

**COMPUTATIONAL CORTICAL SURFACE ANALYSIS FOR STUDY OF EARLY
BRAIN DEVELOPMENT**

Yu Meng

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Computer Science.

Chapel Hill
2017

Approved by:

Dinggang Shen

Marc Niethammer

Martin Styner

Ming Lin

Gang Li

©2017
Yu Meng
ALL RIGHTS RESERVED

ABSTRACT

Yu Meng: Computational Cortical Surface Analysis for Study of
Early Brain Development
(Under the direction of Dinggang Shen)

The study of morphological attributes of the cerebral cortex and their development is very important in understanding the dynamic and critical early brain development. Comparing with conventional studies in the image space, cortical surface-based analysis provides a better way to display, observe, and quantify the attributes of the cerebral cortex. The goal of this dissertation is to develop novel cortical surface-based methods for better studying the attributes of the cerebral cortex during early brain development. Specifically, this dissertation aims to develop methods for 1) estimating the development of morphological attributes of the cerebral cortex and 2) discovering the major cortical folding patterns.

Estimation of the Development of Cortical Attributes. The early development of cortical attributes is highly correlated to the brain cognitive functionality and some neurodevelopmental disorders. Hence, accurately modeling the early development of cortical attributes is crucial for better understanding the mysterious normal and abnormal brain development. This task is very challenging, because infant cortical attributes change dramatically, complicatedly and regionally-heterogeneously during the first year of life. To address these problems, this dissertation proposes a Dynamically-Assembled Regression Forest (DARF). DARF first trains a single decision tree at each vertex on the cortical surface, and then groups nearby decision trees around each vertex as a vertex-specific forest to predict the cortical attribute. Since

the vertex-specific forest can better capture regional details than the conventional regression forest trained for the whole brain, the prediction result is more precise. Moreover, because nearby forests share a large portion of decision trees, the prediction result is spatially smooth. On the other hand, missing cortical attribute maps in the longitudinal datasets often lead to insufficient data for unbiased analysis or training of accurate prediction models. To address this issue, a missing data estimation strategy based on DARF is further proposed. Experiments show that DARF outperforms the existing popular regression methods, and the proposed missing data estimation strategy based on DARF can effectively recover the missing cortical attribute maps.

Discovery of Major Cortical Folding Patterns. The folding patterns of the cerebral cortex are highly variable across subjects. Exploring major cortical folding patterns in neonates is of great importance in neuroscience. Conventional geometric measurements of the cortex have limited capability in distinguishing major folding patterns. Although the recent sulcal pits-based analysis provides a better way for comparing sulcal patterns across individuals of adults or older children, whether and how sulcal pits are suitable for discovering major sulcal patterns in infants remain unknown. This dissertation adapts a sulcal pits extraction method from adults to infants, and validates the spatial consistency of sulcal pits in infants, so that they can be used as reliable landmarks for exploring major sulcal patterns. This dissertation further proposes a sulcal graph-based method for discovering major sulcal patterns, which is then applied to studying three primary cortical regions in 677 neonatal cortical surfaces. The experiments show that the proposed method is able to identify the previously unreported major sulcal patterns. Finally, this dissertation investigates and verifies that the sulcal pattern information could be utilized to help DARF for better estimating cortical attribute maps.

To my mentors and friends, I couldn't have done this without you.
Thank you for all of your support along the way.

AKNOWLEDGEMENTS

It has been a privilege and honor for me to study at the University of North Carolina at Chapel Hill. Doing research and writing a doctoral dissertation was challenging, but I was lucky to get a lot of help and encouragement from my dear colleagues, teachers, and friends.

In particular I thank my advisor Dr. Dinggang Shen for continuously helping me with my research, all the excellent advices and suggestions, and the valuable time he spent with me discussing my works. I also thank him for his support not only in research but also in life.

My committee played an important role in my education, research, and this dissertation. I thank Dr. Marc Niethammer, who is my academic adviser, for his valuable suggestions about my course study and research, and for all his help in my proposal, oral exam, and defense. I thank Dr. Martin Styner for spending a lot of time talking with me about my research, providing valuable advices, and teaching me the basic MR imaging principles. I thank Dr. Ming Lin, who helped me a lot during the first year of my PhD program and always answered to my questions in time. I thank Dr. Gang Li for closely working with me as a mentor. He is an expert in cortical surface-based analysis, and I learned a lot from him.

I thank the faculty, staff, scholars, and students in the Image Display Enhancement and Analysis Group and in the Biomedical Research Imaging Center for creating a great research environment. It has been a pleasure to learn through group meetings, team projects, and one-on-one discussions. I specially thank Dr. Weili Lin, Dr. Li Wang, Dr. Islem Rekik, Dr. John Gilmore,

and Dr. Yaozong Gao for their great help with this work, and thank Dong Nie, Zhengwang Wu, Kim Han Thung, and Xiaofeng Zhu for their indirect help with this work.

I thank the faculty and staff of the Department of Computer Science who offered me a lot of help. I especially thank Jodie Gregoritsch for her always there helping me during my entire PhD program.

A special thank goes to Shuting Zheng, for her invaluable caring and support.

Finally, I thank my parents, Qingyuan Meng and Wenhua Hou, for their love and forever support throughout my life.

TABLE OF CONTENTS

LIST OF TABLES	xii
LIST OF FIGURES	xiii
LIST OF ABBREVIATIONS.....	xv
1 INTRODUCTION	1
1.1 Introduction.....	1
1.2 Predicting the Development of Vertex-Wise Cortical Attributes	2
1.2.1 Significance.....	2
1.2.2 Challenges.....	3
1.2.3 Existing Methods	4
1.3 Estimating the Missing Cortical Attributes.....	7
1.3.1 Significance.....	7
1.3.2 Challenges.....	8
1.3.3 Existing Methods	8
1.4 Discovering Major Cortical Folding Patterns in Neonates	10
1.4.1 Significance.....	10
1.4.2 Challenges.....	11
1.4.3 Existing Methods	12
1.5 Thesis	15

1.6 Overview of Chapters	19
1.7 Summary	20
2 BACKGROUND	21
2.1 Datasets	21
2.2 Data Preprocessing	22
2.2.1 Image Processing	23
2.2.2 Cortical Surface Reconstruction	24
2.2.3 Cortical Surface Registration	25
2.3 Basis of Regression Forest	26
2.3.1 Training and Testing	26
2.3.2 Effects of Parameters	28
2.3.3 Advantages and Disadvantages	30
2.4. Sulcal Pits	31
2.4.1 Sulcal Pit and Its Characteristics	31
2.4.2 Existing Research	32
2.5 Summary	33
3 ESTIMATION OF THE EARLY DEVELOPMENT OF CORTICAL ATTRIBUTES	35
3.1 Predicting the Development of Vertex-wise Cortical Attributes	35
3.1.1 Dynamically-Assembled Regression Forest	36
3.1.2 Input Features	39
3.1.3 Quantitative Evaluation	41

3.1.4 Experiments and Results.....	41
3.2 Estimating the Missing Cortical Attribute Maps	48
3.2.1 Missing Cortical Attribute Estimation Method.....	48
3.2.2 Experiments and Results.....	51
3.3 Summary.....	59
4 DISCOVERY OF THE MAJOR CORTICAL FOLDING PATTERNS IN INFANTS.....	62
4.1 Sulcal Pits.....	62
4.1.1 Sulcal Pits Extraction.....	63
4.1.2 Spatial Distribution and Longitudinal Development	68
4.2 Discovery of Major Sulcal Patterns in Neonates	72
4.2.1 Sulcal Graph and Similarity Measurement	73
4.2.2 Fusion of Sulcal Graph Similarities	76
4.2.3 Sulcal Pattern Clustering.....	77
4.2.4 Experiments and Results.....	78
4.3 Utilization of Sulcal Patterns to Help Predict Cortical Attributes Development.....	86
4.3.1 Motivation.....	86
4.3.2 Method	86
4.3.3 Experiments and Results.....	87
4.4 Summary.....	89
5 SUMMARY AND FUTURE WORK.....	92
5.1 Summary.....	92

5.2 Future Work.....	98
REFERENCES	101

LIST OF TABLES

Table 3.1. Quantitative measures of cortical thickness prediction using mean absolute errors (MAE).....	45
Table 3.2. Quantitative measures of cortical thickness prediction using mean relative errors (MRE).....	46
Table 3.3. Quantitative measures of estimation results for the missing cortical thickness by the normalized mean squared error (NMSE).....	55
Table 3.4. Quantitative measures of estimation results for the missing cortical thickness by the mean absolute error (MAE).....	55
Table 3.5. Quantitative measures of estimation results for the missing cortical thickness by the mean relative error (MRE).....	55
Table 3.6. Quantitative evaluation of the performance of cortical thickness estimation by using MEM, PR, CRF, SLR and DARF.....	59
Table 4.1. Ratios of different reported number of patterns.....	85
Table 4.2. Pattern discovery rate.....	85
Table 4.3. Quantitative comparison of error estimations in 7 ROIs.....	88

LIST OF FIGURES

Figure 1.1. Representations of the cortical surface.....	2
Figure 2.1. Data preprocessing pipeline.....	23
Figure 2.2. Examples of the cortical surfaces and cortical thickness maps.....	24
Figure 2.3. Sulcal pits.....	32
Figure 3.1. Training and testing stages for DARF.....	38
Figure 3.2. Computation of Haar-like features on a resampled spherical surface atlas.....	40
Figure 3.3. Prediction of the cortical thickness maps for a randomly selected subject.....	42
Figure 3.4. Predicted mean cortical thickness and ground truth.....	43
Figure 3.5. The prediction results from multiple available time points.....	45
Figure 3.6. The average prediction errors (mm) in 35 cortical ROIs for 15 infants.....	46
Figure 3.7. Prediction of the vertex-wise cortical thickness map (mm) of a randomly selected infant at 9 months of age by five different methods.....	48
Figure 3.8. Illustration of the longitudinal infant dataset used in this study.....	49
Figure 3.9. Overview of the missing data estimation method.....	50
Figure 3.10. Estimations of the vertex-wise missing cortical thickness at 9 months of age for a randomly-selected infant.....	52
Figure 3.11. Average vertex-wise errors (mm) in estimating the missing cortical thickness at 9-months of age for all subjects at each step of estimation.....	53
Figure 3.12. Average vertex-wise errors (mm) in estimating the missing cortical thickness at 5 time points for all subjects by using the proposed method.....	54
Figure 3.13. Error measures in 36 ROIs for estimation of missing cortical thickness at 9 months of age.....	56
Figure 3.14. The group-average error maps of each step in estimating the missing cortical thickness at 1, 3, 6, 9 and 12 months of age in 36 ROIs.....	57
Figure 3.15. Relationship between estimation errors and the portions of missing data.....	58

Figure 4.1. Sulcal pits in the case of (a) over extraction and (b) expected extraction/result.....	64
Figure 4.2. A 2D schematic illustration of the ridge height. The ridge height is defined as the sulcal depth difference between a ridge point (green point) and a candidate sulcal pit (red or blue point).....	64
Figure 4.3. Relationships between optimal parameters and cortical surface metrics.....	66
Figure 4.4. Sulcal pits extraction using different depth thresholds. T_0 denotes the “optimal” depth threshold.....	67
Figure 4.5. Sulcal pits extraction using different distance thresholds (rings).....	67
Figure 4.6. Sulcal pits extraction using different area thresholds. A_0 denotes the optimal area threshold.....	67
Figure 4.7. Sulcal pits extraction using different ridge height thresholds.....	68
Figure 4.8. Sulcal pits extraction results on the left hemisphere of a representative infant at 0, 1 and 2 years of age.	69
Figure 4.9. Spatial distributions of sulcal pits on both left and right hemispheres from 73 infants at 0, 1, 2 years of age and 64 young adults.....	71
Figure 4.10. Overview of the method for discovering major sulcal patterns.....	73
Figure 4.11. Sulcal patterns in the central sulcus.....	79
Figure 4.12. Sulcal patterns in the superior temporal sulcus.....	80
Figure 4.13. Sulcal patterns in the cingulate sulcus.....	81
Figure 4.14. Examples of different central sulcus patterns.....	82
Figure 4.15. Different numbers of clusters in the cingulate sulcus.....	84
Figure 4.16. Sulcal pattern related feature maps.....	87

LIST OF ABBREVIATIONS

DARF	Dynamically-Assembled Regression Forest
MEM	Mixed Effect Model
NMSE	Normalized Mean Squared Error
MAE	Mean Absolute Error
MRE	Mean Relative Error
PR	Polynomial Regression
CRF	Conventional Regression Forest
SLR	Sparse Linear Regression
ROI	Region of Interest
APC	Affinity Propagation Clustering
SNF	Similarity Network Fusion

1 INTRODUCTION

1.1 Introduction

The early postnatal period is a crucial time for brain development, because many basic brain structures and functions are established rapidly during this period. Understanding the brain development during this period is a fundamental problem in neuroscience. Cerebral cortex, as an important component of the human brain, plays a key role in receiving sensory inputs (e.g., vision and hearing), body movements, and complex intellectual activities (e.g., memory, language, judgement, and emotion), and grows dramatically during the early postnatal period. Therefore, studying the cerebral cortex and its early development is important for understanding early brain development.

One of the challenges in studying cerebral cortex is that cerebral cortex is extremely thin (1~5mm) and highly folded, which makes it difficult to display, observe, measure, and model directly from MR images, as shown in **Figure 1.1(a)**. Therefore, cortical surface-based analysis, in which a cerebral cortex is explicitly represented as a mesh surface, as shown in **Figure 1.1(b)**, has been proposed and widely used in the study of cerebral cortex. By rendering the triangular mesh in 3D space, the cortical surface is clearly displayed and is easy to observe. An important advantage of cortical surface-based analysis is that it enables researchers to study the attributes of cerebral cortex vertex-by-vertex. There are many neuroscientifically and clinically important attributes of the cerebral cortex, such as cortical surface area, cortical thickness, sulcal depth, cortical gyrification, and cortical folding patterns, which may be correlated to the brain functions

and neurodevelopmental disorders, yet have not been well studied. The goal of this dissertation is to develop cortical surface-based methods for better studying the attributes of the cerebral cortex during early brain development. Specifically, this dissertation aims to develop methods for 1) estimating the development of morphological attribute maps of the cerebral cortex and 2) discovering the major cortical folding patterns.

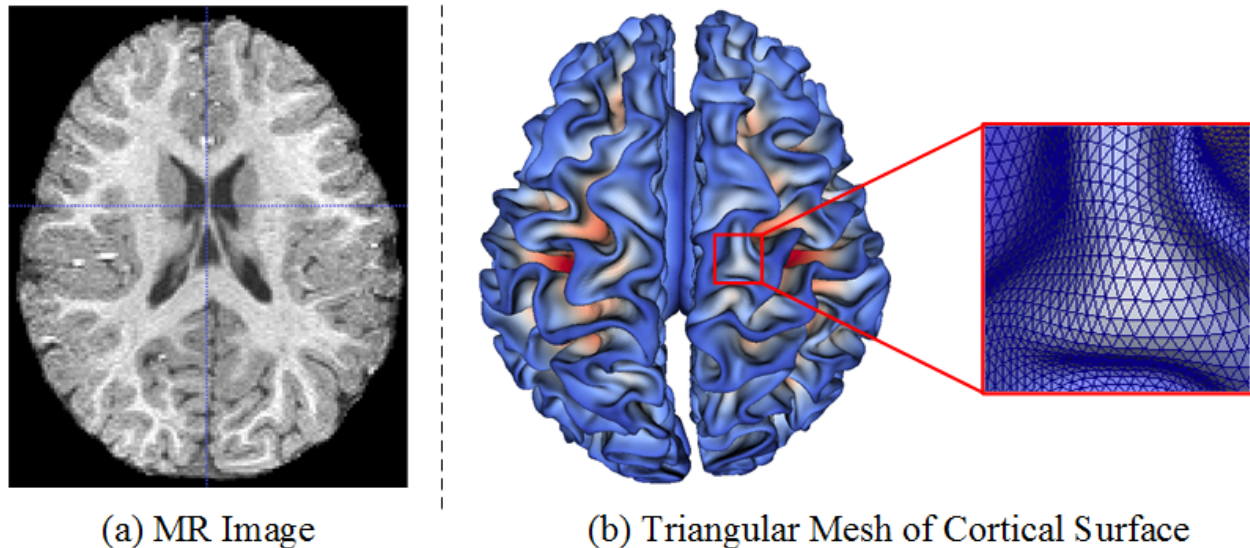


Figure 1.1. Representations of the cortical surface. (a) T1-weighted MR image of an infant brain. (b) The interface between white matter and gray matter is represented by a triangular mesh.

To provide an overview of the above research topics and their related state-of-the-art techniques, the rest of this chapter is organized as follows. The topic of estimating the development of cortical morphological attributes is introduced in **Section 1.2** and **Section 1.3**. Specifically, **Section 1.2** introduces the prediction model of cortical attributes development. **Section 1.3** introduces how to recover the missing cortical attributes in incomplete longitudinal datasets, which is an important sub-problem for estimating the developmental cortical attributes. The topic of exploring the major cortical folding patterns is discussed in **Section 1.4**.

