

**HHS PUBLIC ACCESS**

Author manuscript

*Hum Brain Mapp.* Author manuscript; available in PMC 2017 November 01.

Published in final edited form as:

*Hum Brain Mapp.* 2016 November ; 37(11): 4129–4147. doi:10.1002/hbm.23301.

## Learning-based Subject-specific Estimation of Dynamic Maps of Cortical Morphology at Missing Time Points in Longitudinal Infant Studies

Yu Meng<sup>1,2</sup>, Gang Li<sup>2,\*</sup>, Yaozong Gao<sup>1,2</sup>, Weili Lin<sup>2</sup>, and Dinggang Shen<sup>2,3,\*</sup><sup>1</sup>Department of Computer Science, University of North Carolina at Chapel Hill, NC, USA<sup>2</sup>Department of Radiology and BRIC, University of North Carolina at Chapel Hill, NC, USA<sup>3</sup>Department of Brain and Cognitive Engineering, Korea University, Seoul, Republic of Korea

### Abstract

Longitudinal neuroimaging analysis of the dynamic brain development in infants has received increasing attention recently. Many studies expect a complete longitudinal dataset in order to accurately chart the brain developmental trajectories. However, in practice, a large portion of subjects in longitudinal studies often have missing data at certain time points, due to various reasons such as the absence of scan or poor image quality. To make better use of these incomplete longitudinal data, in this paper, we propose a novel machine learning-based method to estimate the subject-specific, vertex-wise cortical morphological attributes at the missing time points in longitudinal infant studies. Specifically, we develop a customized regression forest, named Dynamically-Assembled Regression Forest (DARF), as the core regression tool. DARF ensures the spatial smoothness of the estimated maps for vertex-wise cortical morphological attributes and also greatly reduces the computational cost. By employing a *pairwise estimation* followed by a *joint refinement*, our method is able to fully exploit the available information from both subjects with complete scans and subjects with missing scans for estimation of the missing cortical attribute maps. The proposed method has been applied to estimating the dynamic cortical thickness maps at missing time points in an incomplete longitudinal infant dataset, which includes 31 healthy infant subjects, each having up to 5 time points in the first postnatal year. The experimental results indicate that our proposed framework can accurately estimate the subject-specific vertex-wise cortical thickness maps at missing time points, with the average error less than 0.23 mm.

### Keywords

Missing data completion; longitudinal study; cortical surface; cortical thickness; early brain development

---

\*Corresponding authors: [dgshen@med.unc.edu](mailto:dgshen@med.unc.edu) or [gang\\_li@med.unc.edu](mailto:gang_li@med.unc.edu).

## 1. Introduction

In recent years, with the advance of pediatric MR imaging acquisition and post-processing techniques, longitudinal neuroimaging analysis of the early postnatal brain development, which can uniquely capture both subject-specific and population-averaged developmental trajectories, has received increasing attention (Almli, et al., 2007; Fan, et al., 2011; Gilmore, et al., 2012; Kaukola, et al., 2009; Li, et al., 2013; Li, et al., 2014b; Meng, et al., 2014; Nie, et al., 2014; Schumann, et al., 2010; Shi, et al., 2010). During the early postnatal stage, the human cerebral cortex develops extremely dynamically, with the surface area increasing 80% and the cortical thickness increasing 40% in the first postnatal year (Li, et al., 2013; Lyall, et al., 2014). Charting longitudinal cortical developmental trajectories in this early postnatal period can help better understand the relationship between normal structural and functional development of the cerebral cortex (Awate, et al., 2010; Dubois, et al., 2008; Gilmore, et al., 2007; Schnack, et al., 2014), and also provide fundamental references for understanding of many neurodevelopmental disorders that are likely the outcomes of abnormal brain development during this period (Gilmore, et al., 2012; Lyall, et al., 2014).

To accurately study the dynamic early brain development, many studies expect using subjects with complete longitudinal scans. However, in practice, missing data at certain time points are unavoidable in longitudinal studies due to various reasons, such as subject's absence from the scheduled scan or poor imaging quality of the scan. On one hand, directly using the incomplete data would introduce biases and consequently reduce precision and power in statistical analysis. For example, when constructing spatial-temporal infant cortical surface atlases (Li, et al., 2014c), each time point will have a different number of subjects because of missing data. As a result, different biases could be introduced to different time points, thus leading to low accuracy and longitudinal inconsistency for subsequent analysis, especially for those time points with limited number of scans. Another example is that, in longitudinal study of the spatial distribution of deep sulcal landmarks on infant cortical surfaces (Meng, et al., 2014), each involved subject needs to be scanned at all time points for consistent and unbiased comparison across ages. Due to missing data, different subjects are scanned at different time points, thus leading to biased and inconsistent comparison. On the other hand, discarding subjects with missing time point(s) is a terrible waste of the useful information in these subjects and also the considerable cost for data acquisition. Hence, accurate estimation of information at missing time points plays an important role in longitudinal analysis of early brain development.

A common strategy to handle missing data is to replace the missing data by the weighted average over the most representative subset of existing data (e.g., k-nearest neighbors) (Ching, et al., 2010; Troyanskaya, et al., 2001; Tsiporkova and Boeva, 2007), but the effectiveness of this strategy often decreases with the increase of the portion of missing data. For recovering a large portion of missing data, several methods based on low-rank matrix completion were proposed (Cai, et al., 2010; Cand and Tao, 2010; Candès and Recht, 2009; Liu, et al., 2013), but these methods work well only if the missing data are distributed randomly and uniformly. Unfortunately, in neuroimaging study, the missing data are usually distributed in blocks and not uniformly (Li, et al., 2014f; Thung, et al., 2014; Yuan, et al., 2012). For example, the missing data can be the entire image, rather than some independent

pixels or clinical scores, thus the methods based on low-rank matrix completion can no longer work properly. Recently, multi-source feature learning methods (Yuan, et al., 2012) and deep convolutional neural networks (Li, et al., 2014f) were proposed to estimate information of missing imaging modalities based on the available modalities, and the estimated modalities were demonstrated to help diagnose neurodegenerative diseases. To deal with the missing modalities in the application of disease diagnosis, another alternative method is to first select the most discriminative features extracted from the available modalities and then estimate only those most discriminative features in the missing modalities (Thung, et al., 2014). However, the estimation of missing vertex-wise cortical morphological maps in this paper has specific characteristics. First, many longitudinal analysis needs information of morphological measurements at every vertex on the cortical surface, rather than just a simple mean value in each region. For example, to extract deep sulcal landmarks from the infant cortex (Meng, et al., 2014), the whole cortical morphological map containing the depth values for all vertices of the cortical surface is needed. Second, cortical surfaces are usually represented by triangular meshes, rather than by 3D volumes. Thus, the ways used to extract features as employed in the image-based methods (Li, et al., 2014f; Yuan, et al., 2012) are not naturally extendable to extraction of features from cortical surfaces. Therefore, the existing image-based methods (Li, et al., 2014f; Thung, et al., 2014; Yuan, et al., 2012) cannot be directly applied to our task of estimating missing vertex-wise morphological maps on the dynamic developing cortical surfaces.

In this paper, we propose a novel learning-based framework for subject-specific estimation of the vertex-wise map for cortical morphological attributes at missing time point(s) in the longitudinal infant brain studies. Of note, this is challenged by the extremely dynamic and regionally-heterogeneous growth of the infant cortex and also by the considerable inter-subject variability of cortical morphology and developmental patterns. Technically, we leverage the regression forest (Breiman, 2001) as our core regression tool to estimate cortical morphological attributes at each vertex of the cortical surface. However, using a *single* conventional regression forest (CRF) to estimate vertex-wise cortical morphological attributes of the entire surface is not suitable. Because the cortical morphological attributes and their developments in infants are highly regionally heterogeneous, using a single CRF cannot precisely estimate values at vertex level. An intuitive way to solve this problem is to partition the whole cortical surface into a set of small regions of interest (ROIs), and then train a local regression forest for each ROI. However, this will lead to spatially unsmooth estimation results around the boundaries of neighboring ROIs. This is because cortical attributes of vertices in the two sides of a ROI boundary are estimated using two different regression forests that are trained independently with different training samples. Intuitively, increasing the overlapping area among ROIs or training a complete regression forest at each vertex could produce smoother estimation results, but it unexpectedly increases the computational cost. Based on our experiments, to make the estimation results as smooth as the real data, more than 90% area of a ROI needs to be overlapped with its neighboring ROIs. Unfortunately, such a large portion of overlap requires a large number of ROIs in order to cover the whole cortex, and thus leads to huge computational workload, since a respective set of individual trees need to be trained for each ROI. Taking account all these

issues, we propose a novel Dynamically-Assembled Regression Forest (DARF). Specifically, by first training one decision tree for each vertex in the training stage and then locally grouping decision trees of neighboring vertices as forests in the testing stage, DARF is able to produce the spatially smooth regression results and *meanwhile* also save a lot of computational cost.

In general, for estimating the vertex-wise cortical attributes maps at missing time point(s), our method includes two major stages. In the first stage, in order to use as many training subjects as possible, the missing data (e.g., vertex-wise cortical attributes maps) at each missing time point of each subject is estimated *multiple times* with each time based on the available data at each existing time point *independently*, and then these estimated results at each missing time of each subject are averaged as an initial estimation. In the second stage, to better use longitudinal information, the missing data at each time point of each subject is refined based on both real data and the initially-estimated missing data at all other time points *jointly*. We have validated our missing data estimation method on an incomplete longitudinal dataset of infants for the estimation of vertex-wise map of cortical thickness, which is an important morphological measure for the cerebral cortex and also correlates with both normal development and neurodevelopmental disorders (Lyll, et al., 2014). The experimental results show that our method achieves high accurate estimations, with the average vertex-wise error of less than 0.23 mm. It is worth mentioning that, among all the cortical attributes, cortical thickness is relatively more difficult to estimate, as it is quite variable in terms of both spatial distribution and dynamic longitudinal development across individuals. Of note, our method is very generic and not limited to only estimating cortical thickness, as it can also be extended to estimate other morphological attributes, such as sulcal depth, surface area, cortical folding, and local cortical gyrification (Li, et al., 2014e).

## 2 Dataset and Image Processing

### 2.1 Subjects and MR Image Acquisition

This study was approved by the Institutional Review Board of the University of North Carolina (UNC) School of Medicine. The UNC hospitals recruited healthy pregnant mothers during their second trimesters of pregnancy. There was no abnormal fetal ultrasound, congenital anomaly, metabolic disease or focal lesion in the infants in the study cohort. For each infant, informed consents were obtained from both parents. All infants were scanned during natural sleep with no sedation used. During each scan, the heart rate and oxygen saturation of the infant were monitored by a physician or a nurse using a pulse oximeter.

For each infant, MRI scans were scheduled at every 3 months in the first year of life, every 6 months in the second year of life, and every 12 months in the third year of life. At each scheduled scan, T1-, T2-, and diffusion-weighted MR images were acquired by a Siemens 3T head-only MR scanner with a 32 channel head coil. T1-weighted images (144 sagittal slices) were acquired with the imaging parameters: TR = 1900 ms, TE = 4.38 ms, flip angle = 7, acquisition matrix =  $256 \times 192$ , and voxel resolution =  $1 \times 1 \times 1 \text{ mm}^3$ . T2-weighted images (64 axial slices) were acquired with the imaging parameters: TR/TE = 7380/119 ms, flip angle = 150, acquisition matrix =  $256 \times 128$ , and voxel resolution =  $1.25 \times 1.25 \times 1.95 \text{ mm}^3$ . Diffusion-weighted images (DWI) (60 axial slices) were acquired with the parameters:

TR/TE = 7680/82 ms, acquisition matrix =  $128 \times 96$ , voxel resolution =  $2 \times 2 \times 2 \text{ mm}^3$ , 42 non-collinear diffusion gradients, and diffusion weighting  $b = 1000 \text{ s/mm}^2$ . More information on image acquisition can be found in (Li, et al., 2014a; Nie, et al., 2012; Wang, et al., 2012).

The whole longitudinal dataset is illustrated in Figure 1, where each black block denotes the missing data at a respective time point for a certain subject. It can be seen that this dataset contained a large portion ( $\approx 25.7\%$ ) of missing data. In this paper, only a subset (enclosed by the red rectangle in Figure 1) of all the data is used, because we need a sufficient amount of real data as the ground truth to evaluate our proposed method. In the selected subset, 31 healthy infants were scanned at 5 time points, i.e., 1, 3, 6, 9, and 12 months of age; 16 infants have complete 5 scans, 10 infants have one missing scan, and 5 infants have two missing scans.

