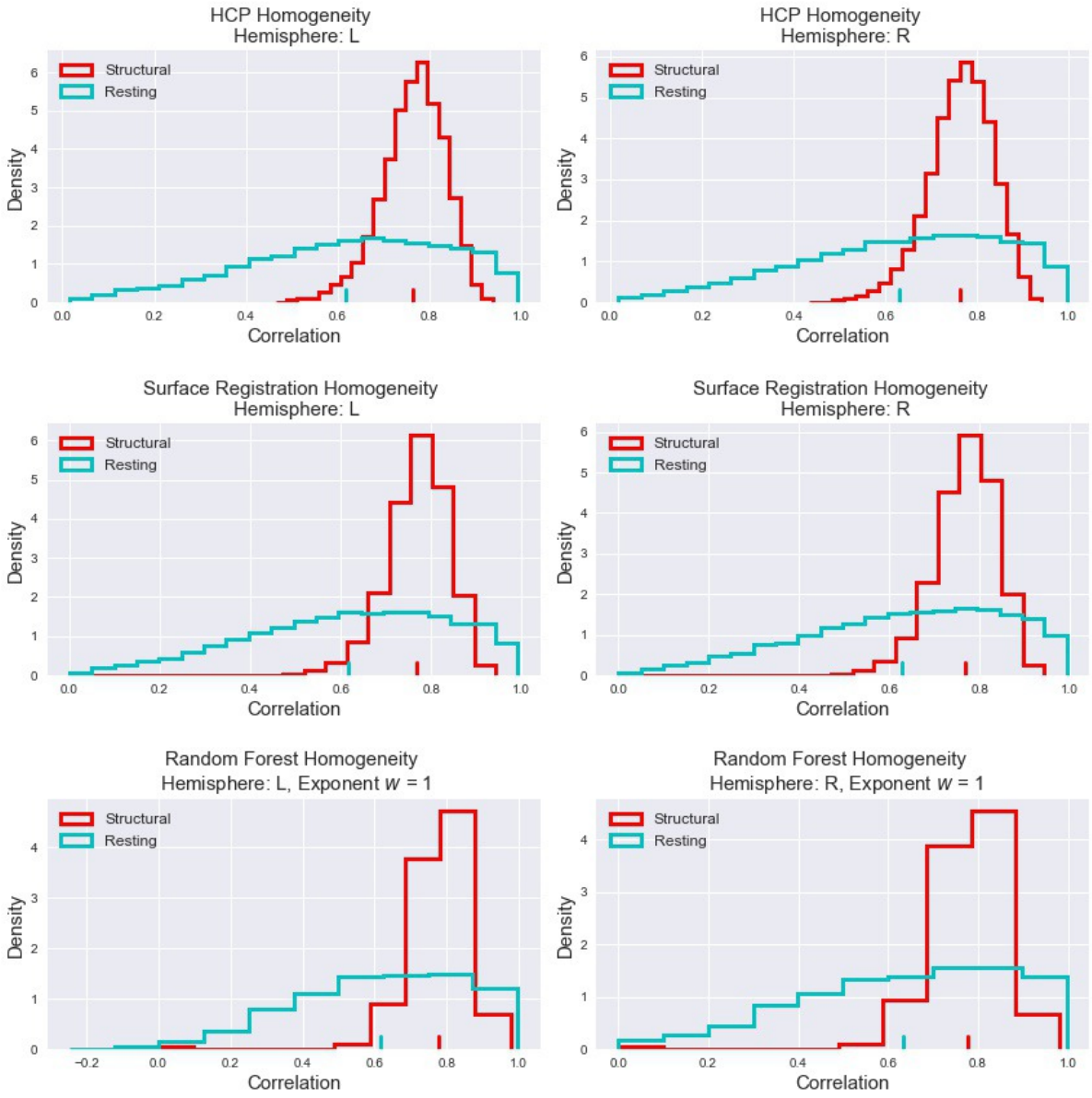

Figure 11. Empirical homogeneity distributions for the ground truth maps, surface registration maps, and single set of Random Forest-derived maps.