1.2 Predicting the Development of Vertex-Wise Cortical Attributes

1.2.1 Significance

Accurately modeling the development of cortical attributes (cortical area, cortical thickness, sulcal depth, gyrification index, etc.) is crucial for better understanding the mysterious and dynamic early brain development. First, the change of cortical attributes is highly correlated to the brain functional development. For example, it has been reported that the changes of both cortical area and cortical thickness over time are related to intelligence (Schnack, et al., 2015); it is also suggested that gyrification index is a symbol that is correlated with the brain capacity of information processing (Tallinen, et al., 2014). Second, the change of cortical attributes could be an important indicator for some neurodevelopmental disorders. For example, the delayed development of cortical thickness is observed in the brains with attention-deficit/hyperactivity disorder (Shaw, et al., 2007); cortical surface area could be used to predict the diagnosis of autism (Hazlett, et al., 2017); and locally increased gyrification index could be a predictor for schizophrenia (Budday, et al., 2014). Therefore, accurately modeling the changes of cortical attributes of infant brains can benefit the studies of early brain development, and is also potentially helpful for predicting the neurodevelopmental disorders in relatively early stage.

1.2.2 Challenges

First, the changes of cortical attributes in infants are dramatic during the first two years of life. For example, from birth to one year of age, the average cortical thickness increases 40%, and cortical surface area increases 80% (Li, et al., 2013; Lyall, et al., 2015). In order to precisely capture such rapid developments, both intensive data acquisition and sophisticated prediction models are required.

Second, the development of cortical attributes is very complex, with different changing rates at different ages. For example, cortical thickness increases very fast from 3 to 6 months, but changes slowly from 9 to 12 months (Li, et al., 2014a). Such a temporally varied developing rate

makes it very difficult to find an explicit mathematical expression that can accurately fit the developmental trajectory of cortical attributes.

Third, the development of cortical attributes is highly regionally heterogeneous. As reported in (Li, et al., 2014a), the increase of cortical thickness in the central sulcus mostly happens from 6 to 9 months; and the increase of cortical thickness in the occipital cortex mostly happens from 0 to 9 months; but for the frontal, temporal, and parietal cortices, cortical thickness grows continuously during the first 18 months of life. Such regional differences make it difficult to build a single model to fully capture the development of cortical attributes for the entire cortex.

Because of the above challenges, accurately modeling the development of cortical attributes vertex-by-vertex is a very difficult problem, where both temporal-varied changes and regional heterogeneity must be carefully encoded.

1.2.3 Existing Methods

The existing methods, which are related to the goal of predicting the vertex-wise development of morphological attributes of the cortical surface, fall into two categories.

The first category is to use mathematical curves to fit the development trajectory of cortical attributes. Usually, such methods try to fit the development trajectory curve for each vertex on the cortical surface using multiple predefined mathematical models and then select the best-fit one. For example, in (Ducharme, et al., 2016), the linear mixed-effects, quadratic mixed-effects, and cubic mixed-effects models were all used to fit the development trajectory of cortical thickness vertex-by-vertex, for the subjects from 4.9 to 22.33 years of age. Then it was found that in most cortical regions the development trajectories of cortical thickness could be fitted using a monotonically decreasing linear model, while in a few cortical regions the trajectories were better fitted using quadratic or cubic models. The similar methods were also applied to modeling the

development of cortical surface area in the same age range (Ducharme, et al., 2015; Wierenga, et al., 2014). Another study used linear, quadratic, and logarithmic models to capture the development curves of cortical thickness and surface area from 1 to 6 year of ages, and found that generally the development followed logarithmic trajectories, while several exceptional regions followed linear and quadratic trajectories (Remer, et al., 2017). All of these modeling studies focused on the period of older children, adolescence, and young adults, when the development of cerebral cortex is much slower compared to that in the first postnatal year, and therefore even a simple linear model could fit the development trajectory well in their studies. However, from birth to one year of age, cerebral cortex grows dramatically, nonlinearly, and even not monotonically in some cortical regions, and thus the simple linear models (or quadratic, cubic, logarithmic, etc.) would fail to precisely represent the complex changes of the cortical attributes.

The second category of methods, instead of directly modeling the development of cortical attributes, are to predict the shape development of the cortical surface, and thus the cortical attributes could be indirectly computed based on the predicted cortical surface. Computational growth model is one of the widely used models for capturing cortical development or simulating the mechanisms of cortical folding (Budday, et al., 2014; Nie, et al., 2010; Nie, et al., 2012). Usually, in a computational growth model, the deformation of the cortical surface is considered as a result of a series of driving forces (e.g., elastic forces and growth forces), and how these forces drive the cortical surface is solved numerically, e.g., using the finite element method. Another way to predict the cortical surface evolution is using the varifold-based/current-based shape morphing-learning method (Rekik, et al., 2015; Rekik, et al., 2016). Specifically, in the training stage, a shape regression model is used to fit the trajectory of the geodesic cortical shape development for each training subject, and then a spatiotemporal surface atlas is built; in the testing stage, using the best

learned features, the cortical surface shape of a given subject is predicted based on the similarity between the given cortical surface and an age-corresponding surface atlas. Although these models are able to predict the shape development of the cortex, which can be used to further derive the development of cortical attributes, such an indirect way of cortical attributes prediction has two marked limitations. First, the cortical surface is a complex manifold and its mechanism of development from birth to 1 year of age has not yet been well understood, and thus predicting the early development of the cortical shape is quite challenging, and also the prediction accuracy of existing methods has not been widely accepted. In particular, the errors introduced from the shape prediction process would directly affect the accuracy of the measurement of cortical attributes. Second, some of the cortical surface attributes (e.g. cortical thickness) can only be measured when both an inner cortical surface (the interface between white matter and gray matter) and an outer cortical surface (the interface between gray matter and cerebrospinal fluid) are given. However, most of the existing methods are designed only for predicting the evolution of inner cortical surfaces, but have not been validated on outer cortical surfaces.

On the other hand, there is an advanced machine learning method, namely random forest (Breiman, 2001) or random decision trees (Criminisi, et al., 2012), which has been widely used to do the nonparametric regression for the complicated data. It has advantages of seeking for the hidden complex nonlinear relationship between the input data and the regression target, which is difficult to describe using explicit mathematical expressions. Random forest is naturally convenient to parallelize and very easy to scale up, and thus quite suitable for handling massive amount of data. Random forest is also relatively easy to train, because it owns only few parameters and the algorithm itself is robust. Unfortunately, random forest has never been utilized for modeling the development of cortical attribute maps. To take these advantages of random forest,

in this dissertation, the conventional techniques of random forest have been specifically adapted for the application of predicting the development of cortical attribute maps. Since understanding random forests is very crucial for comprehending this dissertation, the basis of using random forest to do nonparametric regression will be detailed in the next chapter.

1.3 Estimating the Missing Cortical Attributes

1.3.1 Significance

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 is 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., 2015b), at each time point there are different number of subjects due to the missing data. As a result, different biases could be introduced to different time points, thus leading to low accuracy and also longitudinal inconsistency in the subsequent analysis, especially at the time points with limited number of scans. On the other hand, discarding subjects with missing time point(s) is a terrible waste of the useful information, which is acquired at a considerable cost, and may also lead to insufficient amount of data for launching the study. For example, to predict the development of cortical attributes using learning-based methods, the prediction model can be more effectively trained using a larger complete longitudinal dataset than using a smaller one. But, due to the missing data, even though the total number of subjects is large, the number of subjects with complete scans is much smaller, especially when more time points are involved in the training process. This induces the lack of training data, and consequentially leads to ineffective training

and limits the power of the obtained prediction model. If the missing data can be somewhat recovered, then more available data could be used to train the prediction model, and as a result the prediction precision could be improved. Because of these reasons, accurate estimation of the information at missing time points plays an important role in longitudinal analysis of early brain development.

1.3.2 Challenges

The first challenge is that the portion of missing data could be very large. For instance, in our longitudinal dataset, each subject is scheduled to have 11 MR scans from birth to age 6, with every 3 months during the first year after birth, every 6 months during the second year, and every 12 months till to the sixth year. However, because of the absence from the scheduled scans or poor imaging quality, only 34% subjects have fully available data in the first year, 12% subjects have fully available data in the first two years, and 2% subjects have fully available data for all time.

The second challenge is that the missing data are the cortical attribute maps for the entire brain. In longitudinal cortical attributes dataset, for each subject at each time point, there are several cortical attribute maps (e.g., cortical thickness map, sulcal depth map, gyrification index map, etc.), and in each cortical attribute map there are hundreds of thousands of measured values that are sampled across the whole brain. If a time point is missing for a subject, it means all cortical attribute maps with hundreds of thousands of values at this time point are totally lost. These characteristics of longitudinal cortical attributes datasets leave very few hints for inferring or recovering the missing data.

1.3.3 Existing Methods

An intuitive 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). This strategy works well when the missing data is not too much and randomly distributed, but its efficiency decreases fast with the increase of missing data, and it has very limited ability to address the continuously distributed missing data (e.g., the entire neighbor is missing). One reason of such limitation is that this strategy only uses local information to estimate the missing data, but does not take the advantage of the global relationship between the missing data and the known data. To better utilize the global information to estimate the missing data, several methods based on low-rank matrix/tensor completion were proposed (Cai, et al., 2010; Candes and Recht, 2009; Candes and Tao, 2010; Liu, et al., 2013). Such methods assume that the data matrix/tensor is low-rank and the missing elements are distributed randomly or evenly, thus the missing elements could be filled with the values that minimize the rank of the matrix/tensor. These methods work well for the recovering the missing pixels in the images with repetitive patterns. As reported in (Liu, et al., 2013), although the portion of missing pixels can reach more than 80%, the methods are still able to reconstruct the original image. On the other hand, the low-rank assumption may not be rational for an image with rich information but no repetitive patterns. Another limitation is that if the missing data is not distributed randomly but concentrated into large blocks, these methods would no longer be applicable. Unfortunately, in neuroimaging studies, the missing data are usually distributed in blocks but not uniformly (Thung, et al., 2014; Yuan, et al., 2012). For example, the missing data could be an entire image (Li, et al., 2014d; Yuan, et al., 2012) or an entire cortical attribute map with hundreds of thousands of values. Recently, deep convolutional neural networks (Li, et al., 2014d) were applied to estimating information of missing imaging modalities based on the available modalities for improving brain disease diagnosis. However, this image-based method

cannot be directly applied to our task of estimating missing vertex-wise cortical attribute maps on the dynamic developing cortical surfaces.

1.4 Discovering Major Cortical Folding Patterns in Neonates

1.4.1 Significance

Cortical folding pattern is another important cortical morphological attribute. Although the cortical folding of the human brain exhibits highly variable forms in different brains, certain common folding patterns do exist in some specific cortical regions (Ono, et al., 1990). However, the number of common folding patterns has not been well summarized, and what each common folding pattern looks like has not been well studied. Furthermore, the biological meanings behind these common folding patterns is still not clear. Although several studies targeting adult brains have tried to explore the major folding patterns, to the best of my knowledge, there is no previous study on the cortical folding patterns in neonatal brains, which is of great importance both in neuroscience and methodology development for neuroimaging studies. First, the primary cortical folding is largely genetically determined and has been established at term birth (Li, et al., 2013), thus neonates with the minimal exposure to the complicated postnatal environmental influence are the better candidates for discovering the major cortical patterns than adults. Second, the general structural differences of folding patterns in the cerebral cortex may be related to the distinct cognitive functions (Cachia, et al., 2016; Sun and Hevner, 2014), and may also have some connections with neurodevelopmental disorders (Im, et al., 2013b; Im, et al., 2016). These hypotheses could not be fully tested without knowing the major categories of cortical folding patterns and their differences. Third, mining the major cortical folding patterns can benefit other neuroimaging analysis techniques that rely on the shape of the cortex, such as cortical surface registration and cortical surface atlas reconstruction. For example, in the current group-level

analysis (Li, et al., 2013), typically a single cortical atlas is used for a group of brains. Such an atlas may not be able to reflect some important patterns of cortical folding due to the averaging effect, and thus lead to poor registration accuracy for the subjects that cannot be well characterized by the folding patterns in the atlas. If major cortical folding patterns are well known, then multiple atlases, in which each of them represents one major pattern of cortical folding, could be used. This could boost the accuracy in cortical surface registration and subsequent analysis.

1.4.2 Challenges

First, discovering major cortical folding patterns requires a large-scale dataset, because small datasets may not be able to sufficiently cover all kinds of major cortical patterns and may lead to biased results. However, it may take many years to collect enough subjects to build a large-scale dataset, especially for healthy neonates. Thanks for the years of efforts from UNC School of Medicine and UNC hospitals, the size of the dataset is sufficiently large when I conducted this research.

Second, the structure of the cerebral cortex is so complex and variable across individuals that it is quite difficult to quantitatively characterize the folding patterns using a comprehensive mathematical measurement. Although some geometric metrics such as local mean curvature, local gyrification index (Schaer, et al., 2012), shape index (Koenderink and Vandoorn, 1992), or surface complexity index (Kim, et al., 2016) have been widely used to quantitatively measure the folding shapes and folding degree within a single cortex, they still have very limited abilities to accurately quantify the similarity of major patterns between individuals' cortices. For example, for the two different folding patterns with very similar mean curvature values (or gyrification index, shape index, surface complexity index), the vertex-wise comparison of these geometric metrics is very sensitive to the minor folding variance, and cannot well abstract the differences only in the major

folding patterns. Therefore, such geometric metrics are not suitable for studying the major cortical folding patterns.

Third, recent studies of sulcal pits, which are the locally deepest points on the cortical surface, provide a better way to measure the similarities and varieties of sulcal patterns across adult individuals (Im, et al., 2011b) or older children (Im, et al., 2016). However, this method cannot be directly applied to exploring sulcal patterns in infants, because several sulcal pits-related parameters are fixed for adults and older children and are not applicable for neonates. Moreover, whether sulcal pits are suitable for studying sulcal patterns in infants is still not clear, because the inter-subject consistency of sulcal pits within adults and older children has not been well examined for infants.

1.4.3 Existing Methods

The existence of common sulcal patterns in some specific cortical regions are first reported in (Ono, et al., 1990), which examined 25 autopsy specimen adult brains visually. The sulcal patterns were categorized based on their locations, shapes, sizes, and dimensions, measured on the real brains specimen, instead of using modern advanced MRI techniques. Despite the limited number of samples, this study opened a new window for the future exploration and research on the shape variability of the human cortical surface.

Recently, owing to the successful application of MRI, cortical surface-based methods have been widely used for investigating the folding patterns, freeing researchers from doing body dissection. In a study of 150 adult brains (Sun, et al., 2007), 3D moment invariants were used to represent each sulcus, and major sulcal patterns were categorized using agglomerative clustering algorithm. The authors further found that the discrimination of 3D moment invariants was limited in distinguishing different patterns, and hence proposed a more representative method, which

integrates the Iterative Closest Point (ICP) algorithm (Besl and McKay, 1992) and the Isomap algorithm (Tenenbaum, et al., 2000) in (Sun, et al., 2009). In this method, the corresponding sulci of two individuals were manually identified and aligned using an affine global normalization and then ICP, and their geometric distance was computed, producing a distance matrix. Then, Isomap algorithm was performed to reduce the dimension of the distance matrix. Finally, the major sulcal patterns were categorized using a dedicated hierarchical clustering algorithm. This method was applied to a dataset of 62 adult brains, and found three patterns in the left superior temporal sulcus, four patterns in the left cingulate region, and three patterns in the left inferior frontal gyrus. However, since this method highly relies on the affine registration, using the distance between two aligned cortical regions as a similarity metric is too sensitive to the shape of cortices. For example, if one cortical region is straight and the other is relatively bended, even though they share the same major folding pattern (e.g., both with two equal sized sulcal basins), such distance-based measurement would still report a relatively small similarity, and therefore they cannot be categorized into the same folding pattern.

Later on, an atlas-based method was proposed to study the folding patterns in the left inferior frontal sulcus (Coulon, et al., 2012). Specifically, this method divided the template left inferior frontal sulcus into six pieces. For each piece, the centroid was computed, and the average orientation was estimated using a principal component analysis (PCA). Then, each individual's left inferior frontal sulcus was nonlinearly aligned to the template, and the closest points to the six centroids of the template were identified. The distance between each closest point to its corresponding centroid was computed, as well as the dot product between the normal direction of the closest point and the orientation of the corresponding piece of the template. The six distance values and the six dot products were used as a feature vector for characterizing the individual left

inferior frontal sulcus. Finally, a clustering algorithm was performed to identify the major classes based on the feature vectors of all subjects. This method was applied to a dataset of 151 brains and detected five major sulcal patterns in the left inferior frontal sulcus. The advantage of this method is that the nonlinear registration onto the template (atlas) removes the non-interest variability among the subjects. However, the disadvantage is that the template may introduce biases, because the template may only reflect the most common patterns existing in the majority of the population. For a relatively smaller subset of population with a very different common pattern, the nonlinear registration could unpredictably align the individual local cortices, and thus the feature vectors referring to the template may lose the ability to represent this common pattern.