## 2.2 Cortical Surface Reconstruction and Registration

All infant MR images were processed by using an infant-specific computational pipeline for cortical surface based analysis, which has been extensively verified on over 500 infant MRI scans (Li, et al., 2013; Li, et al., 2015b; Li, et al., 2014c; Lyall, et al., 2014). Of note, our infant-specific computational pipeline was inspired by many existing work for cortical surface analysis of adults (Dale, et al., 1999; Fischl and Dale, 2000; Fischl, et al., 1999b; Han, et al., 2004; Kim, et al., 2005; Liu, et al., 2008; MacDonald, et al., 2000; Mangin, et al., 2004; Shattuck and Leahy, 2002; Shi, et al., 2013; Thompson and Toga, 1996; Van Essen, et al., 2001). First, distortion correction of DWI has been performed and the respective fractional anisotropy (FA) images were then computed. T2 images were first linearly aligned to their respective T1 images using normalized mutual information. Then, FA images were linearly aligned to the warped T2 images. To ensure the quality of alignment, the aligned T1, T2, and FA images were visually inspected. Second, for each set of aligned T1, T2, and FA images, skull stripping was performed using a learning-based method (Shi, et al., 2012), and then brain stem and cerebellum were removed by propagation of their respective masks from the atlas images to the subject images using an in-house developed registration method. Third, intensity inhomogeneity was corrected by N3 (Sled, et al., 1998). Fourth, all longitudinal images of the same infant were rigidly aligned. Fifth, an infant-specific 4D level-set method was used to segment brain tissues by leveraging longitudinal information and complementary multimodal information in T1, T2 and FA images (Wang, et al., 2014a; Wang, et al., 2014b). Finally, non-cortical structures were filled, and each brain was separated into left and right hemispheres.

For cortical surface reconstruction, we used a deformable surface method to reconstruct the topologically correct and geometrically accurate cortical surfaces for each hemisphere (Li, et al., 2012). Specifically, topology correction was first performed for the white matter of each hemisphere. Then, the topology-corrected white matter was tessellated as a triangulated surface mesh. Finally, a deformable surface method was applied to deform the shape of triangulated surface mesh while keeping its initial topology and also spatially-adaptive smoothness, to reconstruct the inner and outer cortical surfaces (Li, et al., 2014a). To prevent surface meshes from self-intersection, in each step of the surface deformation, a fast

triangle–triangle intersection detection method was also performed at each vertex in the local region. Once any triangle–triangle intersection was detected, the deformation was reduced to a location without such intersection (Li, et al., 2014a). Cortical thickness of each vertex was then computed as the mean of the minimum distances from inner to outer surfaces and also from outer to inner surfaces (Li, et al., 2015a; Li, et al., 2014a) as in FreeSurfer (Fischl, 2012). The sulcal depth of each vertex was defined as the shortest distance from the vertex to the cerebral hull surface, and was computed using the method in (Li, et al., 2014b).

For cortical surface registration, all inner cortical surfaces were first mapped onto a spherical surface by minimizing the metric distortion (Fischl, et al., 1999a) using FreeSurfer (Fischl, 2012). The *intra-subject* registration was then performed to unbiasedly align all longitudinal cortical surfaces of the same infant together, using a group-wise surface registration method, namely Spherical Demons (Yeo, et al., 2010). The *inter-subject* registration was then performed by group-wise registration of the mean cortical surfaces of all different infants, using Spherical Demons again. Finally, each cortical surface was resampled to the same mesh tessellation on the spherical space based on the registration results, and thus the vertex-wise correspondences of all cortical surfaces of different infants were established (Li, et al., 2014d). More details on both intra-subject and inter-subject surface registrations can be found in (Li, et al., 2014c).

### 3 Methods

In this section, we first introduce our regression model, named Dynamically-Assembled Random Forest (DARF), and then describe how to use this method for subject-specific estimation of missing vertex-wise cortical morphological attributes at missing time point(s) for an incomplete longitudinal infant dataset.

#### 3.1 Regression Model

**3.1.1 Training and Testing a Decision Tree**—In the training stage, given a set of training samples  $\{(\mathbf{x}_i, y_i) | \mathbf{x}_i \in \mathbb{R}^d, y_i \in \mathbb{R}\}$ , where  $\mathbf{x}_i$  and  $y_i$  are respectively the  $d$ -dimensional feature vector and the scalar target value of the  $i$ -th training sample, each binary decision tree is independently trained by recursively finding a series of optimal partitions of the training samples. Specifically, at the root node, the training samples are optimally partitioned into two subsets by maximizing the following objective function:

$$\arg \max_{k, \theta} H(S) - \sum \frac{|S_l|}{|S|} H(S_l) - \sum \frac{|S_r|}{|S|} H(S_r) \quad (1)$$

where  $S$  is the set of training samples at the current node;  $S_l$  and  $S_r$  are respectively the subsets of  $S$  in the left child node and the right child node after partition;  $H$  is a metric that estimates the consistency of training samples in terms of regression target. Mathematically,  $H$  is defined as:



$$H(S) = \text{trace}(\text{cov}(\{y_i, i \in S\})) \quad (2)$$

where *cov* is the covariance matrix, and *trace* is the matrix trace. The partition is determined by two factors *k* and  $\theta$ . For the *i*-th sample in the training set, if the *k*-th feature  $x_i^k$  in the feature vector  $x_i$  is less than the threshold  $\theta$ , the sample is dispatched to the left child node; otherwise, it is dispatched to the right child node. To maximize the objective function Eq. 1, all dimensions of the feature vector are tested one after another with a certain number of thresholds, which are selected randomly between the minimum and maximum feature values of training samples, and the pair (*k* and  $\theta$ ) with the largest objective value is selected as the optimal parameters and stored in the root node. The partition continues recursively for the subset of training samples in the left and right child nodes, until reaching any of the following two terminal criteria: 1) the tree reaches the specified maximum depth, and 2) the amount of training samples in a node falls below the specified minimum number. For each leaf node, where the partition stops, the target values of all training samples falling into this leaf node are averaged as the regression result and stored in this leaf node.

In the testing stage, for each individual decision tree, the testing sample goes from the root node to a leaf node according to the results of binary tests in the non-leaf nodes, and the output is the regression result stored in the leaf node.

**3.1.2 Dynamically-Assembled Regression Forest (DARF)**—In the *training* stage, an individual binary decision tree  $Tree_p$  is independently trained at each vertex *p* on the resampled spherical surface. Specifically, as shown in Figure 2(a), for a given vertex, all its nearby vertices in a specified neighborhood (the red region) are used together as training samples. For each training sample, we have a feature vector  $x_j \in \mathbb{R}^d$  and a scalar target  $y_j \in \mathbb{R}$ . The feature vector  $x_j$  consists of a set of features (see **Section 3.1.3** below) extracted from the local cortical attribute maps at the input time points, and the scalar target  $y_j$  is the attribute value at the target time point. In the *testing* stage, to estimate the cortical attribute at a given vertex *p*, as shown in Figure 2(b), all trained nearby individual trees in a specified neighborhood (i.e., the green region) are grouped together to form a forest  $F_p = \{Tree_j \mid \|i - p\| < D_t\}$ , where  $D_t$  is a specified threshold. Then, features of the given vertex is computed and fed to each tree in the forest to estimate the cortical attribute at the corresponding vertex of target time point. Finally, the regression results from all the trees in the forest are averaged as the final estimation.

**3.1.3 Feature Computation on Spherical Surface**—For each vertex *i*, its feature vector  $x_i$  includes two types of features: local features and context features. Local features provide local information at each vertex, while context features provide rich neighboring information. In our current implementation for estimating cortical thickness at missing time point(s), the local features include cortical thickness and sulcal depth, and the context features include a number of Haar-like features of cortical thickness and sulcal depth. Herein, Haar-like features provide two types of context information: (1) the mean attribute of a small cortical region, and (2) the difference between mean attributes of two small regions. Next, we introduce how to compute Haar-like features on a spherical surface.

As shown in Figure 3, given a vertex  $i$  on the resampled spherical surface, its neighbor is locally projected to the tangential plane, where a local 2D coordinate system is built at the center of vertex  $i$ . Two blocks  $A$  and  $B$  are *randomly* selected in the neighborhood  $[\pm u_\sigma, \pm v_\sigma]$ , and their sizes  $r_a$  and  $r_b$  are also chosen *randomly* in the interval  $(r_1, r_2]$ , where  $u_\sigma, v_\sigma, r_1$  and  $r_2$  are the user-defined parameters. Letting  $Q_A$  denote the set consisting of all the vertices in block  $A$  and also  $Q_B$  denote the set consisting of all the vertices in block  $B$ , then the Haar-like feature at vertex  $i$  can be mathematically formulated as:

$$f(i) = \frac{1}{|Q_A|} \sum_{(u,v) \in Q_A} M(u,v) - \delta \frac{1}{|Q_B|} \sum_{(u,v) \in Q_B} M(u,v) \quad (3)$$

where  $M(u, v)$  is the value of cortical morphological attribute (i.e., cortical thickness or sulcal depth) at position  $(u, v)$ ; and  $\delta$  is a random coefficient that can only be 0 or 1. In the case of  $\delta = 0$ , Haar-like feature is the mean value of cortical morphological attributes within the block  $A$ . In the case of  $\delta = 1$ , Haar-like feature is the difference between the mean values of cortical attributes within block  $A$  and block  $B$ .

### 3.2 Estimation of Missing Cortical Thickness Map

As shown in Figure 1, in our longitudinal dataset for early brain development study, many time points are missing due to various reasons. Intuitively, to maximize the capability of estimating missing data, we expect using as much available information as possible to train the regression model (i.e., DARF). Specifically, by increasing the number of subjects for training, the regression model can better learn the large diversity among individuals; and also by engaging more time points in training process, the regression model can better capture the longitudinal information of cortex development. Unfortunately, due to the data incompleteness, increasing the number of training subjects and engaging more time points are conflicting with each other. For example, as shown in Figure 1, to estimate the missing data at 6 months, we can use *at most* 28 subjects to train the regression model, since only 28 infants have real data at both 1 and 6 months. Consequently, only 1 time point (i.e., at 1-month-old) is taken account into the training process. In another way, we can engage *at most* 4 time points in the training process, but only 16 infants have real data at all 5 time-points and can be used as training subjects. To eliminate this confliction and fully use the available information, we propose a two-stage missing data completion method. Figure 4 shows the overview of the proposed method, containing the stages of 1) pairwise estimation and 2) joint refinement.

In Stage 1, to use as many subjects as possible for training, the data (i.e., cortical thickness, as introduced in **Section 2.2**) of each subject at each missing time point is estimated based on the existing data at each of other time points *independently*, and then these independent estimations are averaged together to obtain an initial estimation. For example, to obtain the initial estimations at 6 months, we will first use all the subjects with available data at both 1-month-old and 6-months-old as training subjects to train a set of decision trees, by taking the data at 6-months-old as the regression target while the data at 1-month-old as the inputs. After training, for the subjects with available data at 1-month-old but without data at 6-



months-old, these decision trees are dynamically/locally assembled as forests to estimate the missing data at 6-months-old. Similarly, we can also obtain the estimations of missing data at 6-months-old, respectively, using the existing data at each of the 3-, 9-, and 12-months-old. In this way, the available data at all other time points can contribute to the estimation of the data at 6-months-old. Finally, we average all the estimations contributed from different time points together as the initial estimation. Similarly, for the missing data at each of 1-, 3-, 9-, and 12-months-old, the same process can be performed to obtain their own initial estimations. After using Stage 1, the missing data of all subjects at all time points will be approximately recovered, thus providing a pseudo-complete longitudinal dataset.

In Stage 2, to take advantage of the longitudinal information and also to make the estimation temporally consistent, the data at each missing time point is further refined based on the data at all other time points *jointly*. For example, to obtain the final estimation of the missing data at 6-months-old, we use all subjects with real data at 6-months-old as the training subjects to train a set of decision trees, by taking the data at 6-months-old as the regression target and the data at 1-, 3-, 9-, and 12-months-old as the inputs. After training, for each subject with missing data at 6-months-old, the trained decision trees can be dynamically/locally assembled as forests to estimate the missing data. Note that we do not require each training/testing subject to have real data at 1-, 3-, 9-, and 12-months-old, since we already recovered them in Stage 1 for all subjects with missing data at any time point. Similarly, for the missing data at other time points, the same process can be conducted to obtain their final estimations. It is worth noting that, using the above two stages (Stage 1 and Stage 2), our method leverages information from all time points of all available training subjects for the missing data estimation.