As a final comparison of predicted cortical parcellations in relation to the ground truth maps, we also compute the Dice coefficient of each predicted cortical map with the ground truth cortical map to analyze the spatial overlap of the predicted labels with the ground truth HCP maps. We report these results in **Figure 12** below. In correspondence with our previous results, we see that a weight exponent $w = 0$ generates cortical maps with less spatial overlap with the true

map than a weight exponent $w = 1$. The Random Forest models have the highest spatial overlap with the true map, when compared to both the Gaussian Mixture models and the Neural Network models. However, cortical maps generated using only the surface registration results produce larger Dice coefficient estimates than any of the model-based maps.

With regards to the training data types, there is no apparent pattern of increasing or decreasing spatial overlap. However, as with the accuracy measures in **Figure 5**, we see that there is in fact a trend towards larger Dice coefficients for the right hemisphere, when compared to the Dice coefficients for the left hemisphere

Model Dice Coefficient					
w = 0		Random Forest	GMM	Neural Network	
L	Struct.	0.644	0.471	0.643	
	Rest.	0.587	0.534	0.570	
	Comb.	0.639	0.507	0.632	
R	Struct.	0.656	0.498	0.656	
	Rest.	0.596	0.557	0.583	
	Comb.	0.646	0.532	0.650	
w = 1		Random Forest	GMM	Neural Network	Surface Registration
L	Struct.	0.780	0.654	0.702	
	Rest.	0.767	0.790	0.681	0.792
	Comb.	0.782	0.777	0.701	
R	Struct.	0.789	0.657	0.715	
	Rest.	0.772	0.797	0.686	0.800
	Comb.	0.789	0.783	0.715	

Figure 12. Dice coefficient for model-based predicted cortical maps and for surface-registration based cortical maps, stratified by hemisphere.

5. Discussion

In this analysis, we developed a framework to apply existing sets of cortical maps to a new subject's data, driven only by local surface geometry and resting state and structural connectivity features. We tested this framework using 3 classification models, 3 types of training data, and 2 prediction schemes, and assessed the performance of these models using classification accuracy and metrics to characterize the intra-regional feature similarity and regional spatial overlap of predicted cortical maps.

Traditional methods to apply cortical maps to new data typically rely on some form of volumetric registration to spatially normalize the cortices between a template brain and a target brain [9, 10, 18]. However, these methods often suffer from complications and inaccuracies related to voxel-wise interpolation. Additionally, they do not take into account the unique connectivity information related to each point they are trying to label. Our framework does not suffer from either of these drawbacks. While our method does incorporate a surface registration, the results of this registration act primarily as a prior on the possible labels to assign to a given surface point. We are then able to incorporate the connectivity data in a pseudo-Bayesian approach to predict the cortical map of a new test subject.

As it stands, however, our method suffers from a few drawbacks. Referring back to **Figure 2** showing the similarity of feature vectors as a function of geodesic distance, we see that the features we used to train our models, and specifically the regionalized resting state data, is highly variable, with standard deviations close to 0.2, even at short distances. Whether this high variability in feature data is a result of the dimensionality reduction scheme we applied or a manifestation of noise inherent to the original data is unclear, and would be worth examining more closely. Specifically, if we are able generate cleaner data, or perhaps represent our data in a more accurate lower-dimensional embedding, it might be the case that our classification scheme outperforms the surface registration.

Additionally, in considering the classification accuracy and regional homogeneity of our predicted maps, we found that there was a trend towards greater accuracy and regional homogeneity in the right hemisphere. This trend was also present in the Desikan-Killiany and Destrieux atlas maps that were used as controls. While we have not yet studied the possible cause of these right-hemisphere asymmetries, one possible explanation might be related to the left hemisphere language specialization [19]. The hemispheric asymmetry related to language might be correlated with increased variability of the connectivity profiles in the left hemisphere, and therefore could be a driver of the accuracy and homogeneity discrepancies that we saw in our results.