Recently, sulcal graph-based methods were proposed to compare the sulcal pattern similarities between individual subjects (Im, et al., 2011b; Im, et al., 2013b; Im, et al., 2016) and achieved meaningful results. Sulcal graph is an abstract representation of the geometric structure of the cerebral sulci. It is built based on the sulcal pits and the neighboring relationships between the sulcal basin regions of sulcal pits. Specifically, sulcal pits are the locally deepest points on the cerebral cortex, and can be extracted using a watershed algorithm (Im, et al., 2010). Sulcal pits have been suggested to be genetically affected and closely related to functional areas (Im, et al., 2010; Lohmann, et al., 2008; Regis, et al., 2005), and they are thought to be the first occurrences of cortical folds during the cortical folding development (Im, et al., 2011b). Although sulcal folding is highly variable across individuals, the spatial distribution of sulcal pits is relatively consistent, so sulcal pits are well suitable as reliable landmarks for characterizing sulcal folds. In these methods, a sulcal graph was constructed for a local cortical region of an individual brain. The nodes in the sulcal graph were the sulcal pits, and two nodes were linked by an edge if their corresponding sulcal basins were spatially connected. Then, for the pair of sulcal graphs that need

to be compared, a spectral matching technique (Leordeanu and Hebert, 2005) was performed to optimally matching the two sulcal graphs. Finally, the quantitative similarity of the matched parts in these two sulcal graphs was computed. This method was applied to a dataset of 48 young adult healthy twin pairs, and found significantly higher similar sulcal patterns in twin pairs than in unrelated pairs (Im, et al., 2011b). This method has also been used to quantitatively compare the sulcal patterns between the normal and abnormal brains, and successfully found significant differences (Im, et al., 2013b; Im, et al., 2016). All of these results suggest that sulcal graph would be a reliable tool to compare sulcal patterns. However, sulcal graph has not yet been applied to discovering the major common cortical folding patterns in a large dataset.

1.5 Thesis

Thesis: Dynamically-Assembled Regression Forest (DARF) is able to accurately estimate the early development of cortical attribute maps from birth to 1 year of age. Sulcal pits, which have relatively consistent spatial distributions across ages and individuals, can be utilized for discovering the major sulcal patterns. Sulcal pattern information can further improve the performance of DARF for estimating cortical attribute maps.

This dissertation aims to develop methods to address the challenging and important problems in cortical attribute analysis in infant brains. In particular, the dissertation mainly focuses on solving two following problems.

Aim 1: Estimating the Development of Cortical Attribute Maps. *Dynamically-Assembled Regression Forest (DARF) is able to accurately estimate the early development of cortical attribute maps from birth to 1 year of age.* The existing methods of modeling the development of cortex or cortical attributes of young adults, either use linear/nonlinear polynomial expression to fit the development curve or indirectly compute the cortical attributes based on the predicted

cortical shape, as introduced in **Section 1.2.3**. Such methods are not suitable for precisely estimating the vertex-wise cortical attribute development in infants, because cortical attributes change fast, nonlinearly, regionally heterogeneously during the first year of life. Random forest, as a parallel, scalable and flexible machine learning technique, has intrinsic advantages in nonparametric regression. But it cannot be directly applied to the prediction of cortical attribute development, due to the complexity of the development itself. Therefore, this dissertation adapts random forest for estimating the early development of cortical attribute maps. In order to achieve this aim, this dissertation makes the following detailed contributions.

- (1) A novel prediction model, Dynamically-Assembled Regression Forest (DARF), is proposed to accurately estimate the early development of cortical attribute maps from birth to 1 year of age. Conventional regression forests have limited ability in estimating the development of cortical attributes. Different from conventional regression forests, DARF trains a single decision tree at each vertex in the training stage, and locally groups decision trees around each vertex as a vertex-specific forest in the testing stage. Since a DARF handles only a local cortical region, the regression problem is simplified and thus can be solved effectively. Moreover, because one decision tree can be shared by many nearby DARFs, the estimation results would be spatially smooth, and the algorithm is computationally efficient.
- (2) A surface-based feature computation method is proposed for extracting Haar-like features from a spherical surface. Haar-like features provide rich neighboring information of local cortical attribute maps, and help DARF better utilize the correlation between the prediction target and the local input feature maps. Thus, Haar-like features are very useful in training more accurate prediction models.

- (3) A novel missing data estimation strategy is proposed for approximately recovering the missing cortical attribute maps in incomplete longitudinal datasets. The strategy consists of two stages. The first stage, namely pairwise estimation, performs estimation by utilizing the relationship between the missing time point and each of other available time points independently, so that as many subjects as possible can contribute to the estimation. Then the independent outputs are averaged as an initial estimation. Benefiting from this initial estimation, in the second stage, namely joint refinement, the relationship between the missing time point and all other available/initially estimated time points can be utilized together for the estimation. In this way, all available information in the dataset could make a contribution to the estimation, so that the missing data is recovered more precisely.
- (4) Quantitative evaluations and comparisons are performed. First, the results show that the proposed prediction model, DARF, outperforms many existing methods in estimating the development of cortical attributes. Second, experiments indicate that the proposed missing data completion strategy is able to make a better use of the available data and to effectively recover the missing cortical attribute maps.

Aim 2: Discovering the Major Cortical Folding Patterns. *Sulcal pits, which have relatively consistent spatial distributions across ages and individuals, can be utilized for discovering the major cortical sulcal patterns. Sulcal pattern information can further improve the performance of DARF for estimating cortical attribute maps.* The sulcal pattern descriptors, used in the existing methods for major sulcal patterns discovery, have limited ability of distinguishing different patterns. Recent studies of sulcal pits have found that the sulcal graph, consisting of sulcal pits and their connections, is able to measure the similarity between sulcal patterns. Therefore, this

dissertation proposes to utilize sulcal graph as sulcal pattern representation to discover the major sulcal patterns from a large-scale dataset of neonatal brains. For this purpose, the detailed contributions are listed as follows.

- (1) A watershed algorithm is employed and adapted for extracting sulcal pits from cortical surfaces of infants at different ages. Since the parameter configurations in the watershed algorithm is sensitive to the brain size, and the brain sizes of infants are much smaller than those of adults and are growing rapidly, the original configuration tuned for the adult brains cannot be directly applied to the brains of infants at different ages. This dissertation proposes a general way to automatically set the parameters in extraction of sulcal pits. By using the correlation between the parameters and the brain size, the adapted algorithm is able to generate consistent results for the infant brains at any age.
- (2) The spatial distribution and longitudinal development of sulcal pits is studied using a longitudinal dataset of 73 infant brains from birth to 2 year of age. Consistent with previous studies of sulcal pits in adults, sulcal pits distribution is relatively *spatially and temporally stable*, and thus sulcal pits are suitable to be used as landmarks for cortical surface-based analysis.
- (3) A new method is proposed for discovering the major sulcal patterns from a large-scale dataset of 677 neonatal brains. The proposed method takes advantage of sulcal graphs, and evaluates the similarity between the local sulcal patterns of any two neonates in six different points of view. All six similarities are adaptively fused together, and fed into a hierarchical affinity propagation clustering algorithm. By averaging the most representative samples in each cluster, the proposed method successfully discovers

multiple major patterns in three important brain regions, including the central sulcus, superior temporal sulcus, and cingulate regions.

- (4) Whether sulcal pattern information could help better predict the development of cortical attribute maps is also investigated. To this end, five features maps, which encode descriptive information of sulcal patterns, are computed and fed into DARF for the prediction. The results show that the sulcal pattern related feature maps could improve the prediction accuracy in certain cortical regions.

1.6 Overview of Chapters

The rest of this dissertation is organized as follows.

Chapter 2 introduces the necessary background of the related dataset, techniques, and concepts used in this dissertation. For the dataset, the basic information about data acquisition is provided, and then the procedure of data processing is detailed. For techniques, the basic idea of how to use random forest to do nonparametric regression is introduced; then the effects of parameters in random forest is discussed, followed by a summary of the advantages and disadvantages of random forest. Finally, the concept and important characteristics of sulcal pits are presented, and the recent studies of sulcal pits are introduced.

Chapter 3 presents the methods for estimating the development of cortical attributes. First, Dynamically-Assembled Regression Forest (DARF) is proposed for accurately predicting the cortical attribute map at older ages based on the data at younger ages. Second, the method of extracting surface-based Haar-like features is presented. Third, a novel missing data estimation strategy based on DARF is proposed for approximately recovering the missing cortical attribute maps. This chapter also includes extensive experiments for evaluating the proposed methods and comparing with the existing techniques.

Chapter 4 presents the methods for discovering the major sulcal patterns. Specifically, it presents in detail the method of extracting sulcal pits from infant cortical surfaces. Next, it investigates the spatial distribution of sulcal pits and their temporal development. After that, the techniques used for discovering major cortical sulcal patterns are introduced. The discoveries and the results of evaluation experiments are reported. Finally, whether sulcal pattern information could help DARF better predict the development of cortical attribute map is investigated.

Chapter 5 concludes the dissertation, summaries its contributions, and also points out future directions.

1.7 Summary

This chapter provided an overview of the works in this dissertation. This dissertation focuses on addressing two important and challenging problems in cortical surface-based analysis of infant brains. The first problem is to accurately estimate the rapid development of cortical attributes of infants, and it includes two sub-problems: development prediction and missing data estimation. The second problem is to effectively discover the major cortical sulcal folding patterns. For each problem or sub-problem, the significances, challenges, and existing methods were summarized. This chapter also announced the thesis and sketched the solutions proposed in this dissertation. Finally, this chapter presented the organization of the remaining chapters. Next, I will present some background information in **Chapter 2**.

2 BACKGROUND

This chapter presents the necessary background for understanding the rest of this dissertation. **Section 2.1** provides the basic information of the datasets used in this dissertation. **Section 2.2** introduces the processing of raw data for producing the surface-based datasets that can be directly used in the experiments in this dissertation. **Section 2.3** presents an advance regression technique, regression forest. **Section 3.3** introduces sulcal pits and the related recent studies.

2.1 Datasets

Three datasets are used in this dissertation. The first dataset, which is used for estimating the development of cortical attributes (**Chapter 3**), is a longitudinal dataset of 47 infant brains. Each infant has a scheduled scan every three months from birth to 1 year of age. However, due to the missing data, only 36 out of the 47 brains are used in this dissertation. The second dataset, which is used for studying the spatial distribution and longitudinal development of sulcal pits (**Section 3.1**), is a brain dataset mixed with 73 infants and 64 young adults. Each infant in this dataset has three longitudinal scans obtained, respectively, at birth, 1 year of age, and 2 years of age, and each young adult has one scan acquired around 18.9 ± 1.4 years of age. The third dataset, which is used for discovering the major sulcal patterns (**Section 3.2**), is a large-scale dataset with 677 neonatal brains.

All infant scans used in this dissertation were collected by the UNC hospital. Specifically, the UNC hospital 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 without 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.

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 = 256×192 , 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, acquisition matrix = 256×128 , 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 = 7680/82ms, acquisition matrix = 128×96 , voxel resolution = $2 \times 2 \times 2 \text{ mm}^3$, 42 non-collinear diffusion gradients, and diffusion weighting $b = 1000 \text{ s/mm}^2$.

The young adult data were obtained from a subset of the public pediatric MRI data repository (Release 4.0) created for the NIH MRI study of normal brain development (Evans and Brain Development Cooperative, 2006). Each participant was scanned using a 1.5T scanner with T1-weighted Spoiled Gradient Recalled (SPGR) echo sequence. Slice thickness of 1.5 mm was allowed for GE scanners due to their limit of 124 slices. Total acquisition time was about 25 min, and was often repeated when indicated by the scanner-side quality control process. More details on image acquisition can be found in (Evans and Brain Development Cooperative, 2006).

2.2 Data Preprocessing

All infant MR images were preprocessed using an infant-specific cortical surface pipeline, which has been validated on 2000+ infant brain MR images. As shown in **Figure 2.1**, the pipeline

includes three main steps: image processing, cortical surface reconstruction, and cortical surface registration. Each step is introduced briefly next.

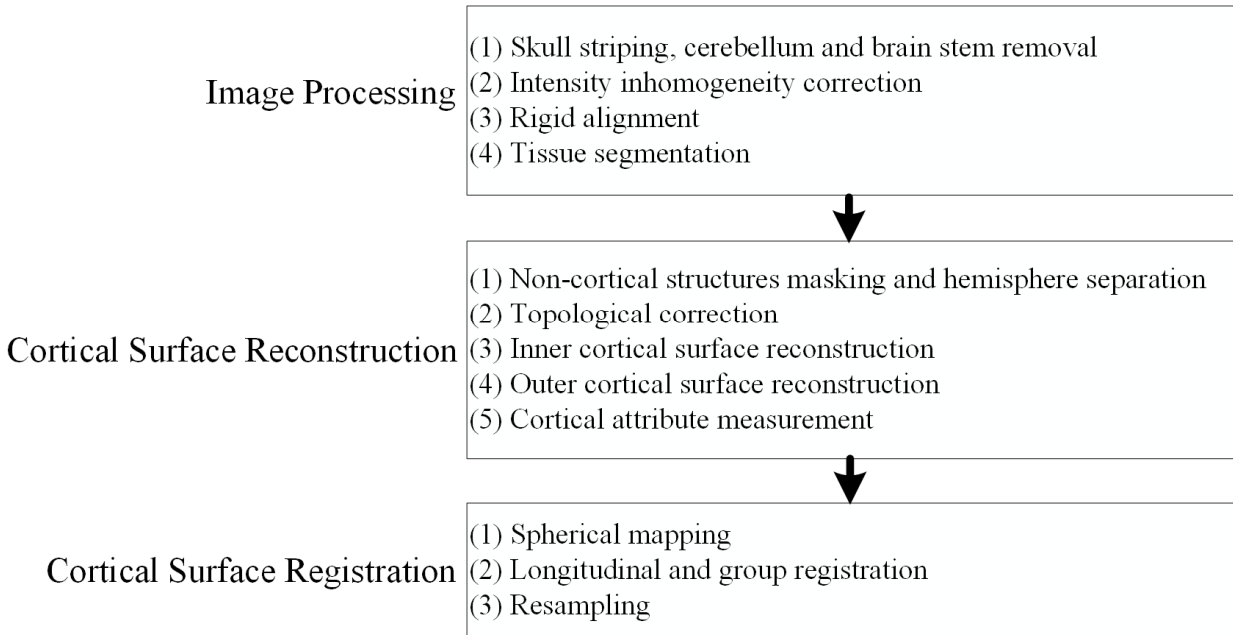


Figure 2.1. Data preprocessing pipeline.

2.2.1 Image Processing

All MR images were processed as follows. First, skull, cerebellum, and brainstem were automatically removed (Shi, et al., 2012). Second, intensity inhomogeneity was corrected using N3 method (Sled, et al., 1998). Third, each image was rigidly aligned to the age-matched brain atlas (Shi, et al., 2011) using FLIRT in FSL (Smith, et al., 2004). Fourth, tissue segmentation of MR images was performed by using a coupled level-sets method, to classify the voxels into white matter, gray matter, and cerebrospinal fluid (Wang, et al., 2014b). Note that, if one subject has multiple MR images that were scanned in different ages, a longitudinally-guided segmentation method is used (Wang, et al., 2013).

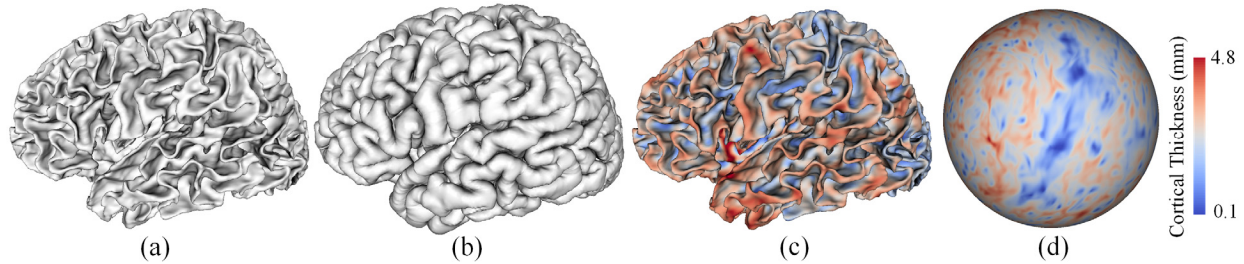


Figure 2.2. Examples of cortical surfaces and cortical thickness maps. (a) A reconstructed inner cortical surface. (b) A reconstructed outer cortical surface. (c) A cortical thickness map displayed on the corresponding inner cortical surface. (d) A cortical thickness map on the corresponding spherical surface.

2.2.2 Cortical Surface Reconstruction

The cortical surface reconstruction method took tissue segmentation results as inputs. Specifically, first, non-cortical structures were automatically masked and filled, and each brain was further separated into a left hemisphere and a right hemisphere (Li, et al., 2012). Second, for each hemisphere, the holes in the white matter segmentation mask, which were inevitable topology defects, were automatically filled using a 4D-level set algorithm (Han, et al., 2003). The other segmentation defects such as missing gyrus and incorrectly-connected handles were visually checked and manually removed, in order to get an anatomically precise and topologically correct white matter segmentation. The cerebrospinal fluid in deep sulcal regions was recovered by an anatomically consistent enhancement method (Han, et al., 2004). Third, the inner surface (the interface between white matter and gray matter) was reconstructed using a marching cube method (Lorenson and Cline, 1987) followed by a deformable surface algorithm for refinement. Fourth, the outer surface (the interface between gray matter and cerebrospinal fluid) was reconstructed by deforming the inner surface outwards while keeping its initial topology and spatially-adaptive smoothness. 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 within a local region. Once any triangle-triangle intersection was detected, the deformation

was reduced to a location without such an intersection. For an infant with multiple scans at different ages, a longitudinal-consistent deformable model was used (Li, et al., 2014a; Li, et al., 2015b) to reconstruct the inner and outer cortical surfaces. Finally, the cortical attributes were measured. For example, cortical thickness of each vertex was computed as the mean of the minimum distances from the inner surface to the outer surface and also from the outer surface to inner surface (Li, et al., 2015a) 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). Some illustrative examples of the reconstructed cortical surfaces and cortical thickness maps are shown in **Figure 2.1**.