### 3.2 Quantitative Evaluation

To quantitatively evaluate the estimation results, we employed three metrics: NMSE (normalized mean squared error) (Faramarzi, et al., 2013), MAE (mean absolute error), and MRE (mean relative error). These metrics are respectively computed as follows:

$$NMSE = \frac{\|\mathbf{T}_{Est} - \mathbf{T}_{True}\|_2^2}{\|\mathbf{T}_{True}\|_2^2} \quad (4)$$

$$MAE = \frac{|\mathbf{T}_{EST} - \mathbf{T}_{True}|}{N} \quad (5)$$

$$MRE = \frac{1}{N} \left| \frac{\mathbf{T}_{EST} - \mathbf{T}_{True}}{\mathbf{T}_{True}} \right| \quad (6)$$

where  $\mathbf{T}_{True}$  and  $\mathbf{T}_{Est}$  are respectively the vectors of ground truth and estimated results, and  $N$  is the number of vertices.

## 4 Results

To evaluate our proposed missing data estimation method, we tested the method by recovering cortical thickness maps at 5 missing time points. Specifically, from the incomplete dataset (Figure 1), we selected 16 subjects with complete data at all 5 time points as the reference subjects. Then, we manually selected one of these 16 subjects, deleted its data at a certain time point, and put it back to the dataset. The deleted data is then recovered using our missing data estimation method. After recovering the missing data, we compared the recovered result with the real data. This experiment was repeated for each of those 16 subjects at each of 5 time points.

Figure 5 shows the error map of our proposed method at each stage for estimation of missing cortical thickness at 9-months-old on a randomly selected infant. Figure 6 further shows the averaged errors for all reference subjects in each step of estimation. From these two figures, we can see that using the data at 6- or 12-months-old as inputs to estimate the cortical thickness at 9-months-old is better than using the data at 1- or 3-months-old. A possible explanation is that the cortical thickness at 9-months-old is more similar to that at 6- and 12-months-old than to that at 1- and 3-months-old. Figure 6 also shows that the result of joint refinement is generally better than all the results in the previous stage (Stage 1), indicating the effectiveness of joint refinement stage (Stage 2). Figure 7 illustrates the average estimation errors of vertex-wise cortical thickness at all 5 time points. We found that the estimation precision is region-specific, with high precision in the unimodal cortex while relatively low precision in the high-order association cortex. A possible explanation is that the unimodal cortex may have less variable cortical thickness patterns across individuals than the high-order association cortex during infancy. Hence, our method can better capture patterns of the unimodal cortex than those of the high-order association cortex, thus producing more accurate prediction in the unimodal cortex. Note that, among all time points, the estimations were relatively less accurate at around 6 months of age. The reason is that, during this development stage, the cortex develops exceptionally rapid (Li, et al., 2014a) and the image contrast is also extremely low. All of the above observations are further proved by the quantitative evaluation given in Tables 1, 2, and 3. From these tables, we can further conclude that our method is able to effectively recover the missing cortical thickness, with the average absolute error of less than 0.23 mm and the average relative error of less than 9.24%. Moreover, we also performed paired t-test to statistically compare between results of pairwise estimation and joint refinement. All p-values are much less than 0.01, indicating that the performance improvement by the joint refinement over the pairwise estimation (reported in Table 1, 2, and 3) is statistically significant. The estimation error by our method is about 10% of cortical thickness. Considering that the MRI resolution is 1mm and the average cortical thickness is around 2 mm in infants, our estimation error is just around 0.2 mm, which is much less than the resolution of a half voxel. Meanwhile, around 96.5%, 95.2%, 85.3%, 91.4%, and 92.0% vertices on the cortical surface have the absolute errors less than 0.5mm (half voxel) for 1<sup>st</sup>, 3<sup>rd</sup>, 6<sup>th</sup>, 9<sup>th</sup>, and 12<sup>th</sup> month, respectively. These results indicate that our prediction is accurate and acceptable. We further computed the estimation errors in each of 36 ROIs. Figure 8 shows both the mean absolute error and the mean relative error in each of 36 ROIs for estimation of the missing cortical thickness at 9-months-old. We

can see that, in all ROIs, the use of joint refinement led to obvious improvements over the use of only the pairwise estimation. Figure 9 shows error measures in each ROI for estimation the missing cortical thickness at all 5 time points. We can see that the joint refinement consistently improved the result of the pairwise estimation in some challenging ROIs, such as caudal anterior-cingulate cortex (ROI 3), cuneus cortex (ROI 6), lateral orbitofrontal cortex (ROI 13), middle temporal gyrus (ROI 16), pars orbitalis (ROI 20), pericalcarine cortex (ROI 22), posterior-cingulate cortex (ROI 24), and superior frontal gyrus (ROI 29).

To demonstrate the advantages of DARF, we compared it with other four representative methods, including mixed effect model (MEM) in FreeSurfer (Bernal-Rusiel, et al., 2013), polynomial regression (PR), conventional regression forest (CRF), and sparse linear regression (SLR). MEM, which explicitly models fixed effects and random effects, is a powerful method for analyzing longitudinal neuroimaging (Bernal-Rusiel, et al., 2013). In our comparison experiments, MEM assumes that the development of cortical thickness increases with age (fixed effect) during the first year, while each subject has individual variant due to personal reasons (random effect), such as genetic and environmental influences. PR method assumes that the development of cortical thickness at each vertex has a two-order polynomial relationship with age. CRF trains a single forest for the entire surface with the spherical location of each vertex as additional features (in addition to the Haar-like features). SLR is an effective method for high-dimensional data analysis (Tibshirani, 1996), which can extract the most “useful” features from a high-dimensional feature representation by setting zero coefficients for irrelevant features. Specifically, given a target vector  $\mathbf{y} = [y_1, y_2, \dots, y_n]^T \in \mathbb{R}^n$  and the feature matrix  $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] \in \mathbb{R}^{d \times n}$ , SLR method finds the optimal coefficients  $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \dots, \alpha_d]^T \in \mathbb{R}^d$  by solving Eq. 7 below, with the constraint that the number of non-zero elements in  $\boldsymbol{\alpha}$  is no more than  $L$ .

$$\arg \min_{\boldsymbol{\alpha} \in \mathbb{R}^d} \frac{1}{2} \|\mathbf{y} - \mathbf{X}^T \boldsymbol{\alpha}\|_2^2 + \lambda \|\boldsymbol{\alpha}\|_1 \quad (7)$$

where  $L$  and  $\lambda$  were optimally set to 12 and 0.001 respectively in our experiments based on a grid search, which was performed on a subset of the training data.

Figure 10 provides a comparison among MEM, PR, CRF, SLR and DARF for estimation of vertex-wise cortical thickness at 9-months-old for a representative subject. As we can see, DARF estimated cortical thickness map more precisely than other four methods. Figure 10 also shows that the error map of CRF is very spotty compared with the error maps of other methods, which indicates that the estimation result of CRF is not as smooth as the real cortical thickness map. Table 4 reports the complete quantitative evaluation for five methods based on leave-one-out cross-validation. It shows that DARF outperforms all other methods. An interesting observation is that PR performs much worse at the first and last time points, but it does relatively better at the intermediate time points. This means that when using a quadratic curve to fit the development trajectory of cortical thickness, it is relatively difficult to precisely estimate the two ends of the curve. We further performed paired t-test between

our method and all other methods, and obtained all p-values less than 0.001, demonstrating the significant advantage of our method.

## 5 Discussion

### Parameter Selection and Robustness

The performance of the proposed method can be affected by several parameters, e.g., maximal depth of trees and the neighborhood size for computing Haar-like features. We chose these parameters by searching their optimal values based on leave-one-out cross-validation on the training data, which could effectively prevent DARF from over-fitting the training data. Specifically, to search the optimal value for the maximal depth of trees, we fixed all other parameters, and then tried all possible values from 10 to 120 with each 10 increment. As shown in Figure 11(a), when the maximal depth is less than 60, the accuracy is improved with increase of the maximal depth; but when the maximal depth is larger than 60, the accuracy is relatively steady. Therefore, we chose 70 as the maximal depth of trees. Figure 11(b) shows the relationship between the minimal sample number in each leaf node and the mean absolute error. It can be seen that the mean absolute error fluctuates slightly when the number is less than 8; however, when the number is larger than 10, the error starts increasing. This is because increasing samples in each leaf node could make more training samples with different target values fall into the same leaf node and thus lead to more rough regression results. Accordingly, we chose 3 as the minimal sample number in each leaf node in all our experiments. For the Haar-like features, four parameters  $u_\sigma$ ,  $v_\sigma$ ,  $r_1$ , and  $r_2$ , as shown in Figure 3, controls their computations. In our experiments, we simply set  $u_\sigma = v_\sigma$ , as there is no specific reason to treat them differently. Assuming the radius of the spherical surface is 100, when  $u_\sigma$  and  $v_\sigma$  are in (7,13), the highest accuracy can be achieved (Figure 11(c)). The main reason is that, when  $u_\sigma$  and  $v_\sigma$  are too small, less neighboring information could be effectively encoded by the Haar-like features; however, when  $u_\sigma$  and  $v_\sigma$  are too large, Haar-like features might fail to precisely describe local information. Thus, we chose 10 for  $u_\sigma$  and  $v_\sigma$ . For  $r_1$ , which is the lower bound for the size of a Haar-like feature block, we simply set it 0, which means that the minimal region described by a Haar-like feature could be of the size of a single vertex. For  $r_2$ , which is the upper bound for the size of a Haar-like feature block, we tested the value from 0.3 to 12.6. As shown in Figure 11(d), the best accuracy can be achieved when  $r_2$  is in the range of (0.9,2.5). When  $r_2$  is larger than 3.1, the estimation error increases with the increase of  $r_2$ . In our experiments, we chose 1.6 for  $r_2$ . Figure 11(e) shows how the estimation error changes with the number of Haar-like features. It can be seen that, when the number of Haar-like features is larger than 500, further increasing does not benefit much. In fact, increasing the number of Haar-like features will lead to high computational cost. Thus, we chose 500 as the number of Haar-like features in all our experiments. In the training process, to maximize the objective function Eq. 1, all dimensions of the feature vector were tested one after another. For each dimension, a certain number of thresholds was selected randomly between the minimum and maximum feature values of all training samples, as suggested in (Criminisi, et al., 2012). We tested different sample numbers for selecting threshold  $\theta$ . As shown in Figure 11(f), increasing the number of sampled thresholds could improve the accuracy. But, when the sample number is larger than 10, the accuracy could no longer be improved if further increasing the sample number.

Moreover, larger sample number could also cause high computational overhead. Thus, we chose 20 in our experiments. In summary, we can see from Figure 11 that all the parameters are selected in a relatively stable range, which means that the slight adjustment of parameters will not affect the estimation accuracy, indicating robustness of our method to the parameter configuration.

### Smoothness

DARF can obtain very smooth estimations because of the following two reasons. First, DARFs of neighboring vertices are very similar, as they share a large number of same decision trees. Second, features of neighboring vertices are also similar. Thus, by feeding the similar DARFs with the similar input features, the outputs at neighboring vertices generally have small differences, and thus the estimated cortical attributes for the whole surface are smooth.

### Computational Cost

To achieve the similar smooth estimation results, DARF significantly saves the computational cost, compared to CRF with highly-overlapped ROIs. For example, if the number of ROIs is  $N_{ROI}$ , the total number of trees we need to train for CRF is  $N_{ROI} \times N_{Tree}$ , where  $N_{Tree}$  is the number of trees in each forest. Based on our experiments, a well-trained forest needs approximately 100 decision trees. Besides, in order to make the estimation as smooth as the real data,  $N_{ROI}$  needs to be close to  $N/5$ , where  $N$  is the number of vertices on the spherical surface. So the total number of decision trees for the whole surface is about  $N/5 \times 100 = 20N$ . If using DARF, the total number of trees to train is  $N$ , since we need to train only one decision tree at each vertex. As a result, DARF reduces the computational cost by approximately 20 times.

### Performance on Large Portion of Missing Data

We randomly removed some existing data, and tested the proposed method with different portions of missing data. Of note, the original missing data in our dataset (enclosed by the red rectangle in Figure 1) is 12%. By randomly removing some existing data the entire dataset, we can get the datasets with 13%, 20%, 40%, and 60% missing data, respectively. From Figure 12, we can see that the estimation errors increase with the portion of missing data increasing. However, even the missing data reaches 40%, the proposed method is still able to produce the average estimation error to less than 0.23 mm. Note that, like most machine learning methods, the estimation precision of DARF depends on the quantity and quality of training data. As long as we have enough quality training data, DARF is able to estimate the missing data precisely. Figure 12 also shows that joint refinement consistently improves the result of pairwise estimation.