It is important to note from our results that the surface registration produced the most accurate predictions of the cortical maps, as well regional similarity measures that were most similar to the ground truth estimates. The weighted Random Forest and weighted Gaussian Mixture models classification models, however, performed only slightly worse than the surface registration. Of particular interest is the fact that the Gaussian Mixture Model performed most similar to the surface registration, in terms of pure accuracy. While we made the assumption of multivariate normality, the feature vectors themselves are also quite noisy, as evidenced by error bars displayed in **Figure 2**. It would be worth exploring whether or not a different mixture model, such as a multivariate Student's-T mixture model, which would be more robust to noise, performs better than a mixture of Gaussians. Likewise, we can also explore more complex neural network architectures. Specifically, it has been shown that there are certain network architectures that perform better on biological data [20] – these architectures might be particularly applicable to cortical mapping. Nevertheless, we have shown that, by using a general framework of training a classification model on connectivity data and using this model to predict the cortical map of new images, we are able to generate cortical parcellations for new data that are reasonably accurate and have similar regional connectivity profiles to the original maps.

References

1. Felleman, D.J. and D.C. Van Essen, *Distributed hierarchical processing in the primate cerebral cortex.* Cereb Cortex, 1991. **1**(1): p. 1-47.
2. Nieuwenhuys, R., *The myeloarchitectonic studies on the human cerebral cortex of the Vogt-Vogt school, and their significance for the interpretation of functional neuroimaging data.* Brain Struct Funct, 2013. **218**(2): p. 303-52.
3. Van Essen, D.C., et al., *Modular and hierarchical organization of extrastriate visual cortex in the macaque monkey.* Cold Spring Harb Symp Quant Biol, 1990. **55**: p. 679-96.
4. Caspers, S., et al., *Organization of the human inferior parietal lobule based on receptor architectonics.* Cereb Cortex, 2013. **23**(3): p. 615-28.
5. Brodmann, K. and L.J. Gary, *Brodmann's localisation in the cerebral cortex : the principles of comparative localisation in the cerebral cortex based on cytoarchitectonics*. 2006, New York, NY: Springer. xv, 298 p.
6. Gordon, E.M., et al., *Generation and Evaluation of a Cortical Area Parcellation from Resting-State Correlations.* Cereb Cortex, 2016. **26**(1): p. 288-303.
7. Glasser, M.F., et al., *A multi-modal parcellation of human cerebral cortex.* Nature, 2016. **536**(7615): p. 171-8.
8. Glasser, M.F., et al., *The minimal preprocessing pipelines for the Human Connectome Project.* NeuroImage, 2013. **80**: p. 105-124.
9. Fischl, B. and A.M. Dale, *Measuring the thickness of the human cerebral cortex from magnetic resonance images.* Proc Natl Acad Sci U S A, 2000. **97**: p. 11050-5.
10. Fischl, B., et al., *Automatically parcellating the human cerebral cortex.* Cereb Cortex, 2004. **14**: p. 11-22.
11. Behrens, T.E., et al., *Characterization and propagation of uncertainty in diffusion-weighted MR imaging.* Magn Reson Med, 2003. **50**(5): p. 1077-88.
12. Jbabdi, S., M.W. Woolrich, and T.E. Behrens, *Multiple-subjects connectivity-based parcellation using hierarchical Dirichlet process mixture models.* Neuroimage, 2009. **44**(2): p. 373-84.
13. Lombaert, H., et al., *FOCUSR: feature oriented correspondence using spectral regularization--a method for precise surface matching.* IEEE Trans Pattern Anal Mach Intell, 2013. **35**(9): p. 2143-60.
14. Destrieux, C., et al., *A sulcal depth-based anatomical parcellation of the cerebral cortex.* Human Brain Mapping, 2009.
15. Desikan, R.S., et al., *An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest.* Neuroimage, 2006. **31**: p. 968-80.

16. Yeo, B.T., et al., *The organization of the human cerebral cortex estimated by intrinsic functional connectivity.* J Neurophysiol, 2011. **106**: p. 1125-65.
17. Chollet, F.a.o. *Keras*. 2015.
18. Avants, B.B., et al., *A reproducible evaluation of ANTs similarity metric performance in brain image registration.* Neuroimage, 2011. **54**(3): p. 2033-44.
19. Morillon, B., et al., *Neurophysiological origin of human brain asymmetry for speech and language.* Proc Natl Acad Sci U S A, 2010. **107**(43): p. 18688-93.
20. Ronnenberger, O., P. Fischer, and T. TBrox, *U-Net: Convolutional Networks for Biomedical Image Segmentation.* Medical Image Computing and Computer-Assisted Intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention, 2015: p. 234-241.