2.2.3 Cortical Surface Registration

For facilitating surface registration and the following surface-based analysis, all cortical surfaces were mapped to a standard sphere using FreeSurfer (Dale, et al., 1999; Fischl, 2012). The spherical surfaces were registered together using a group-wise version of Spherical Demons (Yeo, et al., 2010). Each registered spherical surface was resampled using a standard mesh template with 163842 vertices, thus a vertex-to-vertex correspondence was built between any two cortical surfaces. For longitudinal dataset, the registration and resampling were performed in two stages. In the first stage, an intra-subject registration was first carried out to align all cortical surfaces of the same infant at different ages. Then the aligned surfaces were resampled and averaged to construct a subject-specific mean surface. Since the cortical surfaces of the same infant at different ages are much more similar in cortical folding patterns comparing with the cortical surfaces of other subjects, the intra-subject registration is relatively easier than inter-subject registration, and the established vertex-wise correspondence across the surfaces are more accurate. In the second

stage, the average surfaces of all subjects were registered together and resampled, so the inter-subject vertex-wise correspondence was also established.

It is worth noting that, after the processing, all cortical attribute maps have also been mapped and resampled on the standard sphere. **Figure 2.2(d)** shows an example of a cortical thickness map on the resampled spherical surface.

2.3 Basis of Regression Forest

Random forest is a powerful machine learning tool, which has been widely used in the tasks of classification, regression, density estimation, and semi-supervised learning (Criminisi, et al., 2012). Since in this dissertation the random forest is used only for the regression, I will use the name “regression forest” instead of “random forest” in the rest of this dissertation. A regression forest consists of a number of binary decision trees that are trained independently. Each binary decision tree can recursively split the data into different subgroups according to predefined split functions, and at the end a regression value is computed for each subgroup. Each tree represents a “weak learner” with a limited ability of regression, but a linear combination of many “weak learners” as a forest yields an accurate regression result. It has also been discovered that grouping a set of randomly trained decision trees with slight differences could achieve higher accuracy on previously unseen data comparing than using a single over-trained decision tree, because of the generalization.

The rest of this section is organized as follows. First, in **Section 2.3.1**, the training and testing process of a regression forest is introduced. Then, the effect of parameters is discussed in **Section 2.3.2**. Finally, the advantages and disadvantages of regression forest are summarized in **Section 2.3.3**.

2.3.1 Training and Testing

In the training stage, each decision tree in the forest is trained independently. To prevent over-fitting and increase generalization, usually each tree is trained using a subset of the whole training dataset. 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 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:

$$\arg \max_{k, \theta} H(S) - \frac{|S_l|}{|S|} H(S_l) - \frac{|S_r|}{|S|} H(S_r) \quad (2.1)$$

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 the partition; H is a metric that estimates the consistency of training samples in terms of regression target. Mathematically, H is defined as:

$$H(S) = \frac{1}{|S|} \sum_{(x_i, y_i) \in S} (y_i - \bar{y})^2 \quad (2.2)$$

where \bar{y} is the mean regression target value of all the training samples in S . The partition is determined by two factors k and θ . For the i -th sample in the training set, if the k -th feature x_i^k in the feature vector \mathbf{x}_i is less than the threshold θ , the sample is dispatched to the left child node; otherwise, it is dispatched to the right child node. To maximize the objective function **Equation 2.1**, 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 θ) 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 three terminal criteria: (1) the tree reaches a specified maximum depth, (2) the amount of training samples in a node falls below a specified minimum number, or (3) the maximum value given by objective function is

close to zero. For each leaf node, where the partition stops, a prediction model is used to compute the regression result based on all training samples falling into this leaf node. The prediction model could be in any form. The most popular prediction models includes (1) constant model, which averages the target values of training samples as the regression result, and (2) polynomial and linear model, such as $y(x) = \sum_{i=0}^n \omega_i x_i$.

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. The final result of the regression forest is the average of outputs from all decision trees.

2.3.2 Effects of Parameters

The performance of regression forest is affected by many factors. Next, I will briefly discuss some major factors, including the forest size, tree depth, weak learner model, and prediction model.

Forest Size. Forest size is the number of decision trees in a forest. Typically, increasing the number of trees would consequently make the prediction smoother and more stable. Such smoothness effect is stronger if the testing data moves away from the training data, which is a good thing for both interpolation and extrapolation. But larger number of trees also means more training and testing time. In practice, the forest size could be decided by gradually increasing the number of trees until the prediction errors stop changing.

Tree Depth. The depth of the tree controls how well the tree fits the training data. If the depth is too small, the regression forest will be under-fitting. In the extreme case that the depth is equal to one, the tree will degenerate as a linear model (if linear model is used in the leaf node) or a single constant value (if constant model is used in the leaf node), which is unable to fit a complex

training dataset. Increasing the depth would improve the fitting ability of regression forests. But a forest with too deep trees could yield over-fitting. In practice, there is often no lower bound limitation of the tree depth. The depth will keep increasing until it reaches the upper bound limitation. The upper bound limitation of the tree depth can be explicitly decided by a user-defined threshold, or can be implicitly decided by restricting the minimum sample size/information gain in the leaf node, which is a data-driven way.

Weak Learner Model. Weak learner model is the way of dispatching a training/testing sample into left or right child node. The dispatching method introduced in **Section 2.3.1** is an axis-aligned method, which partitions the feature space using an axis-aligned hyperplane. General oriented hyperplane is another usually-used weak learner model. While the advanced weak learner model could strengthen the learning ability, it would also increase the computation complexity leading to longer training and testing time. Generally speaking, it is not always good to choose an advanced weak learner model, because stronger learning ability may somewhat reduce the variability of different trees and make the regression forest lose generalization. In practice, simple weak learner models are more preferred. Advanced models should be chosen carefully based on the priori knowledge about the dataset.

Prediction Model. Prediction model is used in the leaf node for computing the regression results. Constant model and linear model are the two most popular models. While constant model is computationally efficient and produces smooth interpolation predictions, it is unable to do the extrapolation, because all the regression values stored in the leaf node are the averages of the training data subsets. Linear model could do both interpolation and extrapolation, but the uncertainty increases when the testing data move far away from the training data. In practice, the

prediction model could be chosen based on the characteristics of the data distribution in the feature space.

2.3.3 Advantages and Disadvantages

Regression forest has many advantages over other regression methods, but also has some disadvantages. Both advantages and disadvantages will be briefly summarized in this section.

Advantages

A few advantages of the regression forest are listed below.

- 1) **Parallelization.** Parallelization is an important factor to consider when selecting a machine learning algorithm. Because of the modern techniques of multi-core GPU and distributed computation system, a highly parallel algorithm always means higher computation ability and faster performance. As an ensemble model, each decision tree in a regression forest is trained independently, which makes regression forests easy to be parallelized.
- 2) **Scalability.** In the field of voxel-wise medical image analysis or vertex-wise surface-based analysis, the size of training data could easily reach millions, thus scalability is a very important factor when choosing a regression method. Due to the characteristics of binary decision tree, simply increasing the tree depth by one can double the capacity of a decision tree. In this way, regression forest can scale up efficiently. Moreover, the impact of such scaling up on the testing time is negligible, as only one layer is added in the tree.
- 3) **Flexibility.** The flexibility of regression forest is two-fold. First, the relationship between the input features and the regression target could be in any form. No matter whether the relationship is linear, piecewise linear, sparse linear, quadratic, cubic, or

in other forms, a regression forest could always fit the data without being provided explicit mathematical expressions. Second, the choices of weak learner models and prediction models are open to the users. They can use their prior knowledge about the dataset to design more specific models, and thus improve the overall performance.

Disadvantages

There are two major limitations to keep aware of when using regression forest.

- 1) Storage. When using a massive training dataset with a large-scale and complex distribution, it could take a lot of space to store the trained regression forest. So in the case of limited hardware memory resource, regression forest may not be applicable.
- 2) Interpretability. Regression forests is not a descriptive tool. Though regression forest could learn the relationship between the input features and the output regression value, the learned relationship is quite difficult to interpret. So, if the learning target is to find a descriptive relationship in the data, regression forest is not a good option.

2.4. Sulcal Pits

This section presents some background of the study of sulcal pits. I will first introduce the concept of sulcal pit and its important characteristics in **Section 2.4.1**, and then give an overview of the current research on sulcal pits in **Section 2.4.2**.

2.4.1 Sulcal Pit and Its Characteristics

Sulcal pits are the locally deepest points along sulcal bottom lines in the cerebral cortex (Lohmann, et al., 2008), as shown in **Figure 2.3**. Such deepest points have many important biological characteristics. First, during the human brain development, the deepest parts of primary sulci are thought as the first places to develop in an embryo's brain and then change the least as the cortex grows (Lohmann, et al., 2008). Second, the deepest parts of sulci are more genetically

controlled than the superficial parts (Le Goualher, et al., 1999; Le Guen, et al., 2017; Lohmann, et al., 1999; McKay, et al., 2013). Third, there are particular spatial relationships between the deepest parts of sulci and functional areas (Lohmann, et al., 2008; Piao, et al., 2004; Rakic, 1988; Smart and McSherry, 1986). Fourth, though the human cerebral cortex is highly variable across adult individuals, the spatial distribution of sulcal pits is relatively spatially consistent across human adult individuals (Im, et al., 2010; Lohmann, et al., 2008).

Because of these characteristics, sulcal pits have drawn increasing attention in neuroimaging studies in the past few years. A short review of the existing researches on sulcal pits is given in the next section.

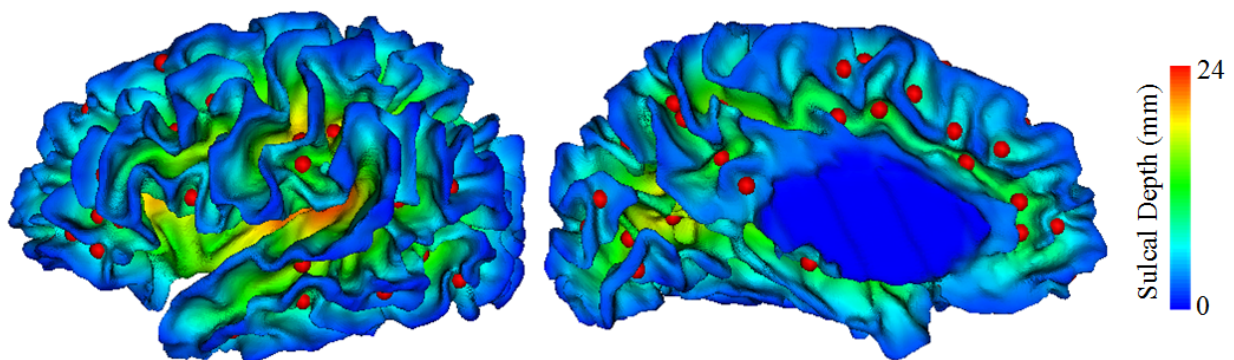


Figure 2.3. Sulcal pits. Sulcal pits are displayed as small red balls. The cortical surface is color-coded by the value of sulcal depth.

2.4.2 Existing Research

Lohmann *et al.* first examined sulcal pits in volumetric MR images and observed that the spatial distribution of sulcal pits in major sulci was strikingly regular across adult individuals, despite their highly variable cortex folds (Lohmann, et al., 2008). Then, Im *et al.* proposed a more reliable approach to extract sulcal pits from the adult cortical surfaces reconstructed from MR images (Im, et al., 2010; Im, et al., 2013a). In this approach, a watershed algorithm based on the sulcal depth was used to partition the cortical surface into many basins, and then the deepest point in each basin was identified as the sulcal pit, after pruning basins with shallow sulcal depths or

small sizes. The results in (Im, et al., 2010) confirmed the observations in (Lohmann, et al., 2008), and further revealed the hemispheric asymmetries of sulcal pits. According to these studies, sulcal pits in major cortical sulci were considered as reliable anatomical landmarks and some of them could be potentially helpful for the challenging problem of inter-subject brain MR image registration. Im *et al.* (Im, et al., 2011a) further investigated the relationships between the presence of sulcal pits and intelligence, and found that in the left posterior inferior frontal sulcus and the right posterior inferior temporal sulcus, the number of sulcal pits of young adults with high IQ was significantly different from that of young adults with average IQ. McKay *et al.* (McKay et al., 2013) specifically studied the central sulcus in adults and found that most adult individuals had two peaks in the sulcal depth position profiles, close to the hand and mouth regions, where the peak genetic heritability of the sulcal depth occurred. By tracking the cortical surface development of four neonates between birth and four weeks of age (Lefevre et al., 2009), Lefèvre et al. found that the cortical surfaces grew in a radial manner from some “growth seeds”. Le Guen *et al.* quantified the degree of how sulcal pits are under genetic controls in various brain regions (Le Guen, et al., 2017). Recently, sulcal pits have been applied to the study of psychological and neurological diseases, such as autism (Brun, et al., 2016), polymicrogyria (Im, et al., 2013b), and dyslexia (Im, et al., 2016).

2.5 Summary

This chapter provides the necessary background for understanding the rest of this dissertation. **Section 2.1** introduced the datasets used in the experiments. **Section 2.2** introduced the data preprocessing pipeline. After the preprocessing, the cortical surfaces have been reconstructed; the cortical attributes at each vertex on the cortical surface have been measured; all cortical surface have been mapped and aligned to the standard spherical surface; and vertex-to-

vertex inter-surface correspondence has been established. **Section 2.3** introduced the basis of regression forests and discussed the effects of parameters as well as the advantages and disadvantages of regression forests. As a highly parallel, scalable, and flexible regression algorithm, regression forests will be used as a core technique in **Chapter 3** for estimating the development of cortical attributes. **Section 2.4** introduced the concept of sulcal pit, its important characteristics, and its current researches. Due to the spatial and longitudinal consistency, sulcal pits will be used as landmarks in **Chapter 4** for studying the major cortical sulcal patterns.

3 ESTIMATION OF THE EARLY DEVELOPMENT OF CORTICAL ATTRIBUTES

Accurately modeling the development of cortical morphological attributes (cortical thickness, sulcal depth, etc.) is crucial for better understanding the mysterious dynamic early brain development and is also potentially helpful for the early diagnosis of neurodevelopmental disorders. This chapter focuses on the technical methods for estimating the early development of cortical attributes. To this end, two sub-problems are addressed in the chapter. **Section 3.1** presents the techniques for addressing the first sub-problem, predicting the cortical attribute maps of future time points based on the known cortical attribute maps of the early time points. **Section 3.2** presents the approach for addressing the second sub-problem, to approximately recover the missing cortical attribute maps in an incomplete dataset based on the available information. Finally, **Section 3.3** summarizes the contributions in this chapter.

3.1 Predicting the Development of Vertex-wise Cortical Attributes

In this section, a novel learning-based approach is proposed for accurately predicting the subject-specific dynamic development of vertex-wise morphological attributes in the first postnatal year, solely based on the MRI features at birth. Of note, developing such a method is challenged by the extremely dynamic and regionally-heterogeneous growth of the infant cortex, as well as by the considerable inter-subject variability of cortical morphology and developmental patterns. Technically, a Dynamically-Assembled Regression Forest (DARF) is proposed as the core regression tool to estimate cortical morphological attribute(s) at each vertex of the cortical surface.

The rest of this subsection is organized as follows. First, some background and the motivation of DARF is introduced, and then the concept of DARF is introduced by illustrating the training and testing processes. Second, I will introduce how to extract surface-based Haar-like features from cortical attribute maps to support DARFs. Third, the proposed method is tested on an application of predicting the early development of cortical thickness maps, and the advantage of the proposed method is demonstrated by comparing with the existing popular techniques.

3.1.1 Dynamically-Assembled Regression Forest

Motivation

Regression forest is a powerful regression tool, especially when it is difficult to find an explicit mathematical expression for modeling the complex relationship between the input values and output values. This intrinsic characteristic makes the regression forest quite suitable for predicting the rapid and nonlinear early development of cortical attributes. However, a single conventional regression forest (CRF) may not be sufficient for estimating vertex-wise cortical morphological attributes of the entire cortical surface. Because the cortical morphological attributes and their developments in infants are highly regionally heterogeneous, using a single CRF cannot make a precise estimation at the 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 strategy will lead to spatially unsmooth estimations around the boundaries of neighboring ROIs. This is because the cortical attributes of the vertices by the two sides of a ROI boundary are estimated using two different regression forests, which are trained independently with different training samples. Intuitively, increasing the overlapping area among ROIs could produce smoother estimation results, but it unexpectedly increases the computational cost. Based on my experiments, to make the estimation results as

smooth as the real data, more than 90% of the area of an ROI needs to be overlapped with its neighboring ROIs. Unfortunately, such a large portion of overlap requires quite a large number of ROIs in order to cover the whole cortex, and thus leads to a large computational workload, as a respective set of individual trees need to be trained for each ROI. Taking account of all these issues, I propose a Dynamically-Assembled Regression Forest (DARF). By first training a single 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 spatially smooth regression results and meanwhile also save a lot of computational cost.

Training Stage

In the training stage, one individual binary decision tree is trained at each vertex on the cortical surface. As shown in **Figure 3.1**, for the given vertex on the spherical cortical surface (mapped from the original cortical surface), one individual tree is trained using the nearby vertices within a training-neighborhood (i.e., the red region) as training samples. Each training sample can be denoted as a pair of a feature vector and a scalar regression target ($\mathbf{x}_i \in \mathbb{R}^d, y_i \in \mathbb{R}$). The feature vector \mathbf{x}_i consists of a set of features extracted from the local cortical attribute maps around the vertex i at the input time point(s) (see **Section 3.1.2**), and the scalar target y_i is the cortical attribute value of vertex i at the target time point.