### Estimation of Abnormal Cortical Attribute Maps

Our current study focuses on the normal early brain development. Note that the abnormal cortical attribute map could have much larger variability across subjects/diseases than the normal ones, thus needing larger-scale datasets to train our estimation model for effectively covering all the possibilities. If training DARF by the normal subjects and then directly

using it to estimate the abnormal cortical attributes, the estimation results may not reflect the true abnormal development. However, this can indirectly benefit the detection of abnormal brain development. For example, given a new subject, we can first estimate its cortical attribute maps using DARF, which has been trained by the normal subjects. Then, the estimated maps can be compared with the real maps. If their difference is significantly larger than the estimation errors reported in Table 2, this may indicate an abnormal brain development in this subject. Moreover, for a certain disease, if DARF could be trained by subjects with this disease, the estimation accuracy would likely be better than the results estimated by DARF that is trained with a general dataset.

### Potential Applications

After recovering the missing cortical attribute maps in a longitudinal dataset by our method, the “completed” dataset could be used for studying early brain development. For example, the recovered dataset can be used to better build longitudinal cortical surface atlases. Note that, before recovering the missing data, building such atlases could introduce different biases for different time points since the atlas at each time point is constructed by a different number of subjects. While, after recovering the missing data, the atlases at all time points can be constructed based on the same subjects, thus introducing less bias and leading to more accurate and longitudinally-consistent atlases. Another application example is that the recovered dataset can be used to carry out less biased and longitudinally more consistent groupwise comparisons of cortical attributes across ages, since each pair of ages will have the same subjects. Also, the recovered dataset can be used for more accurate modeling of early brain developmental trajectories.

## 6 Conclusion

There are two major contributions in this paper. First, we proposed a Dynamically-Assembled Regression Forest (DARF). By independently training one decision tree at each vertex on the spherical surface and also dynamically grouping trees in the local neighborhood as a forest, the smoothness of regression results can be guaranteed, and also the computational cost can be largely reduced, compared with using the conventional regression forests with highly-overlapped ROIs. Second, we proposed a novel two-stage method to recover the missing data of vertex-wise cortical morphological attributes in an incomplete longitudinal dataset. The proposed method can effectively exploit all available information in the incomplete dataset to help estimate the missing information. Specifically, in the stage of *pairwise estimation*, as many training subjects as possible are used in the training process, thus better learning the huge diversity among subjects by the regression model; in the stage of *joint refinement*, all time points are taken into account simultaneously, thus better capturing longitudinal information. Our proposed missing data estimation method has been extensively tested on an incomplete dataset with 31 infants, each having up to 5 time points in the first postnatal year, and obtained promising performance for estimation of the cortical thickness maps. In our future work, we will test our method for estimation of other cortical morphological attributes, such as sulcal depth, surface area, cortical folding, and cortical local gyrification.



## Acknowledgments

This work was supported in part by National Institutes of Health (grant numbers EB006733, EB008760, EB008374, EB009634, MH088520, MH070890, MH064065, MH100217, MH108914, NS055754, and HD053000). Dr. Gang Li was supported by NIH K01MH107815.

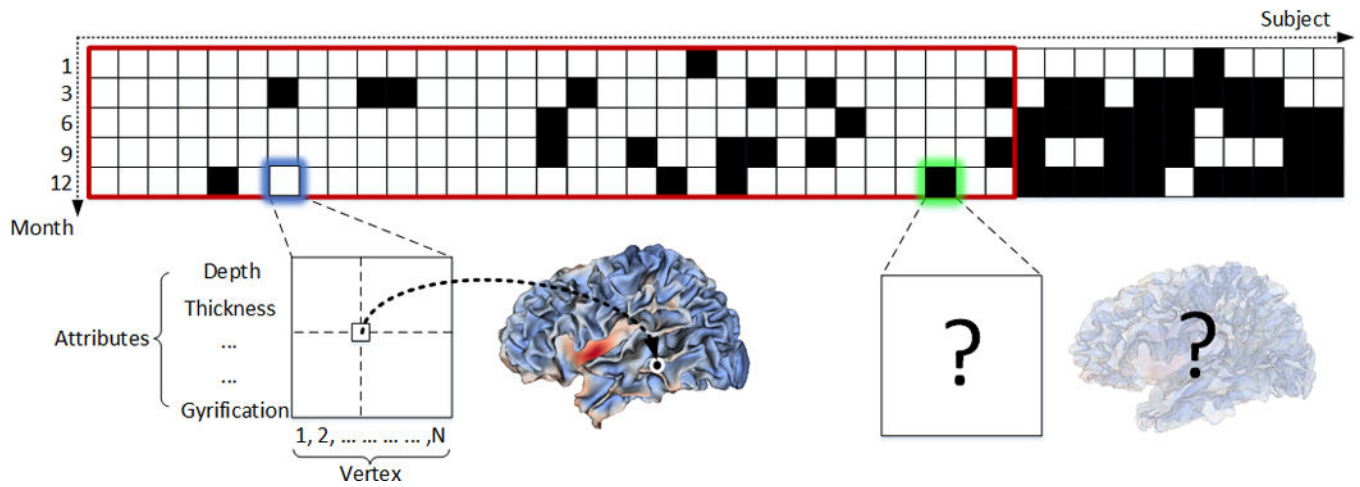
## Reference

- Almli CR, Rivkin MJ, McKinstry RC. Brain Development Cooperative, G. The NIH MRI study of normal brain development (Objective-2): newborns, infants, toddlers, and preschoolers. *NeuroImage*. 2007; 35:308–325. [PubMed: 17239623]
- Awate SP, Yushkevich PA, Song Z, Licht DJ, Gee JC. Cerebral cortical folding analysis with multivariate modeling and testing: Studies on gender differences and neonatal development. *Neuroimage*. 2010; 53:450–459. [PubMed: 20630489]
- Bernal-Rusiel JL, Greve DN, Reuter M, Fischl B, Sabuncu MR. Alzheimer's Disease Neuroimaging, I. Statistical analysis of longitudinal neuroimage data with Linear Mixed Effects models. *NeuroImage*. 2013; 66:249–260. [PubMed: 23123680]
- Breiman L. Random Forests. *Machine Learning*. 2001; 45:5–32.
- Cai J-F, Cand EJ, Shen Z. A Singular Value Thresholding Algorithm for Matrix Completion. *SIAM J. on Optimization*. 2010; 20:1956–1982.
- Cand EJ, Tao T. The power of convex relaxation: near-optimal matrix completion. *IEEE Trans. Inf. Theor*. 2010; 56:2053–2080.
- Candès E, Recht B. Exact Matrix Completion via Convex Optimization. *Found Comput Math*. 2009; 9:717–772.
- Ching WK, Li L, Tsing NK, Tai CW, Ng TW, Wong AS, Cheng KW. A weighted local least squares imputation method for missing value estimation in microarray gene expression data. *International journal of data mining and bioinformatics*. 2010; 4:331–347. [PubMed: 20681483]
- Criminisi A, Shotton J, Konukoglu E. Decision Forests: A Unified Framework for Classification, Regression, Density Estimation, Manifold Learning and Semi-Supervised Learning. *Found. Trends. Comput. Graph. Vis*. 2012; 7:81–227.
- Dale AM, Fischl B, Sereno MI. Cortical surface-based analysis. I. Segmentation and surface reconstruction. *NeuroImage*. 1999; 9:179–194. [PubMed: 9931268]
- Dubois J, Benders M, Borradori-Tolsa C, Cachia A, Lazeyras F, Ha-Vinh Leuchter R, Sizonenko SV, Warfield SK, Mangin JF, Huppi PS. Primary cortical folding in the human newborn: an early marker of later functional development. *Brain : a journal of neurology*. 2008; 131:2028–2041. [PubMed: 18587151]
- Fan Y, Shi F, Smith JK, Lin W, Gilmore JH, Shen D. Brain anatomical networks in early human brain development. *NeuroImage*. 2011; 54:1862–1871. [PubMed: 20650319]
- Faramarzi E, Rajan D, Christensen MP. Unified blind method for multi-image super-resolution and single/multi-image blur deconvolution. *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*. 2013; 22:2101–2114. [PubMed: 23314775]
- Fischl B. FreeSurfer. *NeuroImage*. 2012; 62:774–781. [PubMed: 22248573]
- Fischl B, Dale AM. Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proceedings of the National Academy of Sciences of the United States of America*. 2000; 97:11050–11055. [PubMed: 10984517]
- Fischl B, Sereno MI, Dale AM. Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. *NeuroImage*. 1999a; 9:195–207. [PubMed: 9931269]
- Fischl B, Sereno MI, Tootell RB, Dale AM. High-resolution intersubject averaging and a coordinate system for the cortical surface. *Human brain mapping*. 1999b; 8:272–284. [PubMed: 10619420]
- Gilmore JH, Lin W, Prastawa MW, Looney CB, Vetsa YS, Knickmeyer RC, Evans DD, Smith JK, Hamer RM, Lieberman JA, Gerig G. Regional gray matter growth, sexual dimorphism, and cerebral asymmetry in the neonatal brain. *The Journal of neuroscience : the official journal of the Society for Neuroscience*. 2007; 27:1255–1260. [PubMed: 17287499]

- Gilmore JH, Shi F, Woolson SL, Knickmeyer RC, Short SJ, Lin W, Zhu H, Hamer RM, Styner M, Shen D. Longitudinal development of cortical and subcortical gray matter from birth to 2 years. *Cerebral cortex*. 2012; 22:2478–2485. [PubMed: 22109543]
- Han X, Pham DL, Tosun D, Rettmann ME, Xu CY, Prince JL. CRUISE: Cortical reconstruction using implicit surface evolution. *NeuroImage*. 2004; 23:997–1012. [PubMed: 15528100]
- Kaukola T, Kapellou O, Laroche S, Counsell SJ, Dyet LE, Allsop JM, Edwards AD. Severity of perinatal illness and cerebral cortical growth in preterm infants. *Acta paediatrica*. 2009; 98:990–995. [PubMed: 19302092]
- Kim JS, Singh V, Lee JK, Lerch J, Ad-Dab'bagh Y, MacDonald D, Lee JM, Kim SI, Evans AC. Automated 3-D extraction and evaluation of the inner and outer cortical surfaces using a Laplacian map and partial volume effect classification. *NeuroImage*. 2005; 27:210–221. [PubMed: 15896981]
- Li G, Lin W, Gilmore JH, Shen D. Spatial Patterns, Longitudinal Development, and Hemispheric Asymmetries of Cortical Thickness in Infants from Birth to 2 Years of Age. *The Journal of neuroscience : the official journal of the Society for Neuroscience*. 2015a; 35:9150–9162. [PubMed: 26085637]
- Li G, Nie J, Wang L, Shi F, Gilmore JH, Lin W, Shen D. Measuring the dynamic longitudinal cortex development in infants by reconstruction of temporally consistent cortical surfaces. *NeuroImage*. 2014a; 90:266–279. [PubMed: 24374075]
- Li G, Nie J, Wang L, Shi F, Lin W, Gilmore JH, Shen D. Mapping region-specific longitudinal cortical surface expansion from birth to 2 years of age. *Cerebral cortex*. 2013; 23:2724–2733. [PubMed: 22923087]
- Li G, Nie J, Wang L, Shi F, Lyall AE, Lin W, Gilmore JH, Shen D. Mapping longitudinal hemispheric structural asymmetries of the human cerebral cortex from birth to 2 years of age. *Cerebral cortex*. 2014b; 24:1289–1300. [PubMed: 23307634]
- Li G, Nie J, Wu G, Wang Y, Shen D. Alzheimer's Disease Neuroimaging, I. Consistent reconstruction of cortical surfaces from longitudinal brain MR images. *NeuroImage*. 2012; 59:3805–3820. [PubMed: 22119005]
- Li G, Wang L, Shi F, Gilmore JH, Lin W, Shen D. Construction of 4D high-definition cortical surface atlases of infants: Methods and applications. *Medical image analysis*. 2015b; 25:22–36. [PubMed: 25980388]
- Li G, Wang L, Shi F, Lin W, Shen D. Constructing 4D infant cortical surface atlases based on dynamic developmental trajectories of the cortex. *Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention*. 2014c; 17:89–96.
- Li G, Wang L, Shi F, Lin W, Shen D. Simultaneous and consistent labeling of longitudinal dynamic developing cortical surfaces in infants. *Medical image analysis*. 2014d; 18:1274–1289. [PubMed: 25066749]
- Li G, Wang L, Shi F, Lyall AE, Lin W, Gilmore JH, Shen D. Mapping longitudinal development of local cortical gyrification in infants from birth to 2 years of age. *The Journal of neuroscience*. 2014e; 34:4228–4238. [PubMed: 24647943]
- Li R, Zhang W, Suk HI, Wang L, Li J, Shen D, Ji S. Deep learning based imaging data completion for improved brain disease diagnosis. *Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention*. 2014f; 17:305–312.
- Liu J, Musialski P, Wonka P, Ye J. Tensor completion for estimating missing values in visual data. *IEEE transactions on pattern analysis and machine intelligence*. 2013; 35:208–220. [PubMed: 22271823]
- Liu T, Nie J, Tarokh A, Guo L, Wong ST. Reconstruction of central cortical surface from brain MRI images: method and application. *NeuroImage*. 2008; 40:991–1002. [PubMed: 18289879]
- Lyall AE, Shi F, Geng X, Woolson S, Li G, Wang L, Hamer RM, Shen D, Gilmore JH. Dynamic Development of Regional Cortical Thickness and Surface Area in Early Childhood. *Cerebral cortex*. 2014