Testing Stage

In the testing stage, to predict the cortical attribute value of a given vertex at the target time point, as shown in **Figure 3.1**, all nearby individual trees within a testing-neighborhood (i.e., the blue region) are grouped together to form a vertex-specific forest. The feature vector of the given vertex is computed and then fed into each individual tree of the formed forest. The prediction

result is finally computed as the average of regression outputs from all trees of the formed forest. Of note, the optimal size of testing-neighborhood can be learned via a cross validation.

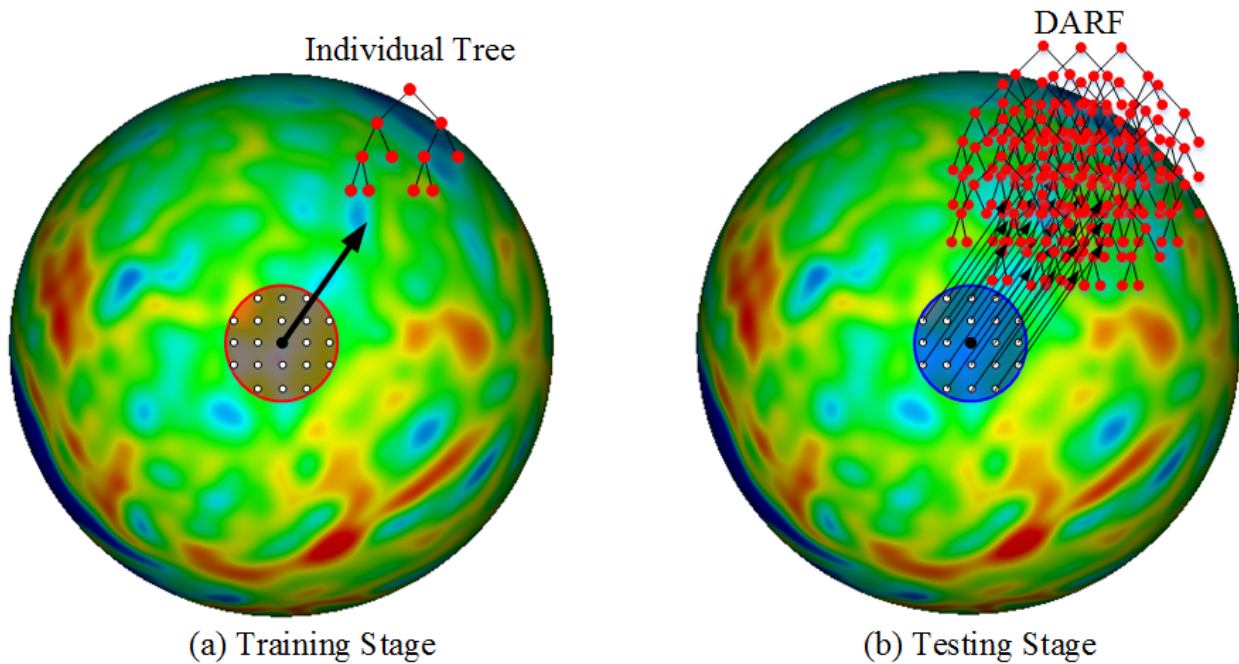


Figure 3.1. Training and testing stages for DARF. (a) The red region is the training neighborhood, where all the vertices on the spherical space are used as training samples. (b) The blue region is the testing neighborhood, where all the individual trees are combined together to form a forest in the testing stage. Note that the red and blue regions in (a) and (b) could have different sizes.

Different from the original way of using a regression forest, which trains a set of trees and makes them a fixed forest at the training stage, our method does not assign trees to any forest in the training stage. Instead, the forest is formed by group neighboring trees during the testing stage, and thus it is named “Dynamically-Assembled” Regression Forest. This novel way of forming forest has two advantages.

The first advantage is that, by using DARFs, the predicted cortical attribute map is smooth without any border-artifact near the boundary between ROIs as that in the ROI-based method. The reason is two-fold. First, as the neighboring DARFs share a large number of the same decision trees, they are very similar. Second, features of neighboring vertices, which are extracted from the

local cortical attribute maps, are also similar. By feeding the similar input features to the similar DARFs, the outputs at neighboring vertices generally have small differences, and thus the predicted cortical attribute map is smooth.

The second advantage is that, since one single decision tree can be shared by many nearby forests, the computational cost for training forests is significantly reduced. For instance, if in each forest there are 100 trees, using DARF would save 99% computational cost compared to the case of using the original regression forests (i.e., training a forest at each vertex). Even compared to the ROI-based strategy, DARF still significantly saves the computational cost in the training stage. Based on my experiments, to achieve the smooth predicted cortical attribute maps, highly-overlapped ROIs are needed, with the number of ROIs being nearly 20% of the number of vertices. If each forest owns 100 trees, using DARF would still save 19% computational cost.

3.1.2 Input Features

For each vertex i , its feature vector \mathbf{x}_i could include three types of features: demographic features, vertex-wise morphological features, and context features. Demographic features could be postnatal age, gestational age, birth weight, gender, or any other information that is believed to be correlated to the development of the target cortical attribute. Note that, the demographic feature is subject/time specified, which means all the training samples from the same subject/time point have the same demographic features. The vertex-wise morphological features are the values of cortical attribute maps at the vertex i , such as cortical thickness value, sulcal depth, gyrification index, etc. The context features, also named Haar-like features, are the randomly-extracted mean values or regional differences of the local cortical attribute maps, as introduced next.

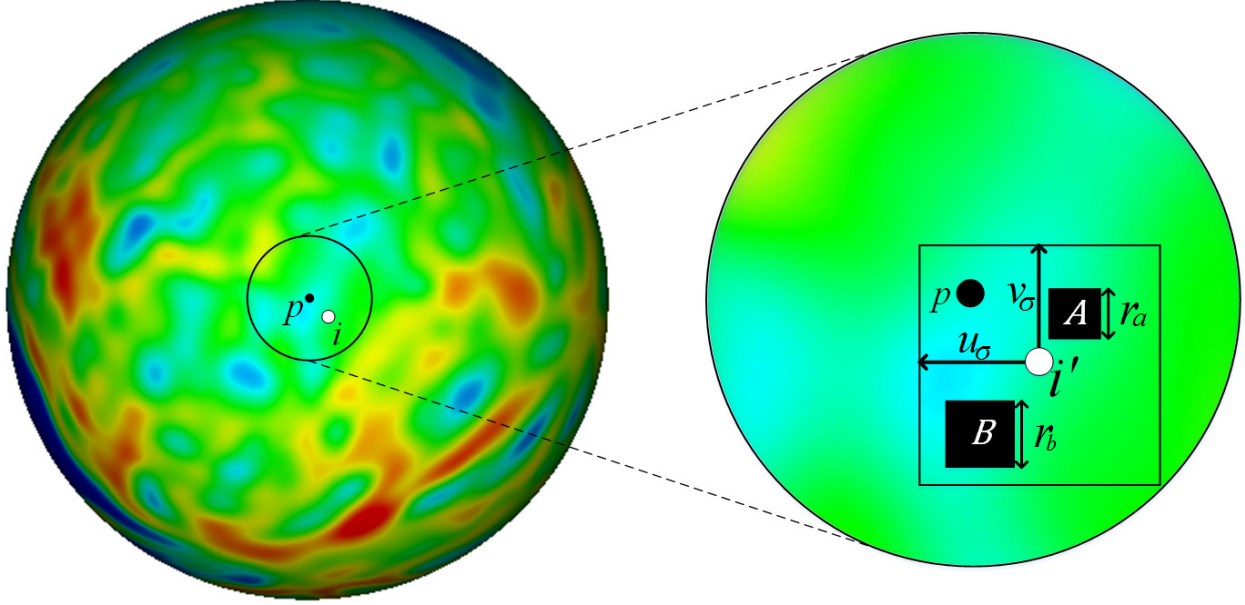


Figure 3.2. Computation of Haar-like features on a resampled spherical surface atlas. The blocks A and B are the two randomly selected regions. The value of Haar-like feature is defined as 1) the mean value of the cortical attributes in the block A , or 2) the mean value of the cortical attributes in the block A subtracting that in the block B .

Figure 3.2 illustrates how to compute context features from a local cortical thickness map for a vertex i on the resampled sphere. Assuming now we are training a decision tree at the vertex p , and without loss of generality (i could be equal to p), we assume vertex i is a neighbor of vertex p . The local cortical thickness map around the vertex p is projected onto the tangential plane, where a local 2D coordinate system is built at the center of the vertex p , and i' is the projection of vertex i on the tangential plane. Two blocks A and B are randomly selected in the neighborhood $[\pm u_\sigma, \pm v_\sigma]$, with their sizes r_a and r_b chosen randomly within the interval $[r_1, r_2]$, where u_σ , v_σ , r_1 , and r_2 are the user-defined parameters. Letting Q_A denote the set of all vertices in block A and Q_B denote the set of all vertices in block B , the context feature at vertex i can be defined as:

$$f(i) = \frac{1}{|Q_A|} \sum_{(u,v) \in Q_A} T(u,v) - \delta \frac{1}{|Q_B|} \sum_{(u,v) \in Q_B} T(u,v) \quad (3.1)$$

where $T(u, v)$ is the value of cortical thickness at position (u, v) , and δ is a random coefficient that takes either 0 or 1.

3.1.3 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:

$$NMSE = \frac{\sum_{i=1}^N (y'_i - y_i)^2}{\sum_{i=1}^N y_i^2} \quad (3.2)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |y'_i - y_i| \quad (3.3)$$

$$MRE = \frac{1}{N} \sum_{i=1}^N \left| \frac{y'_i - y_i}{y_i} \right| \quad (3.4)$$

where y_i and y'_i are respectively the ground truth and the estimated result, and N is the number of vertices.

3.1.4 Experiments and Results

The proposed method was tested on a longitudinal MRI dataset of 15 infants, each with 5 serial scans at around 1, 3, 6, 9 and 12 months of age during the first postnatal year. At each time point of each infant, the cortical thickness map and sulcal depth map were generated. The experiments were to use the cortical thickness map and sulcal depth map at early time points to predict the cortical thickness map at later time points. Specifically, the feature vector includes 1) the accurate postnatal age at MRI scan (in days), 2) the cortical thickness and sulcal depth values at early time points, and 3) the Haar-like features extracted from the local cortical thickness map and sulcal depth map at early time points. To evaluate the method, two nested cross-validation

loops were used. The inner cross-validation loop was used to tune the parameters of DARF, while the outer loop leave-one-out cross validation was used to evaluate the prediction results.

Individual-level Inspection

For each individual, the predicted cortical thickness map was compared with its ground truth, which was obtained using the method in **Section 2.2.2**, and the prediction error map was computed. **Figure 3.3** provides an example of the predicted cortical thickness maps based on the data at 1 month of age for a randomly selected subject. It is clear that the predicted map is generally quite similar to the ground truth.

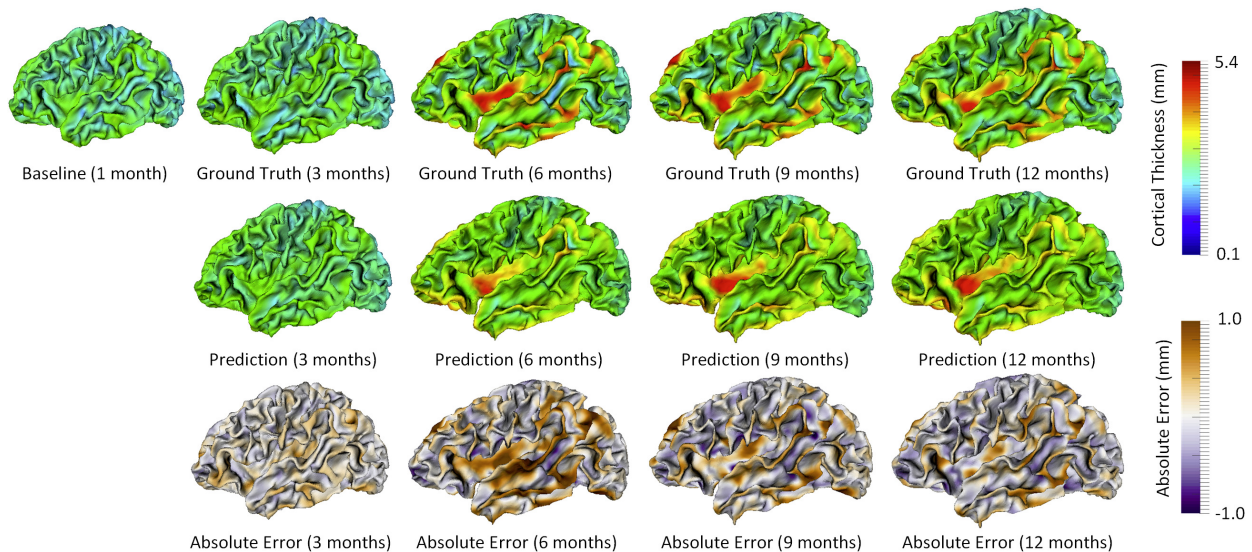


Figure 3.3. Prediction of the cortical thickness maps for a randomly selected subject.

Group-level Inspection

For each individual at each time point, I computed the mean value of the predicted cortical thickness over the whole cortical surface, and then explored the longitudinal distribution of the predicted mean cortical thickness. **Figure 3.4** shows a comparison between the longitudinal distribution of ground-truth mean cortical thickness and its corresponding prediction for 15

subjects. As shown in **Figure 3.4**, the distribution of the predicted cortical thickness is generally similar to the distribution of ground-truth cortical thickness.

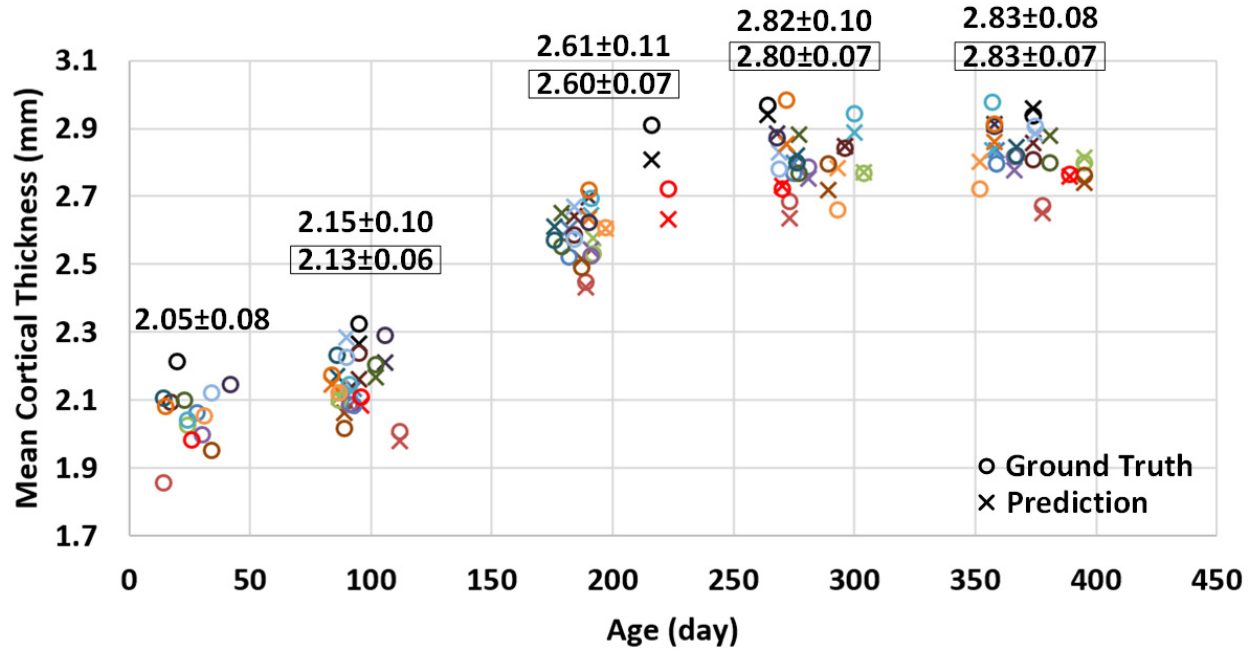


Figure 3.4. Predicted mean cortical thickness and ground truth. This figure plots the longitudinal distribution of ground-truth mean cortical thickness and its corresponding prediction (over the whole cortical surface) for 15 subjects. Different subjects are distinguished by different colors. For each time point, the average ground-truth mean cortical thickness over all 15 subjects and standard deviation are provided on the top of data distribution, and the values within each black rectangle denote the average and standard deviation of the corresponding prediction results.

Furthermore, I examined whether adding more early time points could better predict the cortical thickness maps at future time points. To do this, I gradually included more time points in the training data, and compared the respective prediction results. Specifically, I first used the baseline data (i.e., at 1 month of age) as inputs to predict the cortical thickness maps at 3, 6, 9 and 12 months of age, and then combined the data at 1 and 3 months of age and used them together as inputs to predict the cortical thickness maps at 6, 9, and 12 months of age. More time points were gradually added into the inputs, i.e., up to using the combination of data at 1, 3, 6, and 9 months of age to predict the cortical thickness map at 12 months of age. **Figure 3.5** shows that using more

time points for prediction would generally achieve better results. Additionally, we can see that the error map at each time point is highly correlated with the corresponding standard deviation map of cortical thickness across individuals, with the averaged correlation coefficient of 0.8 ± 0.05 . Another observation based on **Figure 3.5** is that the standard deviations at the 6th and 9th months are relatively larger than those at other time points, and accordingly the prediction errors at the 6th and 9th months are also larger compared with other time points. Of note, the large standard deviation of cortical thickness estimation errors at the 6th and 9th months might be caused by the extremely low tissue contrast of infant MRI at these ages, which makes both cortical surface reconstruction and measurement more challenging and less accurate. In **Figure 3.5**, we can also observe that the prediction accuracy peaks in the unimodal cortex, e.g., the precentral gyrus (primary sensory cortex), postcentral gyrus (primary motor cortex) and occipital cortex (visual cortex), while the prediction accuracy drops in the high-order association cortex, e.g., the prefrontal cortex, temporal cortex, insula cortex, and inferior parietal cortex.