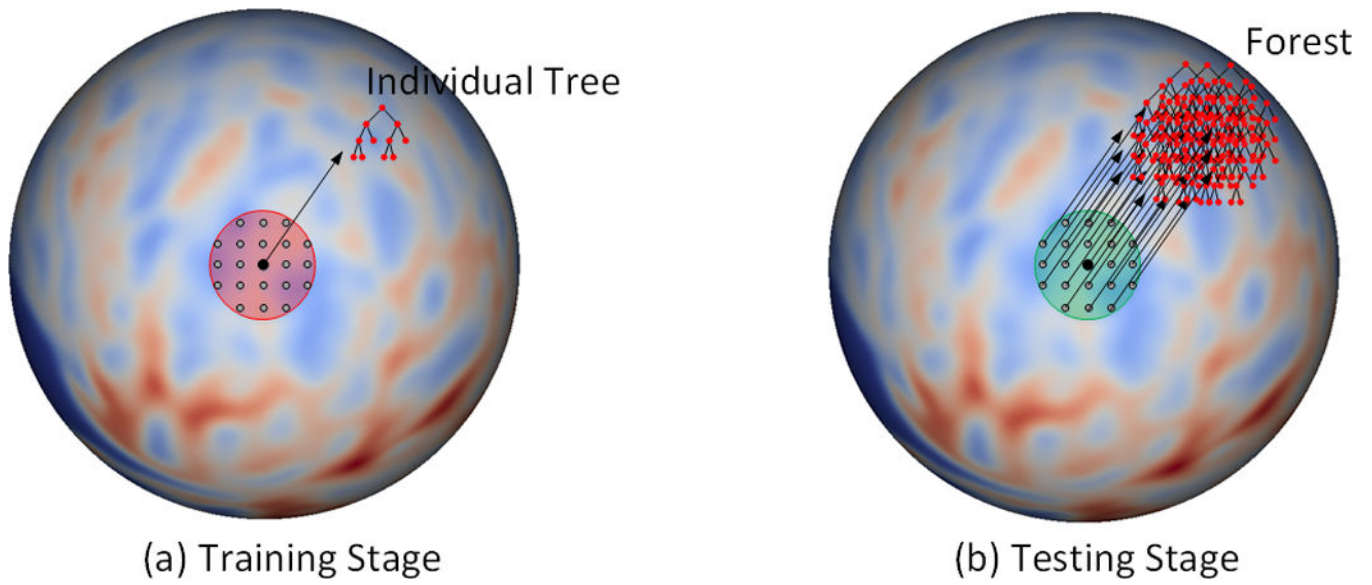
- MacDonald D, Kabani N, Avis D, Evans AC. Automated 3-D extraction of inner and outer surfaces of cerebral cortex from MRI. *NeuroImage*. 2000; 12:340–356. [PubMed: 10944416]
- Mangin JF, Riviere D, Cachia A, Duchesnay E, Cointepas Y, Papadopoulos-Orfanos D, Scifo P, Ochiai T, Brunelle F, Regis J. A framework to study the cortical folding patterns. *NeuroImage*. 2004; 23:S129–S138. [PubMed: 15501082]
- Meng Y, Li G, Lin W, Gilmore JH, Shen D. Spatial distribution and longitudinal development of deep cortical sulcal landmarks in infants. *NeuroImage*. 2014; 100:206–218. [PubMed: 24945660]
- Nie J, Li G, Wang L, Gilmore JH, Lin W, Shen D. A computational growth model for measuring dynamic cortical development in the first year of life. *Cerebral cortex*. 2012; 22:2272–2284. [PubMed: 22047969]
- Nie J, Li G, Wang L, Shi F, Lin W, Gilmore JH, Shen D. Longitudinal development of cortical thickness, folding, and fiber density networks in the first 2 years of life. *Hum Brain Mapp*. 2014; 35:3726–3737. [PubMed: 24375724]
- Schnack HG, van Haren NE, Brouwer RM, Evans A, Durston S, Boomsma DI, Kahn RS, Hulshoff Pol HE. Changes in Thickness and Surface Area of the Human Cortex and Their Relationship with Intelligence. *Cerebral cortex*. 2014
- Schumann CM, Bloss CS, Barnes CC, Wideman GM, Carper RA, Akshoomoff N, Pierce K, Hagler D, Schork N, Lord C, Courchesne E. Longitudinal magnetic resonance imaging study of cortical development through early childhood in autism. *The Journal of neuroscience : the official journal of the Society for Neuroscience*. 2010; 30:4419–4427. [PubMed: 20335478]
- Shattuck DW, Leahy RM. BrainSuite: an automated cortical surface identification tool. *Medical image analysis*. 2002; 6:129–142. [PubMed: 12045000]
- Shi F, Fan Y, Tang S, Gilmore JH, Lin W, Shen D. Neonatal brain image segmentation in longitudinal MRI studies. *NeuroImage*. 2010; 49:391–400. [PubMed: 19660558]
- Shi F, Wang L, Dai Y, Gilmore JH, Lin W, Shen D. LABEL: pediatric brain extraction using learning-based meta-algorithm. *NeuroImage*. 2012; 62:1975–1986. [PubMed: 22634859]
- Shi Y, Lai R, Toga AW. Alzheimer's Disease Neuroimaging, I. Cortical surface reconstruction via unified Reeb analysis of geometric and topological outliers in magnetic resonance images. *IEEE transactions on medical imaging*. 2013; 32:511–530. [PubMed: 23086519]
- Sled JG, Zijdenbos AP, Evans AC. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE transactions on medical imaging*. 1998; 17:87–97. [PubMed: 9617910]
- Thompson P, Toga AW. A surface-based technique for warping three-dimensional images of the brain. *IEEE transactions on medical imaging*. 1996; 15:402–417. [PubMed: 18215923]
- Thung KH, Wee CY, Yap PT, Shen D. Alzheimer's Disease Neuroimaging, I. Neurodegenerative disease diagnosis using incomplete multi-modality data via matrix shrinkage and completion. *NeuroImage*. 2014; 91:386–400. [PubMed: 24480301]
- Tibshirani R. Regression shrinkage and selection via the Lasso. *J Roy Stat Soc B Met*. 1996; 58:267–288.
- Troyanskaya O, Cantor M, Sherlock G, Brown P, Hastie T, Tibshirani R, Botstein D, Altman RB. Missing value estimation methods for DNA microarrays. *Bioinformatics*. 2001; 17:520–525. [PubMed: 11395428]
- Tsiporkova E, Boeva V. Two-pass imputation algorithm for missing value estimation in gene expression time series. *Journal of bioinformatics and computational biology*. 2007; 5:1005–1022. [PubMed: 17933008]
- Van Essen DC, Drury HA, Dickson J, Harwell J, Hanlon D, Anderson CH. An integrated software suite for surface-based analyses of cerebral cortex. *Journal of the American Medical Informatics Association : JAMIA*. 2001; 8:443–459. [PubMed: 11522765]
- Wang L, Shi F, Gao Y, Li G, Gilmore JH, Lin W, Shen D. Integration of sparse multi-modality representation and anatomical constraint for iso-intense infant brain MR image segmentation. *NeuroImage*. 2014a; 89:152–164. [PubMed: 24291615]
- Wang L, Shi F, Li G, Gao Y, Lin W, Gilmore JH, Shen D. Segmentation of neonatal brain MR images using patch-driven level sets. *NeuroImage*. 2014b; 84:141–158. [PubMed: 23968736]

- Wang L, Shi F, Yap PT, Gilmore JH, Lin W, Shen D. 4D multi-modality tissue segmentation of serial infant images. *PLoS One*. 2012; 7:e44596. [PubMed: 23049751]
- Yeo BT, Sabuncu MR, Vercauteren T, Ayache N, Fischl B, Golland P. Spherical demons: fast diffeomorphic landmark-free surface registration. *IEEE transactions on medical imaging*. 2010; 29:650–668. [PubMed: 19709963]
- Yuan L, Wang Y, Thompson PM, Narayan VA, Ye J. Alzheimer's Disease Neuroimaging, I. Multi-source feature learning for joint analysis of incomplete multiple heterogeneous neuroimaging data. *NeuroImage*. 2012; 61:622–632. [PubMed: 22498655]



**Figure 1.**

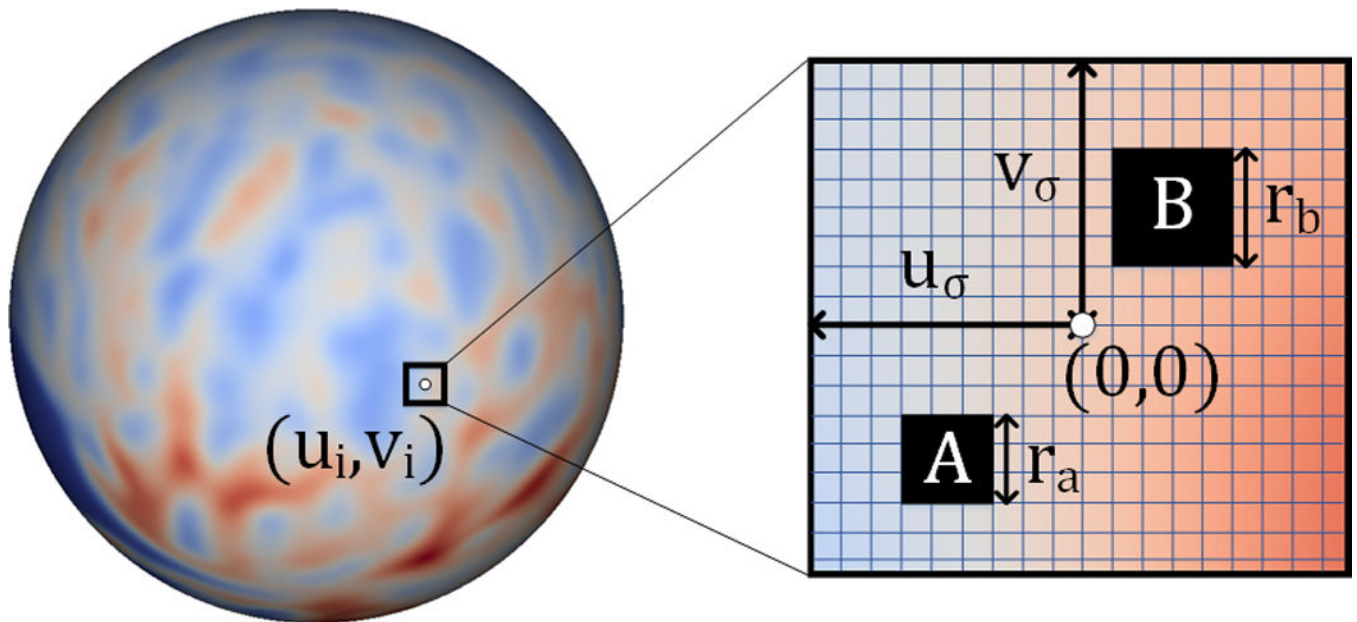
Illustration of the longitudinal infant dataset used in our study. Each block indicates the cortical morphological attributes of all vertices of the entire cortical surface for a specific subject (column) at a specific time point (row). The black blocks indicate the missing data at the respective time point. The blocks enclosed by the red rectangle indicate the dataset used in this paper.