Table 3.1 and **Table 3.2** report the quantitative evaluations of using the data at different available time points to predict cortical thickness maps at future time points. When predicting cortical thickness maps from a single time point at 1 month of age, the prediction error at 3 months of age was the smallest, followed by a second small error at 12 months of age, and the prediction errors at 6 and 9 months of age are relatively larger. This is consistent with the observations in **Figure 3.5**. From these tables, we can also conclude that integrating more time points into the prediction framework generally induces higher accuracy.

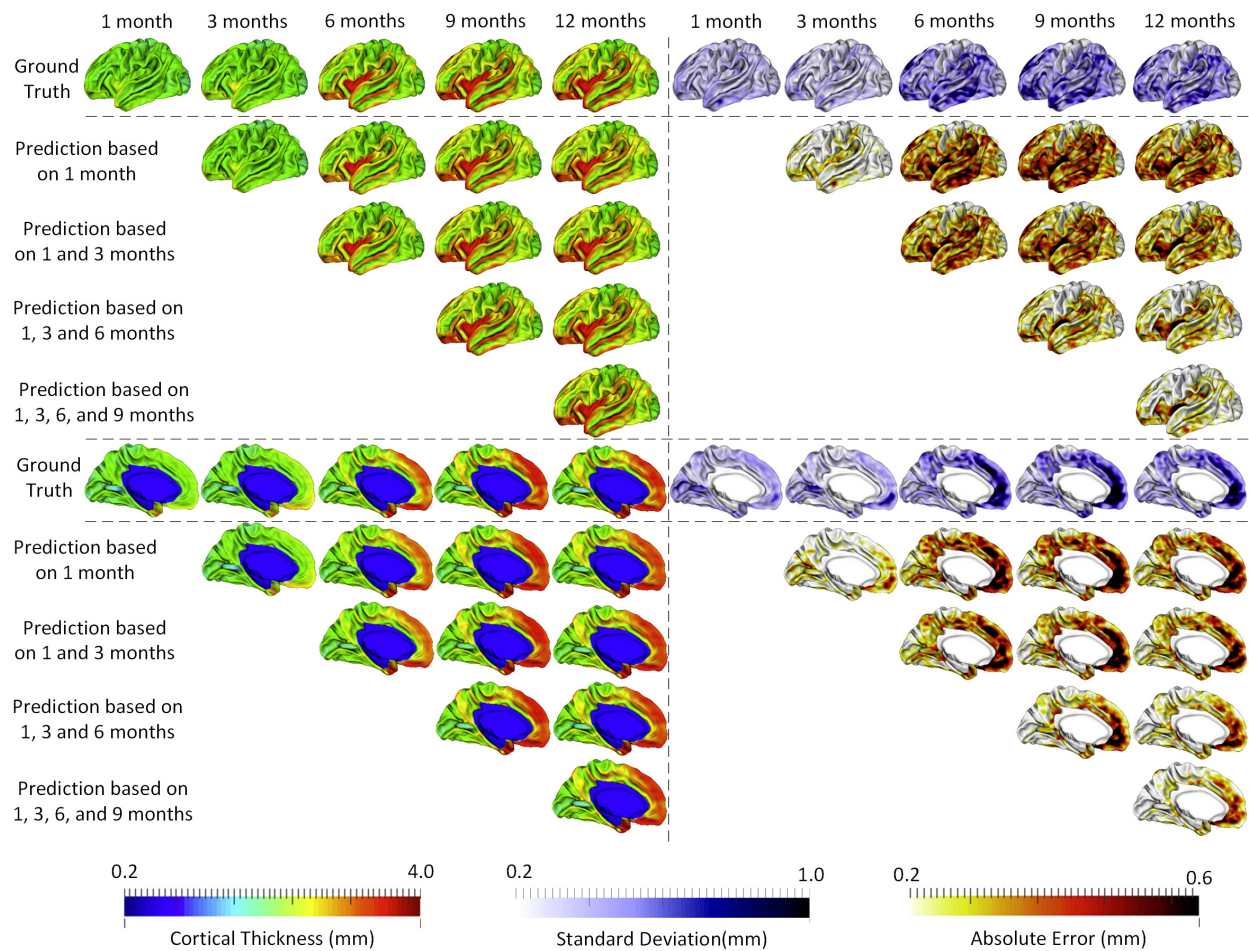


Figure 3.5. The prediction results from multiple available time points, averaged across 15 subjects. The 1st row and the 6th row show, respectively, the averaged cortical thickness maps (left half columns) and corresponding standard deviation maps (right half columns) at all 5 time points. The 2nd–5th rows and the 7th–10th rows show the predicted cortical thickness maps (left half columns) and the corresponding error maps (right half columns).

Table 3.1. Quantitative measures of cortical thickness prediction using mean absolute errors (MAE).

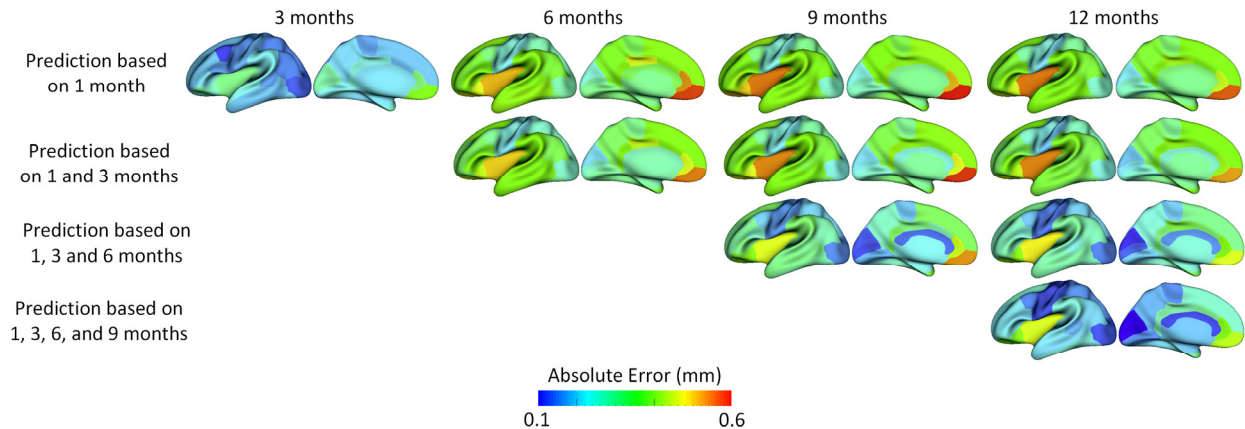
MAE (mm)	Target Time Point			
Used Time Point(s)	3 rd month	6 th month	9 th month	12 th month
1 st month	0.209±0.026	0.332±0.037	0.340±0.030	0.321±0.028
1 st , 3 rd months	-	0.313±0.036	0.321±0.025	0.301±0.025
1 st , 3 rd , 6 th months	-	-	0.026±0.023	0.247±0.022
1 st , 3 rd , 6 th , 9 th months	-	-	-	0.219±0.020

Table 3.2. Quantitative measures of cortical thickness prediction using mean relative errors (MRE).

MRE (%)	Target Time Point			
	3 rd month	6 th month	9 th month	12 th month
Used Time Point(s)				
1 st month	9.9±1.1	13.1±0.9	12.4±0.8	11.7±0.9
1 st , 3 rd months	-	12.3±0.8	11.7±0.7	11.0±0.7
1 st , 3 rd , 6 th months	-	-	9.5±0.8	9.0±0.7
1 st , 3 rd , 6 th , 9 th months	-	-	-	7.9±0.6

Region-based Evaluation

I parcellated each cortical surface into 35 regions using the method developed in (Li, et al., 2014c), and then computed the average prediction error in each region. As shown in **Figure 3.6**, the regions with smaller errors generally included the unimodal cortex, such as the sensorimotor region (precentral gyrus and postcentral gyrus) and visual area (including cuneus cortex, pericalcarine cortex, lingual gyrus, and lateral occipital cortex), while the regions with larger prediction errors represented the high-order association cortex, such as the prefrontal, lateral temporal, cingulate, and insula cortices.

**Figure 3.6.** The average prediction errors (mm) in 35 cortical ROIs for 15 infants.

Comparison with Other Methods.

To demonstrate the advantages of DARF, I compared it with other four representative methods, including the mixed effect model (MEM) in FreeSurfer (Bernal-Rusiel, et al., 2013), polynomial regression (PR), conventional regression forests (CRF), and sparse linear regression

(SLR). MEM, which explicitly models fixed effects and random effects, is a powerful method for analyzing longitudinal neuroimaging data. 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 variations (random effect), such as genetic and environmental influences. PR method assumes that the development of cortical thickness at each vertex has a second-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 the equation below:

$$\arg \min_{\boldsymbol{\alpha} \in \mathbb{R}^d} \frac{1}{2} \|\mathbf{y} - \mathbf{X}^T \boldsymbol{\alpha}\|_2^2 + \lambda \|\boldsymbol{\alpha}\|_1 \quad (3.5)$$

where L and λ 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 3.7 provides a comparison among MEM, PR, CRF, SLR and DART for predicting vertex-wise cortical thickness at 9 months of age for a representative subject. As we can see, DART predicted the cortical thickness map more precisely than the other four methods. **Figure 3.7** 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. A more comprehensive quantitative evaluation will be given in the **Section 3.2.2**.

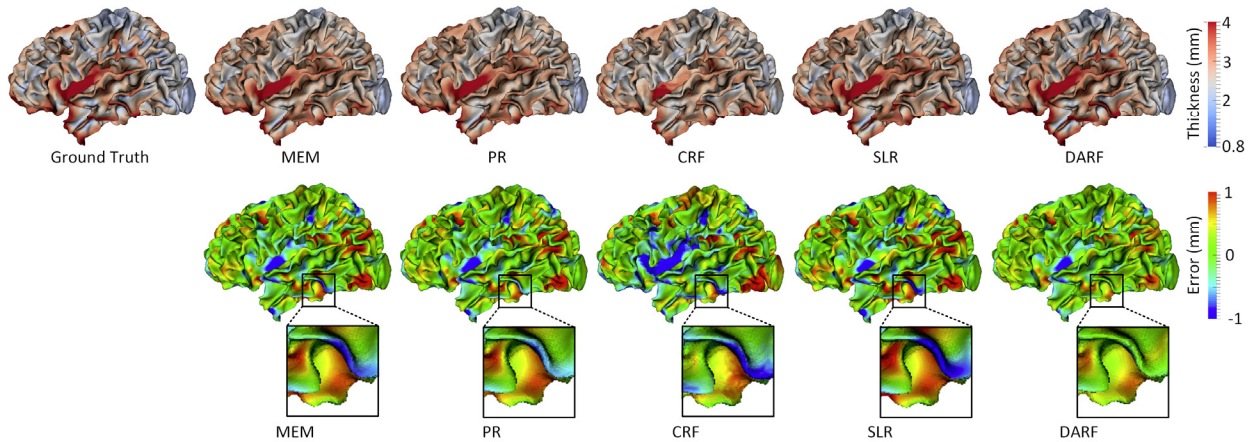


Figure 3.7. Prediction of the vertex-wise cortical thickness map (mm) of a randomly selected infant at 9 months of age by five different methods. The first row shows the ground truth and the estimated thickness maps by five different methods. The second row shows the estimation error maps (mm), along with the zoomed-up ROIs.

3.2 Estimating the Missing Cortical Attribute Maps

This section presents the techniques of how to make a better use of the available data in an incomplete longitudinal dataset to estimate the missing data. **Section 3.2.1** introduces the approach in detail. **Section 3.2.2** describes all the evaluations and comparison experiments and reports the results.

3.2.1 Missing Cortical Attribute Estimation Method

Motivation

As shown in **Figure 3.8**, in the 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, it is expected to use 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 diversity among individuals; and by engaging more time points in the training process, the regression model can better capture the longitudinal information of the 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 3.8**, to estimate the missing data at 6 months of age, I can use at most 28 subjects to train the regression model, since 28 infants have real data at both 1 and 6 months of age. Consequently, only 1 time point (i.e., at 1 month of age) is taken account into the training process. In another way, I 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 conflict and fully utilize the available information, a two-stage missing data estimation strategy is proposed. **Figure 3.9** shows the overview of the proposed strategy, containing the stages of 1) pairwise estimation and 2) joint refinement.

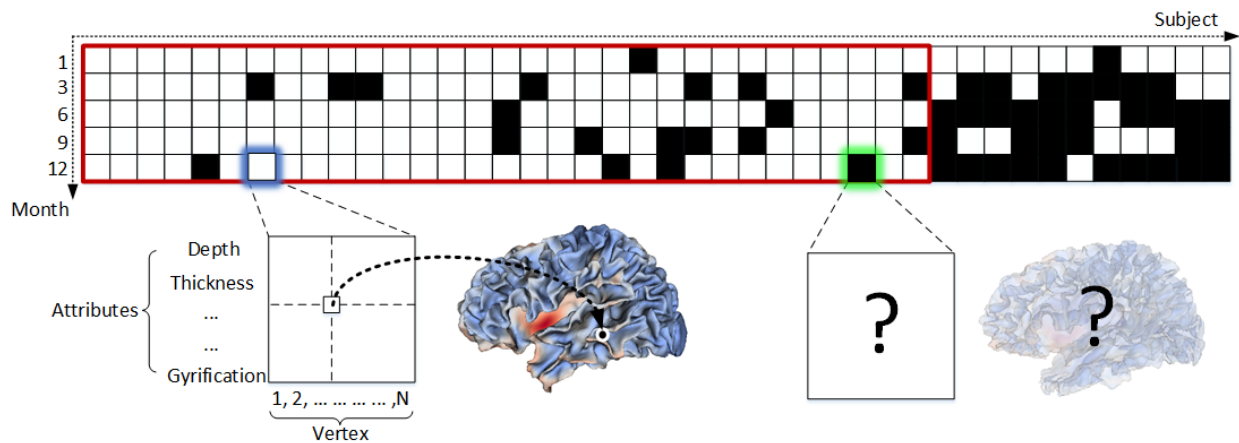


Figure 3.8. Illustration of the longitudinal infant dataset used in this study. Each block indicates the cortical morphological attributes of all vertices of the entire cortical surface for a specific subject (column) at a specific time point (row). The black blocks indicate the missing data at the respective time point. The blocks enclosed by the red rectangle indicate the dataset used in this dissertation.

In **Stage 1**, to use as many subjects as possible for training, the data (i.e., cortical thickness, as computed in **Section 2.2.2**) 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, all the subjects with available data at both 1 month and 6 months of age are used as

training subjects to train a set of decision trees. During the training, the data at 6 months of age is the regression target and the data at 1 month of age are the inputs. After training, for the subjects with available data at 1 month of age but without data at 6 months of age, these decision trees are dynamically/locally assembled as the forest to estimate the missing data at 6 months of age. Similarly, the estimations of missing data at 6 months of age can be obtained, respectively, using the existing data at each of the 3, 9, and 12 months of age. In this way, the available data at all other time points can contribute to the estimation of the data at 6 months of age. Finally, all the estimations contributed from different time points are averaged together as the initial estimation. Similarly, for the missing data at each of 1, 3, 9, and 12 months of age, the same process can be performed to obtain their initial estimations. After performing Stage 1, the missing data of all subjects at all time points will be approximately recovered, thus providing a pseudo-complete longitudinal dataset.

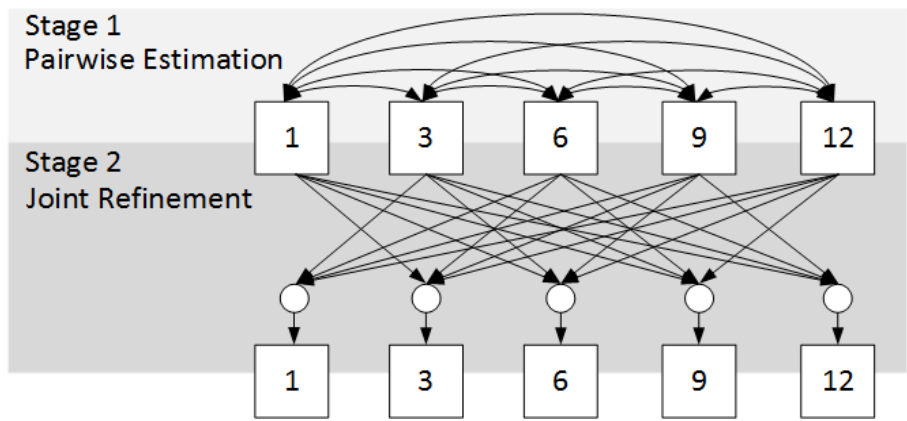


Figure 3.9. Overview of the missing data estimation method. The box with number stands for the data at the corresponding time point. The directed edges represent the processes of estimating the missing data at the target time points (as pointed by the arrowhead) based on the data at the available time 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.

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 of age, all subjects with real data at 6 months of age are used as the training subjects to train a set of decision trees. During the training, the data at 6 months of age is the regression target and the data at 1, 3, 9, and 12 months of age are the inputs. After training, for each subject with missing data at 6 months of age, the trained decision trees can be dynamically/locally assembled as the forest to estimate the missing data. Note that it is not required that each training/testing subject must have real data at 1, 3, 9, and 12 months of age, since the missing data have already been recovered in Stage 1. 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), the proposed strategy can effectively leverage the information from all time points of all available training subjects for the missing data estimation.