**Figure 2.**

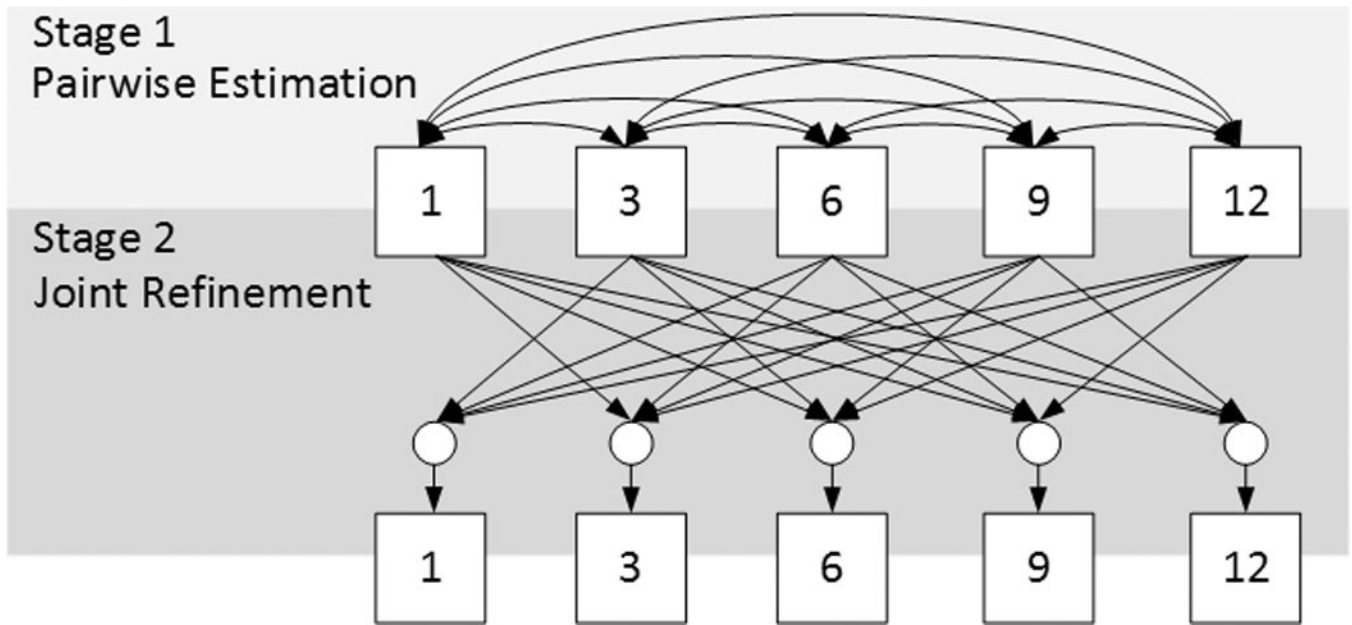
Training stage (a) and testing stage (b). (a) The red region is the neighborhood, where all the vertices on the spherical space are used as training samples. (b) The green region is the neighborhood, where all the individual trees are combined together to form a forest in the testing stage. Note that the red and green regions in (a) and (b) could have different sizes.



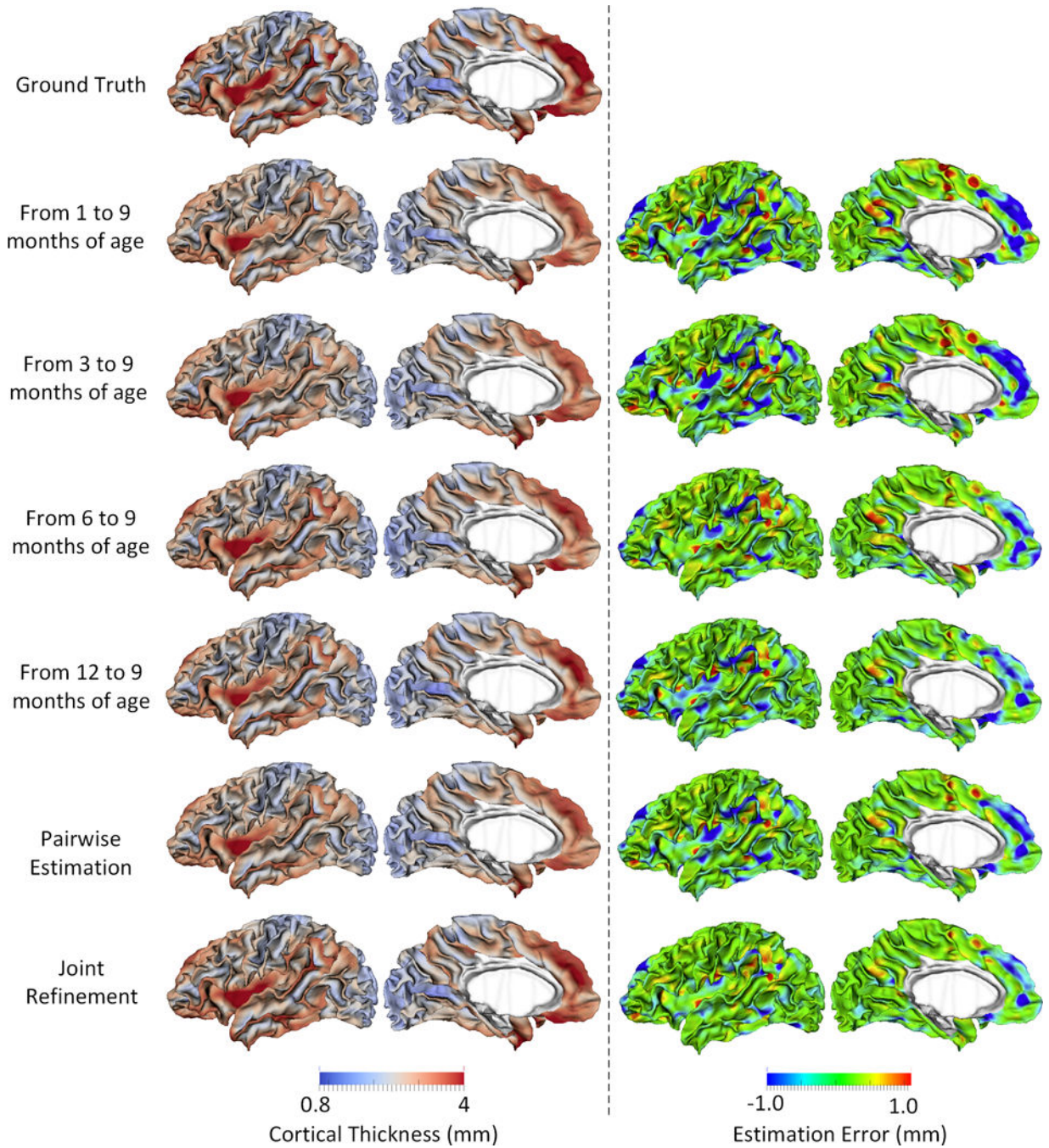


**Figure 3.**

Computation of Haar-like features on a resampled spherical surface atlas. The horizontal and vertical lines in the rectangular patch denote for latitude and longitude directions, respectively. The blocks A and B are the two randomly selected regions. The value of Haar-like feature is defined as 1) the mean value of the cortical attributes in the block A, or 2) the mean value of the cortical attributes in the block A subtracting that in the block B.

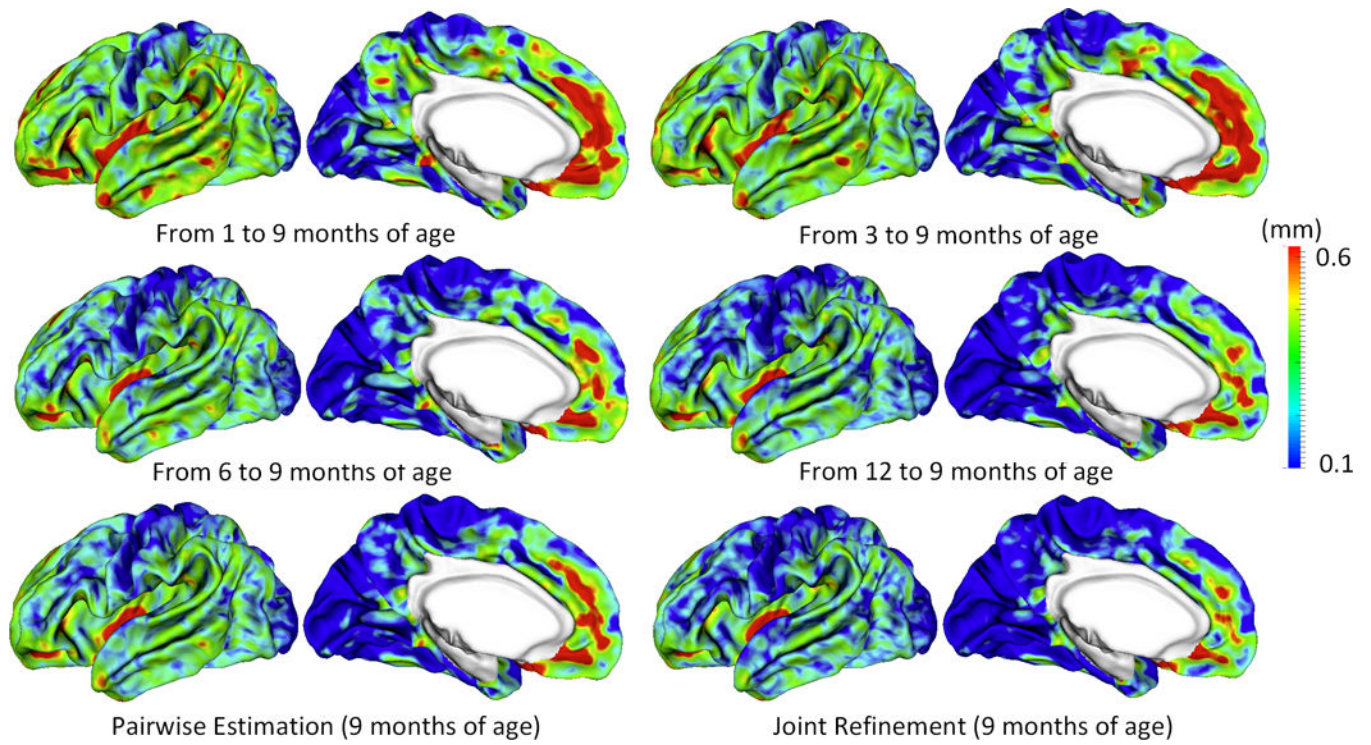


**Figure 4.** Overview of the proposed method for estimation of missing data. The box with number stands for the data at the corresponding time point. The directed edges represent the processes of estimation of the missing data at the target time points (as pointed by the arrowhead) based on the data at the available time points (at the tail side). In Stage 1, the edges are bidirectional, which means that the estimation is performed twice by exchanging between the input and the output time points. The circles in Stage 2 denote the use of multiple time points jointly.

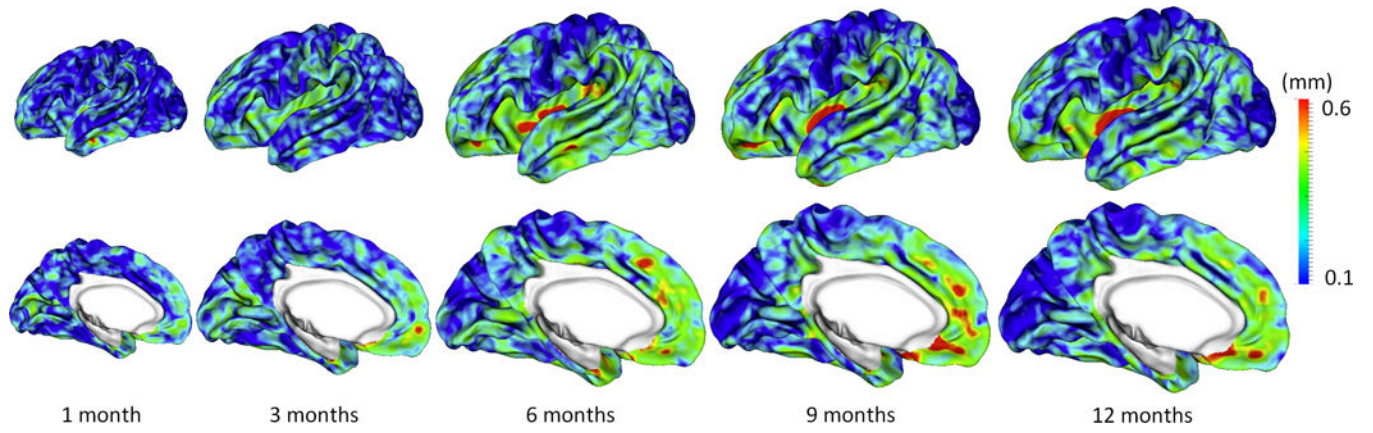


**Figure 5.** Estimations of the vertex-wise missing cortical thickness at 9-months-old for a randomly selected infant. The first two columns show the maps of ground truth and the estimation of cortical thickness at each step. The last two columns show the maps of estimation errors at each step.