3.2.2 Experiments and Results

To evaluate the proposed missing data estimation method, the method was tested by recovering the data of cortical thickness at 5 missing time points. Specifically, from the incomplete dataset (**Figure 3.8**), I selected 16 subjects with complete data at all 5 time points as the reference subjects. Then, I manually selected one of the 16 subjects, deleted the data at one time point, and put it back to the dataset. The deleted data is then recovered using the missing data estimation method. After recovering the missing data, I compared the recovered result with the real data. This experiment was repeated for each of 16 subjects at each of 5 time points.

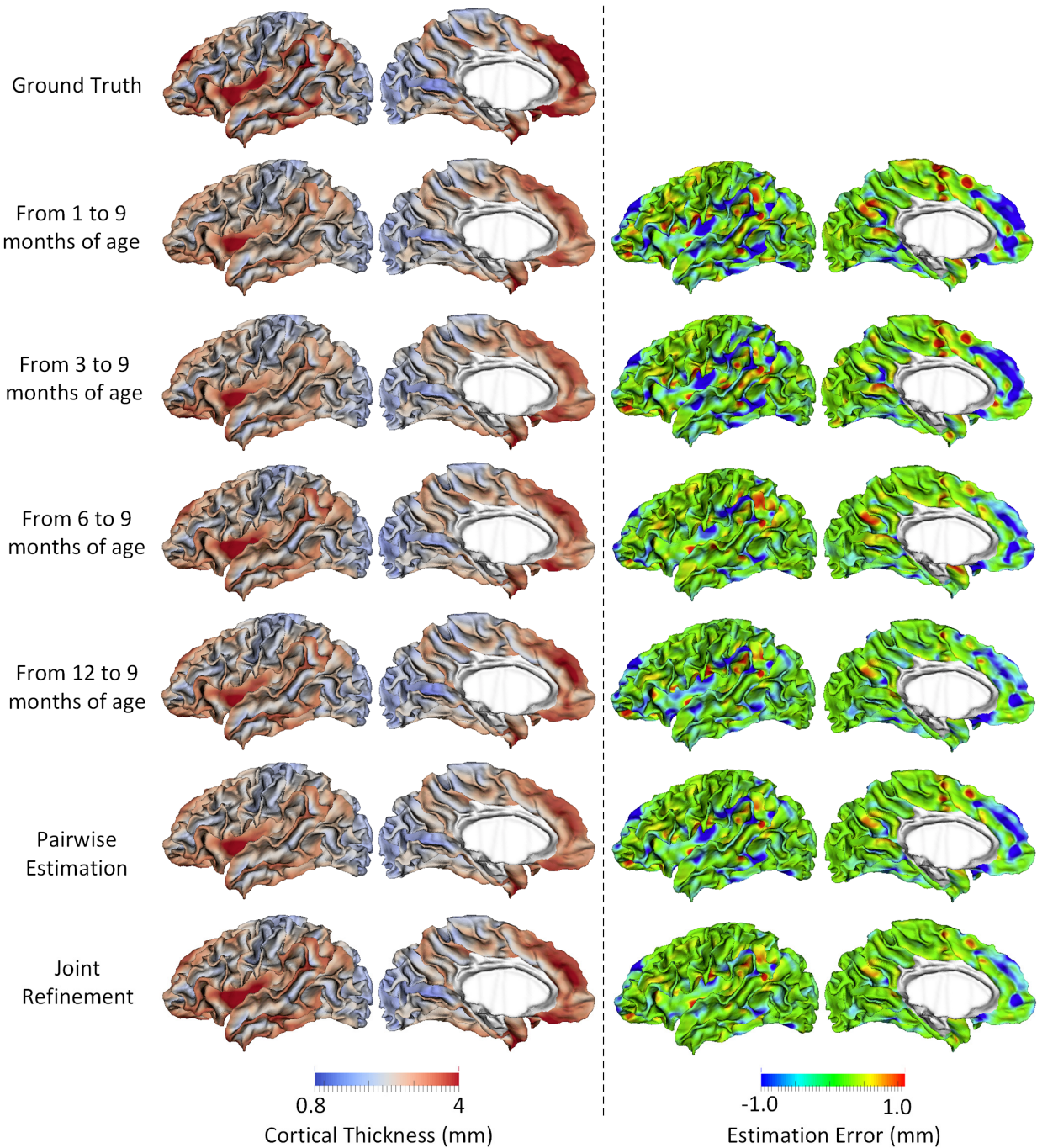


Figure 3.10. Estimations of the vertex-wise missing cortical thickness at 9 months of age for a randomly-selected infant. The first two columns show the maps of ground truth and the estimated cortical thickness at each step. The last two columns show the maps of estimation errors at each stage.

Figure 3.10 shows the error map of our proposed method at each stage of estimating the missing cortical thickness map at 9 months of age on a randomly selected infant. **Figure 3.11**

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 of age as inputs to estimate the cortical thickness at 9 months of age is better than the case of using the data at 1 or 3 months of age. A possible explanation is that the cortical thickness at 9 months of age is more similar to cortical thickness at 6 and 12 months of age, compared to cortical thickness at 1 and 3 months of age. **Figure 3.11** 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 3.12** illustrates the average estimation errors of vertex-wise cortical thickness at all 5 time points.

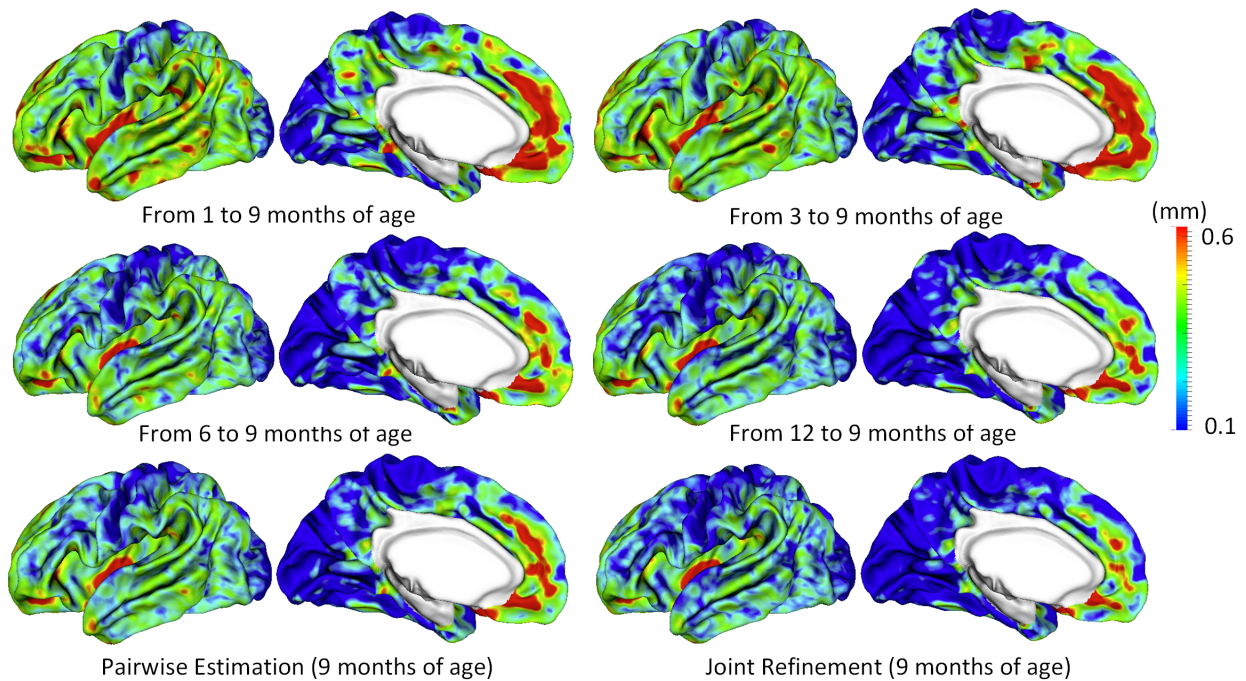


Figure 3.11. Average vertex-wise errors (mm) in estimating the missing cortical thickness at 9-months of age for all subjects at each step of estimation.

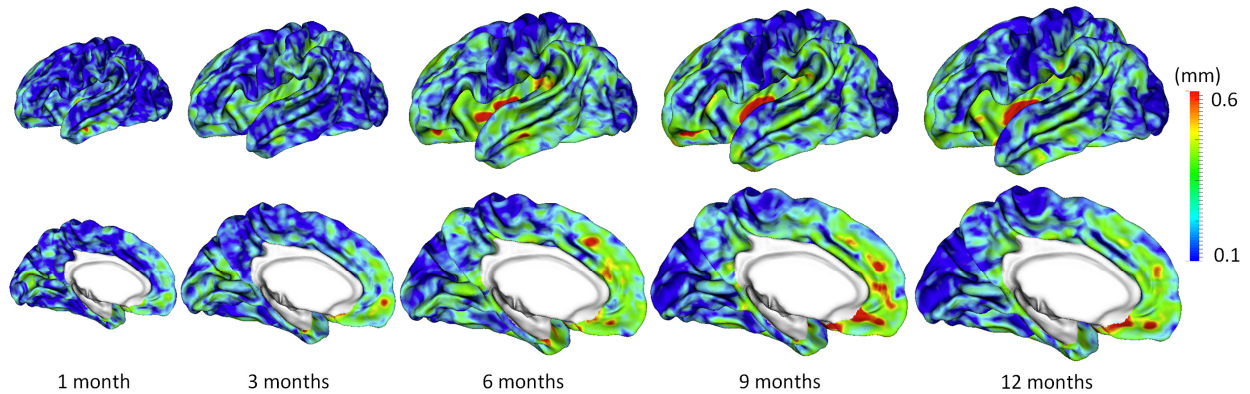


Figure 3.12. Average vertex-wise errors (mm) in estimating the missing cortical thickness at 5 time points for all subjects by using the proposed method.

It is 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, the proposed method can better capture patterns of the unimodal cortex than those of the high-order association cortex, thus leading to more accurate predictions in the unimodal cortex. All of the above observations are further confirmed by the quantitative evaluation given in **Tables 3.3, 3.4, and 3.5**. From these tables, I can further conclude that the proposed method is able to effectively recover the missing cortical thickness map, with the average absolute error of less than 0.23 mm and the average relative error of less than 9.24%. Moreover, I also performed paired t-test to statistically compare between results of pairwise estimation and the results of 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 3.3, 3.4, and 3.5**) is statistically significant. Considering that the MRI resolution is 1 mm and the average cortical thickness is around 2 mm in infants, the estimation error by the proposed method is 10% of cortical thickness and thus is 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 1st, 3rd, 6th, 9th, and 12th month of age, respectively. All these indicate that the prediction is accurate.

Table 3.3. Quantitative measures of estimation results for the missing cortical thickness by the normalized mean squared error (NMSE). PE and JR respectively stand for two stages of pairwise estimation and joint refinement.

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

Table 3.4. Quantitative measures of estimation results for the missing cortical thickness by the mean absolute error (MAE). PE and JR respectively stand for two stages of pairwise estimation and joint refinement.

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

Table 3.5. Quantitative measures of estimation results for the missing cortical thickness by the mean relative error (MRE). PE and JR respectively stand for two stages of pairwise estimation and joint refinement.

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

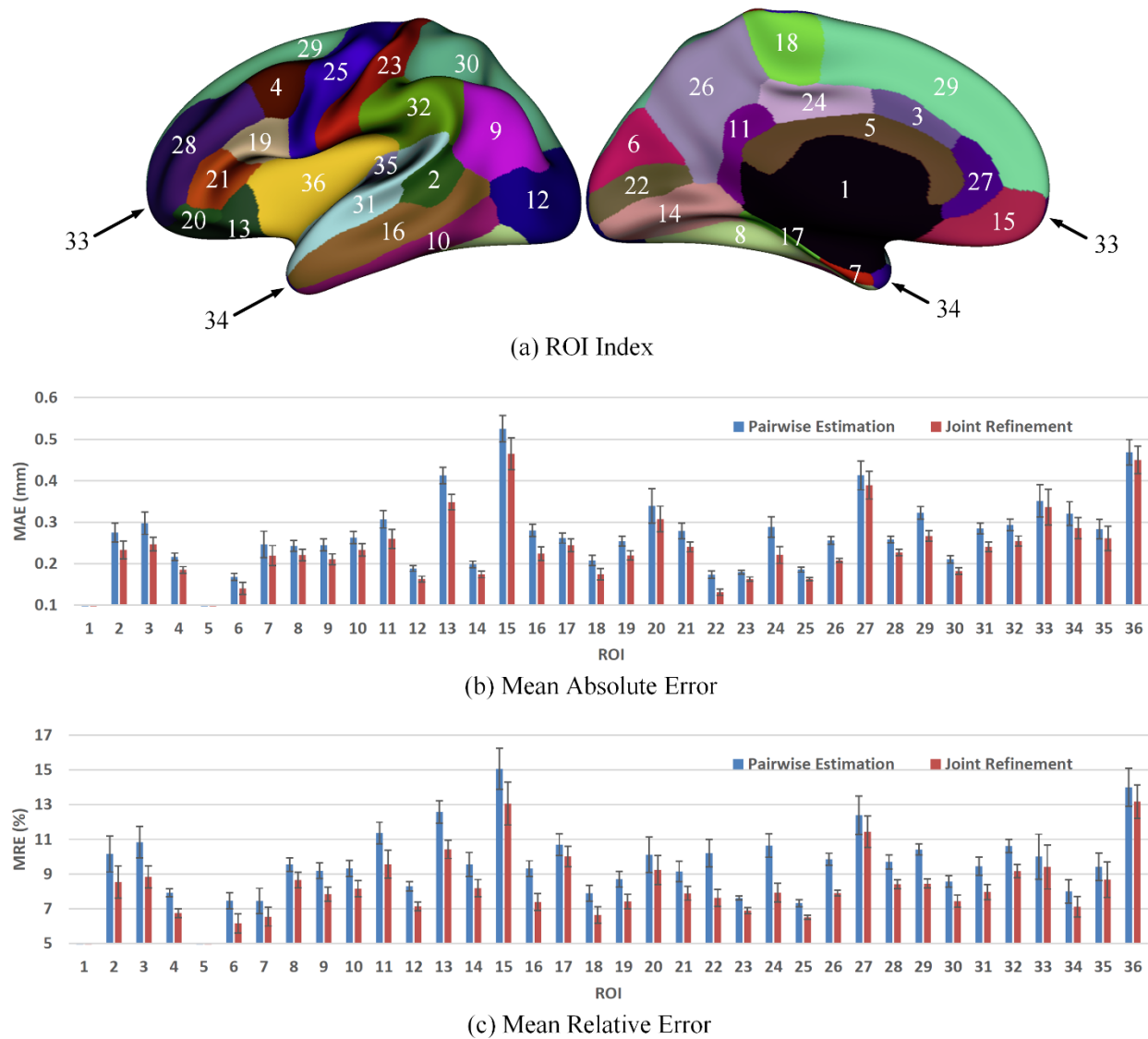


Figure 3.13. Error measures in 36 ROIs for estimation of missing cortical thickness at 9 months of age. ROIs 1 and 5 are excluded, as there is no definition of cortical thickness for these two regions.

I further computed the estimation errors in each of 36 ROIs. **Figure 3.13** 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 of age. We can see that, in all ROIs, the use of joint refinement led to obvious improvements over the case of using only the pairwise estimation. **Figure 3.14** shows error measures in each ROI for estimation the missing cortical thickness at all 5 time points. It can be seen 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).

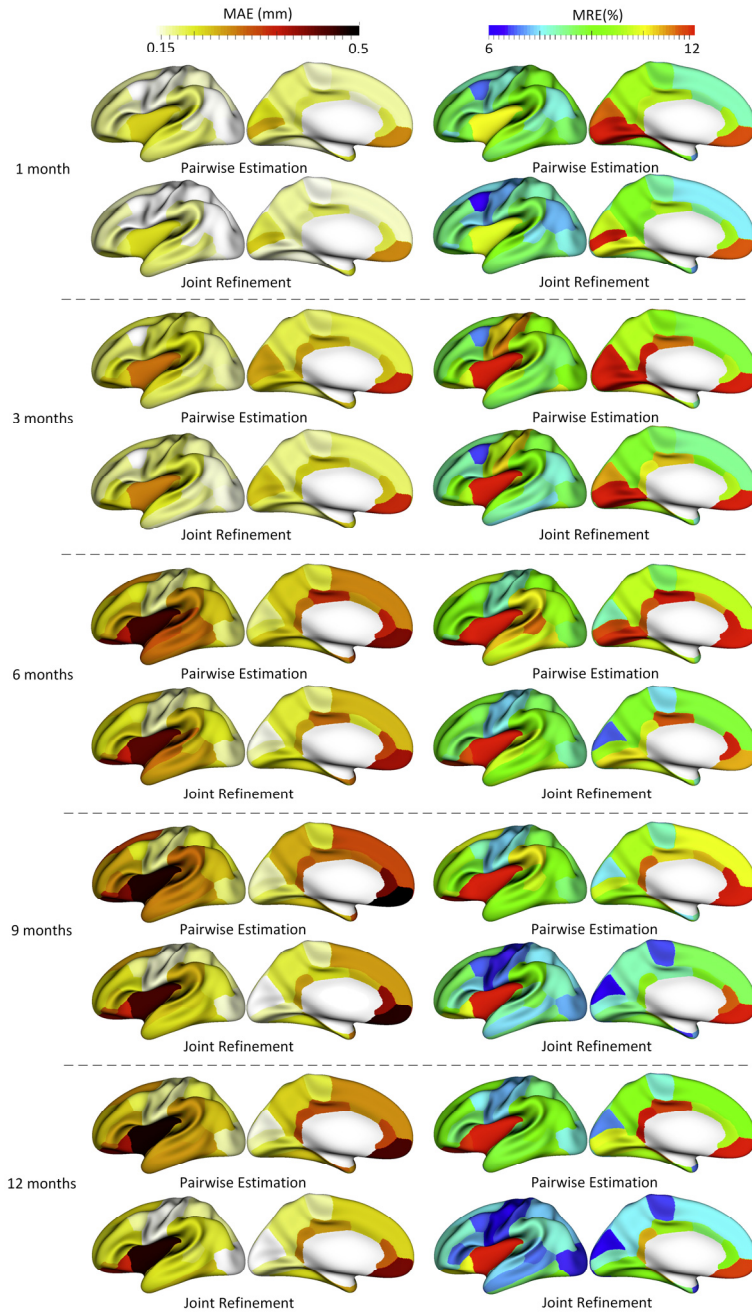


Figure 3.14. The group-average error maps of each step in estimating the missing cortical thickness at 1, 3, 6, 9 and 12 months of age in 36 ROIs.