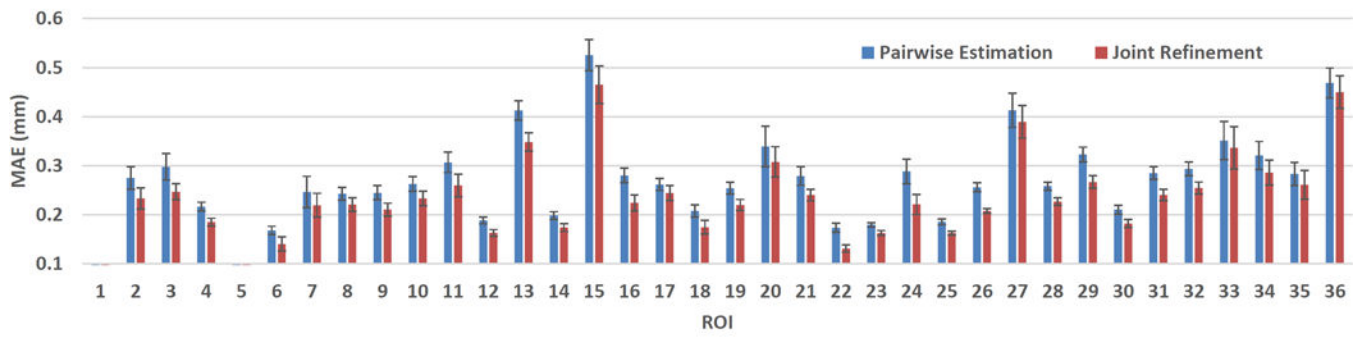
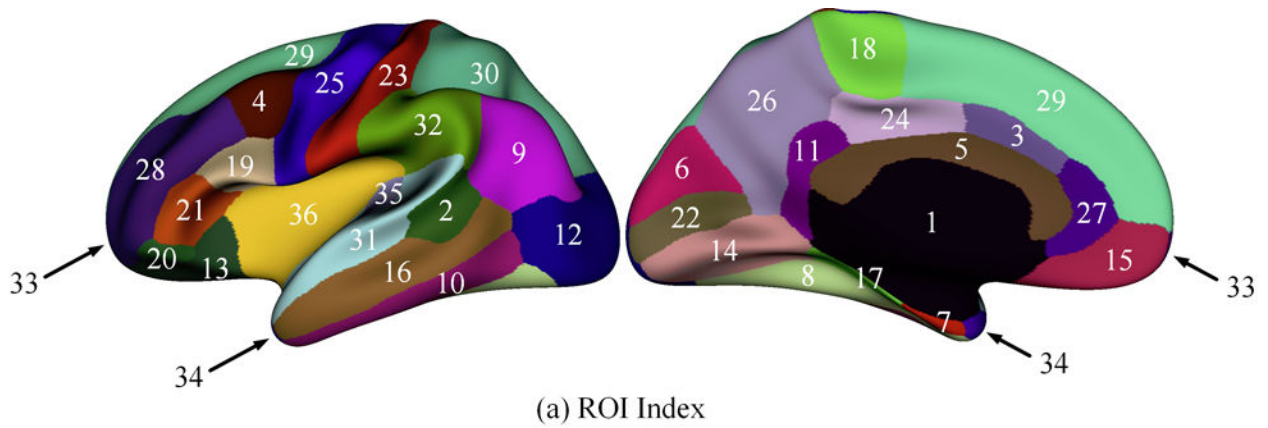




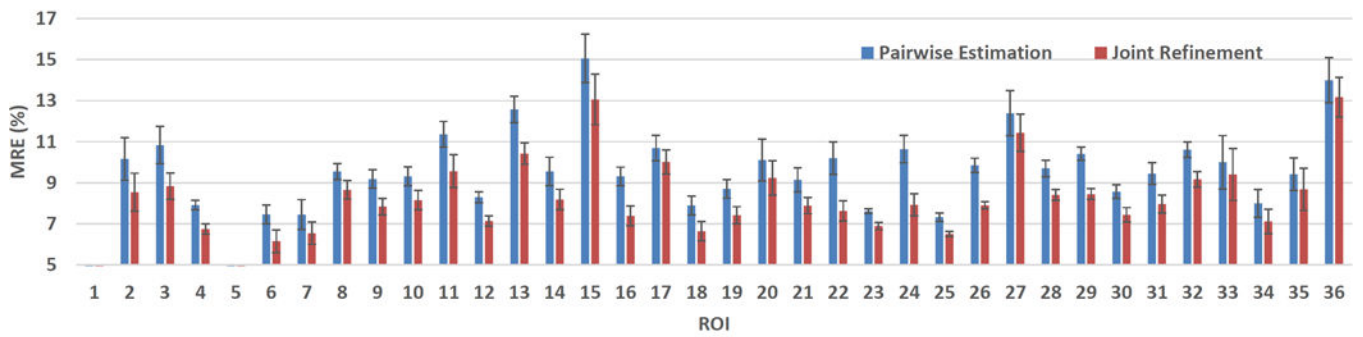
**Figure 6.** Vertex-wise average estimation errors (mm) of missing cortical thickness at 9-months-old for all subjects at each step of estimation.



**Figure 7.** Vertex-wise average estimation errors (mm) of missing cortical thickness at 5 time points for all subjects by using the proposed method.



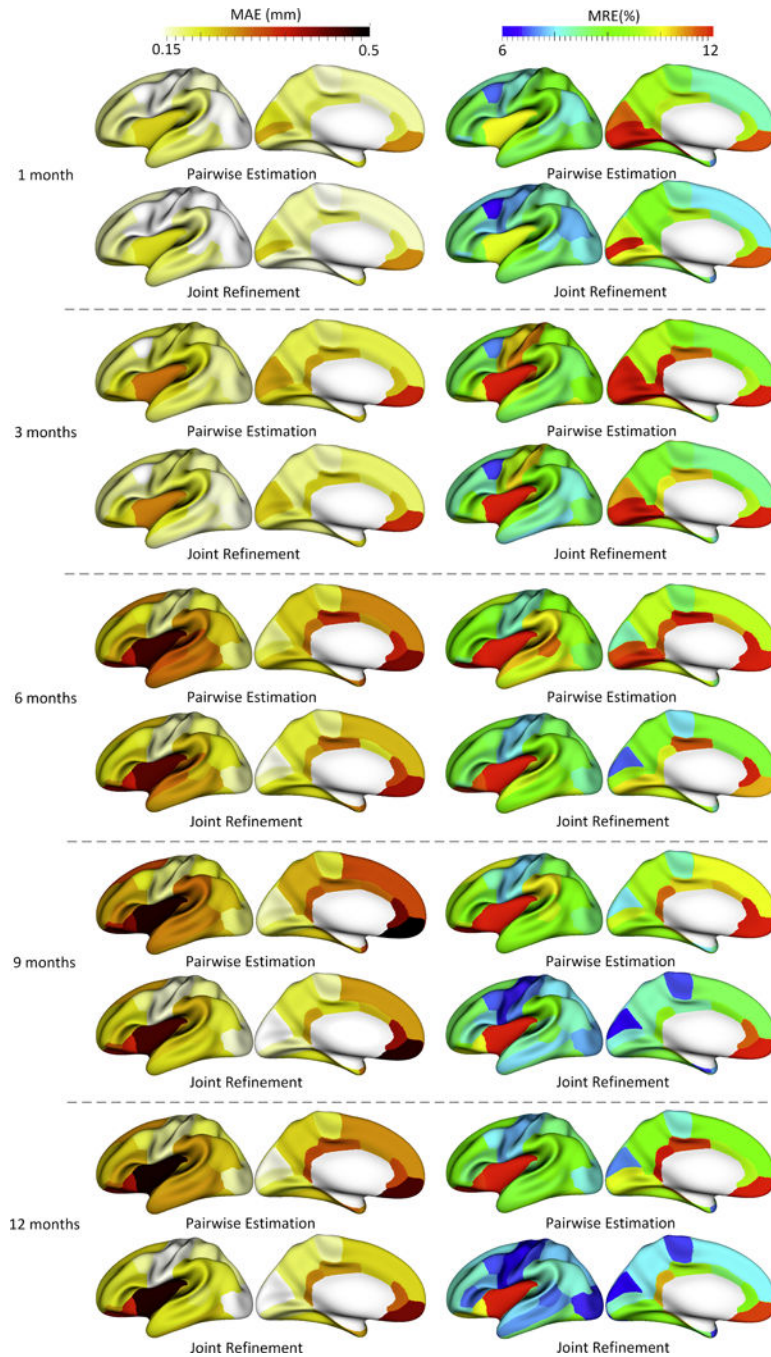
(b) Mean Absolute Error



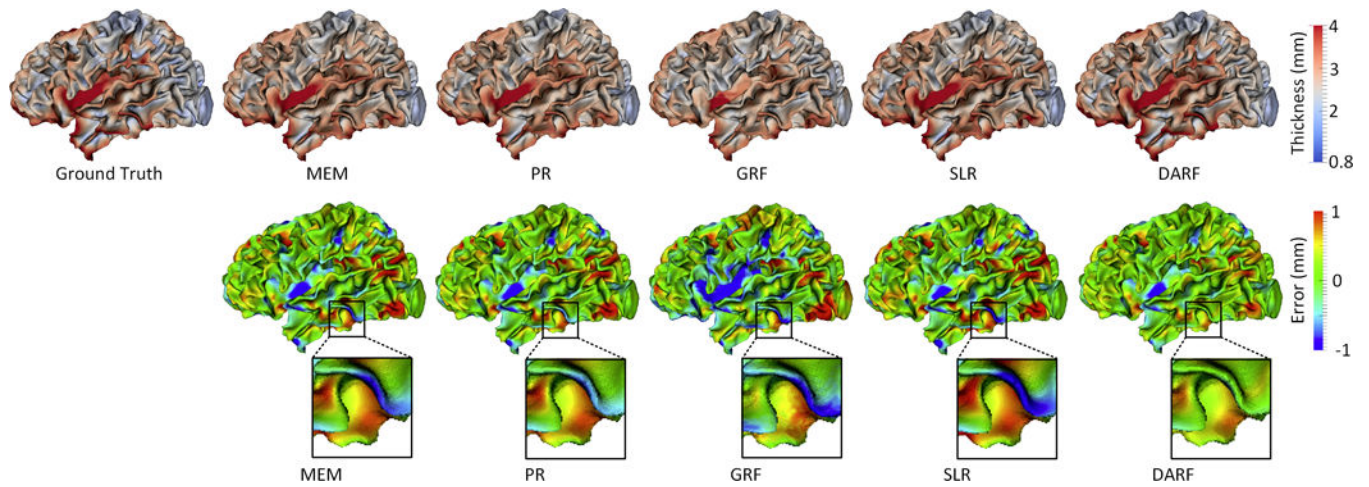
(c) Mean Relative Error

**Figure 8.** Error measures in 36 ROIs for estimation of missing cortical thickness at 9-months-old. ROIs 1 and 5 are excluded as there is no definition of cortical thickness for these two regions.



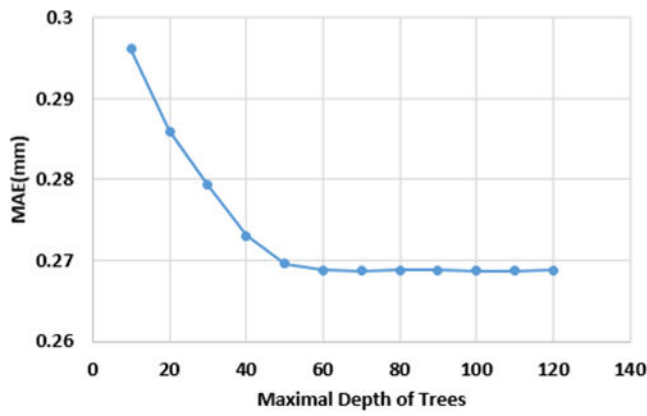


**Figure 9.** The accuracy of each step in estimating the missing cortical thickness at 1-, 3-, 6-, 9- and 12-months-old in 36 ROIs.

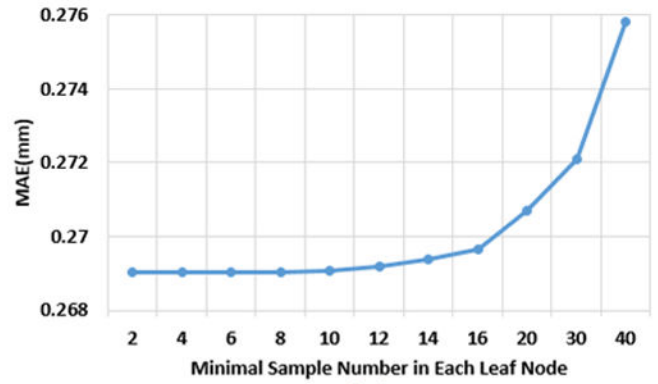


**Figure 10.**

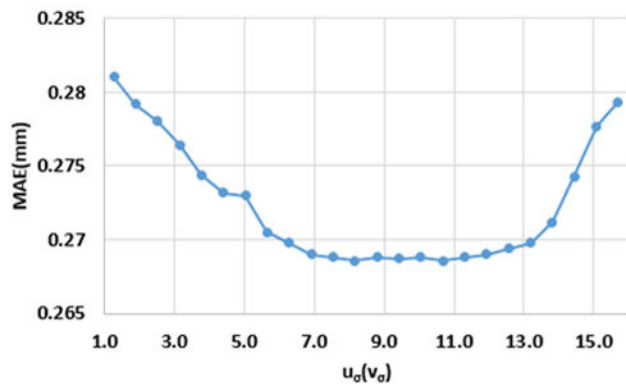
Estimation of the vertex-wise cortical thickness map (mm) of a randomly selected infant at 9 months of age by five different methods. The first row shows the ground truth and the estimation of cortical thickness map. The second row shows the estimation error maps (mm).



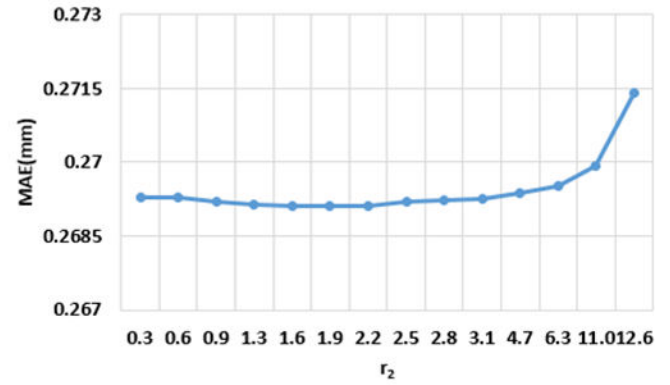
(a)



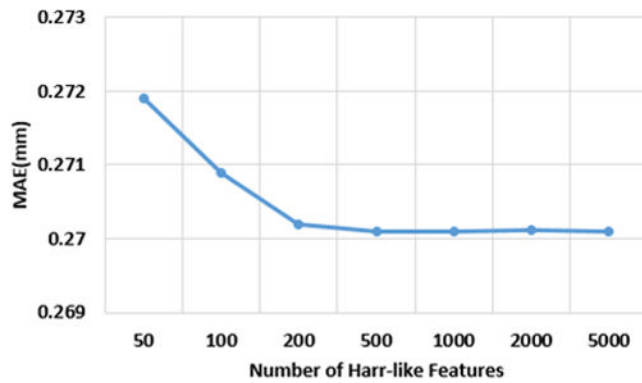
(b)



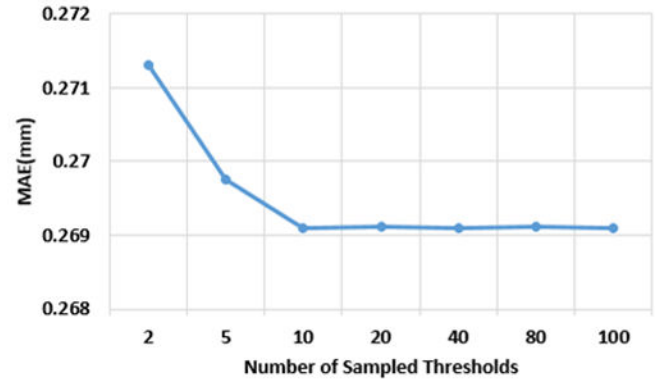
(c)



(d)

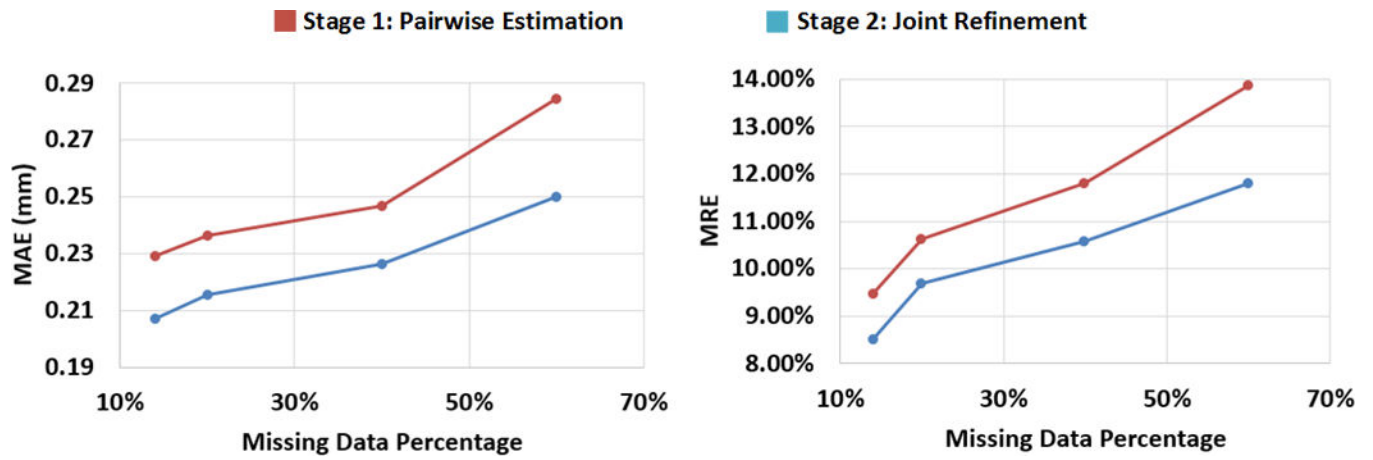


(e)



(f)

**Figure 11.**  
The influence of 6 parameters to the estimation accuracy (mm).



**Figure 12.**  
Relationship between estimation errors and the portions of missing data.

Quantitative measures of estimation results for the missing cortical thickness by normalized mean squared error (NMSE), PE and JR respectively stand for two stages of pairwise estimation and joint refinement.

**Table 1**

NMSE	Baseline Time Points (months)						PE	JR	p-value
	1	3	6	9	12	12			
1	-	0.0129 ±0.0022	0.0153 ±0.0028	0.0156 ±0.0024	0.0147 ±0.0025	0.0128 ±0.0022	<b>0.0117</b> ±0.0019	<b>9×10<sup>-4</sup></b>	
3	0.0163 ±0.0032	-	0.0187 ±0.0035	0.0186 ±0.0039	0.0175 ±0.0035	0.0157 ±0.0035	<b>0.0140</b> ±0.0032	<b>6×10<sup>-6</sup></b>	
6	0.0231 ±0.0037	0.0221 ±0.0049	-	0.0157 ±0.0049	0.0158 ±0.0035	0.0155 ±0.0040	<b>0.0139</b> ±0.0030	<b>1×10<sup>-3</sup></b>	
9	0.0220 ±0.0023	0.0211 ±0.0025	0.0155 ±0.0025	-	0.0133 ±0.0015	0.0147 ±0.0019	<b>0.0121</b> ±0.0021	<b>2×10<sup>-3</sup></b>	
12	0.0191 ±0.0024	0.0182 ±0.0021	0.0139 ±0.0021	0.0124 ±0.0022	-	0.0128 ±0.0018	<b>0.0107</b> ±0.0019	<b>7×10<sup>-8</sup></b>	

Quantitative measures of estimation results for the missing cortical thickness by mean absolute error (MAE), PE and JR respectively stand for two stages of pairwise estimation and joint refinement.

**Table 2**

MAE (mm)	Input Time Points (months)						PE	JR	p-value
	1	3	6	9	12	12			
1	-	0.176 ±0.015	0.195 ±0.014	0.198 0.017	0.192 ±0.016	0.168 ±0.013	0.179 ±0.015	5×10 <sup>-4</sup>	
3	0.209 ±0.026	-	0.230 ±0.027	0.231 ±0.029	0.222 ±0.026	0.193 ±0.025	0.209 ±0.027	8×10 <sup>-6</sup>	
6	0.307 ±0.035	0.306 ±0.046	-	0.252 ±0.043	0.252 ±0.040	0.233 ±0.034	0.264 ±0.041	2×10 <sup>-4</sup>	
9	0.326 ±0.027	0.317 ±0.025	0.264 ±0.023	-	0.241 ±0.016	0.224 ±0.016	0.261 ±0.021	5×10 <sup>-8</sup>	
12	0.308 ±0.025	0.301 ±0.021	0.253 ±0.022	0.232 ±0.017	-	0.215 ±0.016	0.245 ±0.019	7×10 <sup>-11</sup>	



Quantitative measures of estimation results for the missing cortical thickness by mean relative error(MRE), PE and JR respectively stand for two stages of pairwise estimation and joint refinement.

**Table 3**

MRE (%)	Input Time Points (months)						PE	JR	p-value
	1	3	6	9	12	12			
1	-	8.75 ±0.75	9.71 ±0.84	9.92 ±1.01	9.65 ±1.01	9.65 ±1.01	8.93 ±0.95	<b>8.32</b> ± <b>0.69</b>	<b>3×10<sup>-4</sup></b>
3	9.94 ±1.14	-	11.05 ±1.49	11.07 ±1.34	10.90 ±1.39	10.02 ±1.33	10.02 ±1.33	<b>9.23</b> ± <b>1.25</b>	<b>6×10<sup>-7</sup></b>
6	12.25 ±0.67	11.99 ±0.82	-	9.81 ±0.77	9.83 ±0.70	9.92 ±0.74	9.92 ±0.74	<b>9.08</b> ± <b>0.79</b>	<b>6×10<sup>-6</sup></b>
9	11.91 ±0.62	11.64 ±0.69	9.54 ±0.85	-	8.72 ±0.33	9.51 ±0.56	9.51 ±0.56	<b>8.14</b> ± <b>0.51</b>	<b>3×10<sup>-9</sup></b>
12	11.33 ±0.59	11.08 ±0.52	9.25 ±0.67	8.43 ±0.52	-	9.04 ±0.55	9.04 ±0.55	<b>7.81</b> ± <b>0.52</b>	<b>9×10<sup>-12</sup></b>

Table 4

Quantitative evaluation of the performance of cortical thickness estimation by using MEM, PR, CRF, SLR and DARE.

Metric	Method	1 month	3 months	6 months	9 months	12 months
NMSE	MEM	0.0177±0.0023	0.0244±0.0175	0.0185±0.0033	0.0156±0.0012	0.0154±0.0010
	PR	0.2321±0.0506	0.0389±0.0100	0.0223±0.0053	0.0142±0.0027	0.0888±0.0171
	CRF	0.0165±0.0035	0.0188±0.0031	0.0285±0.0051	0.0189±0.0031	0.0163±0.0023
	SLR	0.0139±0.0037	0.0161±0.0039	0.0183±0.0044	0.0153±0.0034	0.0154±0.0042
MAE (mm)	DARE	<b>0.0117±0.0019</b>	<b>0.0140±0.0032</b>	<b>0.0139±0.0030</b>	<b>0.0121±0.0021</b>	<b>0.0107±0.0019</b>
	MEM	0.215±0.016	0.269±0.049	0.264±0.035	0.270±0.015	0.258±0.020
	PR	0.764±0.091	0.328±0.043	0.295±0.044	0.245±0.024	0.634±0.072
	CRF	0.181±0.023	0.224±0.028	0.330±0.041	0.306±0.024	0.204±0.027
MRE (%)	SLR	0.175±0.023	0.205±0.027	0.263±0.038	0.268±0.026	0.259±0.026
	DARE	<b>0.168±0.013</b>	<b>0.193±0.025</b>	<b>0.233±0.034</b>	<b>0.224±0.016</b>	<b>0.215±0.016</b>
	MEM	12.25±0.73	13.68±0.44	11.05±0.98	10.46±0.45	10.68±0.33
	PR	37.17±0.048	15.96±2.30	11.12±1.05	8.5±0.67	22.52±2.42
MRE (%)	CRF	9.23±0.91	11.51±1.73	12.63±1.25	11.65±0.95	11.42±0.76
	SLR	9.05±0.83	10.05±1.52	11.01±1.12	10.44±0.98	10.69±0.77
	DARE	<b>8.32±0.69</b>	<b>9.23±1.25</b>	<b>9.08±0.79</b>	<b>8.14±0.51</b>	<b>7.81±0.52</b>