Performance on Large Portion of Missing Data

I randomly removed some existing data, and tested the proposed method with different portions of missing data. Of note, the original missing data in the dataset (enclosed by the red rectangle in **Figure 3.8**) is 13%. By randomly removing some existing data from the entire dataset, I created datasets with 13%, 20%, 40%, and 60% missing data, respectively. From **Figure 3.15**, we can see that the estimation errors increase with the increased portion of missing data. However, even the missing data reaches 40%, the proposed method is still able to produce an average estimation error of 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 having enough quality training data, DARF is able to estimate the missing data precisely. **Figure 3.15** also shows that joint refinement consistently improves the results of pairwise estimation.

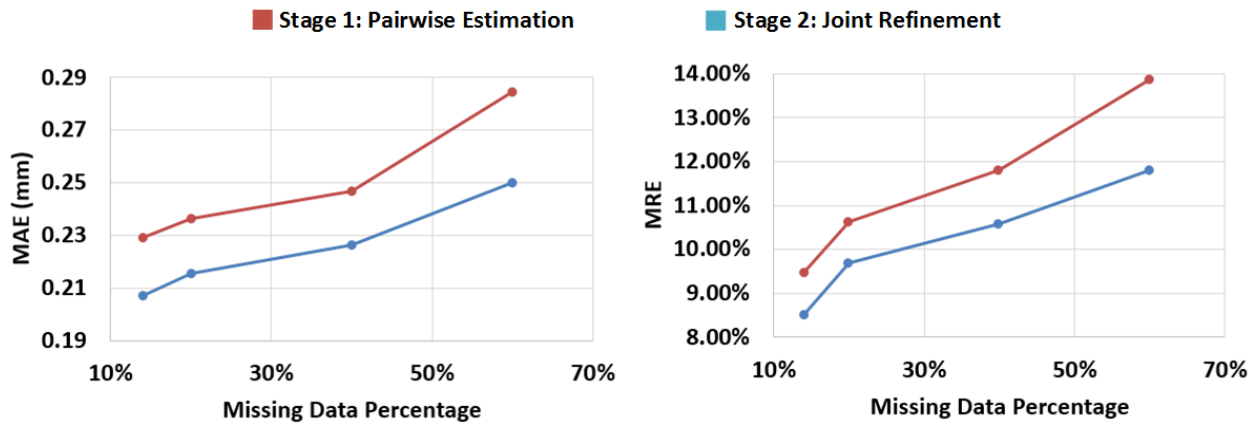


Figure 3.15. The relationship between the estimation errors and the portions of missing data.

More Quantitative Comparisons

The proposed missing data estimation framework is not limited to using DARF. In fact, the framework is open to any regression method. Here I compared five different methods, including the mixed effect model (MEM), polynomial regression (PR), conventional regression forest (CRF), and sparse linear regression (SLR), for estimation of vertex-wise cortical thickness.

Table 3.6 reports complete quantitative evaluation for the five methods based on the leave-one-out cross-validation. It shows that DARF outperforms all the other methods. An interesting observation is that PR performs much worse at the first and last time points, but does relatively better at the intermediate time points. That means, when using a quadratic curve to fit the development of cortical thickness, it is relatively difficult to precisely estimate the two ends of the curve. I further performed paired t-test between our method and all the other methods. With all obtained p-values being less than 0.001, the significant advantage of our method is demonstrated.

Table 3.6. Quantitative evaluation of the performance of cortical thickness estimation by using MEM, PR, CRF, SLR and DARF.

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
	DARF	0.0117±0.0019	0.0140±0.0032	0.0139±0.0030	0.0121±0.0021	0.0107±0.0019
MAE (mm)	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
	SLR	0.175±0.023	0.205±0.027	0.263±0.038	0.268±0.026	0.259±0.026
	DARF	0.168±0.013	0.193±0.025	0.233±0.034	0.224±0.016	0.215±0.016
MRE (%)	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
	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
	DARF	8.32±0.69	9.23±1.25	9.08±0.79	8.14±0.51	7.81±0.52

3.3 Summary

This chapter presented the proposed techniques for estimating the cortical attribute maps. These techniques were used for addressing two sub-problems: (1) predicting the future development of cortical attributes, and (2) estimating the missing cortical attributes in incomplete longitudinal datasets. For the first sub-problem, a Dynamically-Assembled Regression Forest (DARF) was proposed in **Section 3.1**. By training a single decision tree at each sample point on the spherical cortical surface and grouping local decision trees as the forest, DARF is able to

achieve accurate and spatially smooth predictions of cortical attribute maps. Moreover, by sharing common decision trees for the neighboring DARFs, this method is also time-efficient. **Section 3.1** further introduced the extraction of surface-based Haar-like features. Since Haar-like features provide rich neighboring information of the local cortical attribute maps and enable DARF to encode the correlation between the prediction target and the local feature distribution, better prediction accuracy can be obtained. For the second sub-problem, **Section 3.2** presented a novel missing data estimation strategy, which consists of two stages, i.e., pairwise estimation and joint refinement. In the pairwise estimation stage, missing data were estimated by utilizing the relationship between the missing time point and each of the other available time points independently, in order to use as many training subjects as possible. Then the independent outputs were averaged as initial estimations. In the joint refinement stage, the initial estimation was refined by utilizing the relationship between the missing time points and all other available or initially estimated time points together. In this way, the available information in the dataset can be effectively used, so that the missing cortical attributes can be estimated more precisely. The experiments and comparisons with existing approaches demonstrate the effectiveness of the proposed techniques.

The work in this section also has some limitations. First, the proposed methods have been evaluated for estimating only the cortical thickness map, though the proposed methods are generally designed for all kinds of cortical attributes, such as sulcal depth, gyrification index, and myelin content. Second, due to the lack of data, the proposed model is trained only on normal brain datasets with healthy infants. However, in order to understand the development of cortical attributes of specific neurodevelopmental disorders or perform early diagnosis, DARF should be trained using a dataset of a certain disease or a general dataset with both normal and abnormal

brains. Third, since a regression forest is not a descriptive model, the learned relationships between the input cortical attribute maps and the output cortical attribute maps are hard to interpret. This problem might be partially solved by integrating parametric models with DARF. However, more experiments are required to fully validate this idea.

4 DISCOVERY OF THE MAJOR CORTICAL FOLDING PATTERNS IN INFANTS

Cortical folding pattern is another important attribute of the cerebral cortex. This chapter presents a sulcal pits-based framework to study the folding patterns in sulcal regions. Specifically, **Section 4.1** first introduces how to extract sulcal pits from an infant cortical surface, followed by an investigation of the spatial distribution and longitudinal development of sulcal pits in infants. **Section 4.2** introduces how to use sulcal pits to discover the major cortical folding patterns in local sulcal regions. Finally, **Section 4.3** investigates whether the sulcal pattern information can be encoded into DARF to better estimate the development of cortical thickness map.

4.1 Sulcal Pits

Sulcal pits, the locally deepest points in the sulci of the highly convoluted cerebral cortex, are found to be spatially consistent across human adult individuals. Researches also suggested that sulcal pits are genetically controlled and have close relationships with functional areas. Sulcal pits can be extracted from a cortical surface using a watershed algorithm (Im, et al., 2010). However, this method is controlled by several important parameters, which are sensitive to the brain size, and thus the watershed algorithm that is tuned on adult brains in previous studies cannot be directly applied to the dynamically developing infant brains. The first content of this section is to introduce the methods of how to adapt the watershed algorithm to extract sulcal pits from cortical surfaces in a variety of brain sizes.

The distribution of sulcal pits is relatively consistent across adult brains. This is one of the important reasons why sulcal pits can be used as reliable landmarks for comparing cortical folding

patterns between adult individuals. However, whether such spatial consistency does still hold for the distribution of sulcal pits in fast growing infant brains is unknown. Therefore, the second content of this section is to investigate the spatial distribution and longitudinal development of sulcal pits in infants.

4.1.1 Sulcal Pits Extraction

If the sulcal depth of each vertex on the cortical surface is known, the sulcal pits can be extracted using a watershed algorithm. The algorithm can be performed as follows: 1) from a set of vertices that have not been processed yet, a vertex with the largest sulcal depth is picked out; 2) if none of its one-ring neighboring vertices has already been processed, this vertex is then selected as a sulcal pit and is assigned with a new label; if one of its one-ring neighboring vertices has been processed, this vertex is assigned with the same label as this neighboring vertex; if more than one of its one-ring neighboring vertices have been processed, this vertex is assigned with the label of the neighbor that owns the most similar sulcal depth to this vertex; 3) the steps 1 and 2 are repeated until the sulcal depth of the currently picked vertex is less than a predefined threshold T , as only the deep sulcal pits are needed.

Cortical surface is not a perfectly smooth sheet, even the regions that look flat are still slightly uneven. Since the watershed algorithm is very sensitive to the degree of smoothness, directly applying the above algorithm would extract many meaningless (false positive) sulcal pits, as shown in **Figure 4.1**. This phenomenon is called over extraction. To avoid over extraction, three threshold parameters (i.e., area threshold A , distance threshold D and ridge height threshold R) are introduced. Specifically, area threshold A is used to exclude the sulcal pits in very small basins. Distance threshold D is used to prevent two sulcal pits from being too close to each other. Ridge height threshold R is used to check whether a candidate sulcal pit is significantly different from its

neighboring sulcal pit. Note that, a ridge point is the position where two sulcal basins first meet, and ridge height is the sulcal depth difference between a candidate sulcal pit and the ridge point of two neighboring basins, as explained in **Figure 4.2**. For a candidate sulcal pit, if it fails to pass the threshold A (area) or D (distance), and also fails to pass the threshold R (ridge height), it would not be included as a sulcal pit.

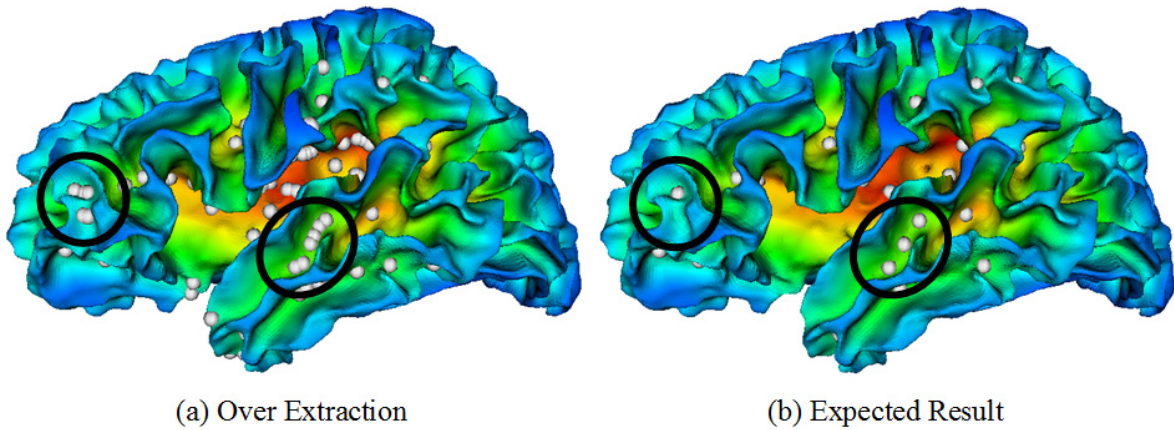


Figure 4.1. Sulcal pits in the case of (a) over extraction and (b) expected extraction/result.

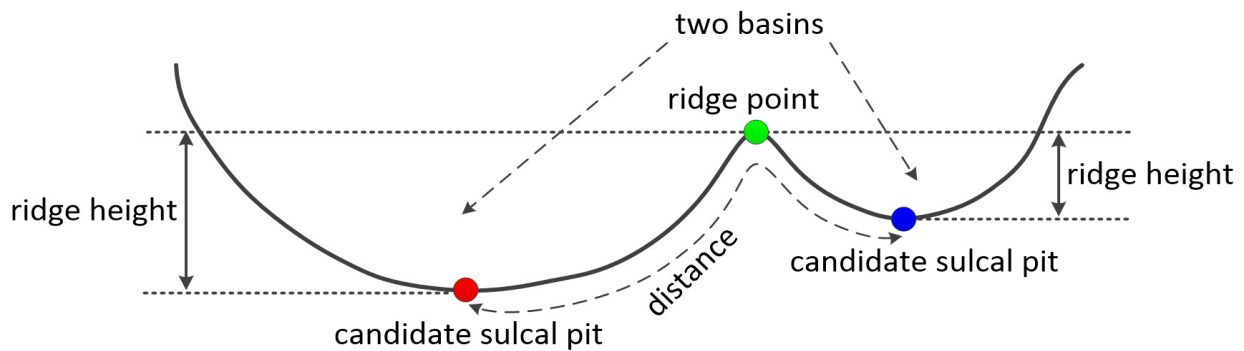


Figure 4.2. A 2D schematic illustration of the ridge height. The ridge height is defined as the sulcal depth difference between a ridge point (green point) and a candidate sulcal pit (red or blue point).

For the adult brains with relatively similar brain sizes across individuals, these threshold parameters could be set as fixed values for all subjects as in (Im, et al., 2010). However, for infant brains, the brain size varies a lot across ages, thus using fixed values to set these threshold parameters is not suitable any more. Intuitively, these parameters might have relationships with some measurements of the cortex, such as volume, surface area, mean depth, maximum depth, etc.

In this dissertation, in order to find these relationships and set suitable parameters for infant brains, exhaustive searches were performed within a predefined range for each parameter based on the 10 randomly-selected infants with manually-labeled “ground truth” of sulcal pits.

Specifically, 10 infants were randomly selected. Each infant had three reconstructed cortical surfaces, respectively, at 0 year, 1 year, and 2 years of age. Then, watershed algorithm was applied to these subjects with very loose parameters, to make sure that sulcal pits were over extracted. In the implementation, more than 90 sulcal pits for each hemisphere were extracted. Next, the results were manually edited by removing “inappropriate” sulcal pits. The basic principles of manual edition were two-fold. First, the sulcal pits in major sulci were kept, but the sulcal pits in minor sulci were discarded. Second, in each sulcal basin, only one sulcal pit is kept. If two or more pits were extracted in a sulcal basin, the deepest one was kept. After manually removing the “inappropriate” sulcal pits, the results were used as “ground truth”.

After having the “ground truth”, a grid search was performed to look for the optimal parameters, with which the sulcal pits extraction results were the closest to the “ground truth”. Since there are four parameters (T , A , D , R) to determine, the grid search was executed in four-layer loops. Each layer tested one parameter with a predefined searching range and incremental step. Let the searching ranges for T , A , D , and R be respectively denoted as $[T_1, T_2]$, $[A_1, A_2]$, $[D_1, D_2]$, and $[R_1, R_2]$. Specifically, for each cortical surface, T_1 was set as the depth of the shallowest point on the cortical surface and T_2 was set as the depth of the shallowest sulcal pits in the “ground truth” of the same age; A_1 was set as 0 (mm^2) and A_2 was as 30 (mm^2); D_1 was set as 5 (rings) and D_2 was set as 20 (rings); R_1 was set as 0 (mm) and R_2 was set as 2.5 (mm). The incremental steps for searching the optimal T , A , D , and R were respectively set as 0.2 (mm), 0.5 (mm^2), 1 (ring) and 0.5 (mm). For any combination of the four parameters, sulcal pits extraction

results were compared with the “ground truth”. If a sulcal pit in the ground truth did not appear in the result, it counted as a false negative. If a sulcal pit in the result was not extracted in the “ground truth”, it counted as a false positive. The optimal setting of parameters was obtained when the numbers of false positive and false negative were minimal.

The results show that changing D from 8 (rings) to 12 (rings) achieves similar low errors. Thus, 10 (rings) is chosen as the distance threshold. Note that, since every cortical surface was resampled as a triangular mesh with the same number of vertices, the triangle size in the surface mesh was adaptive to the brain size. As a result, setting the distance threshold as a fixed ring number could adaptively fit infant cortical surfaces with any size.

The results also indicate that increasing R from 0 mm to 2.0 mm leads to better results and further increasing R to 2.5 mm generates similar results. Therefore, R could be set as 2.5 mm, without depending on the brain size.

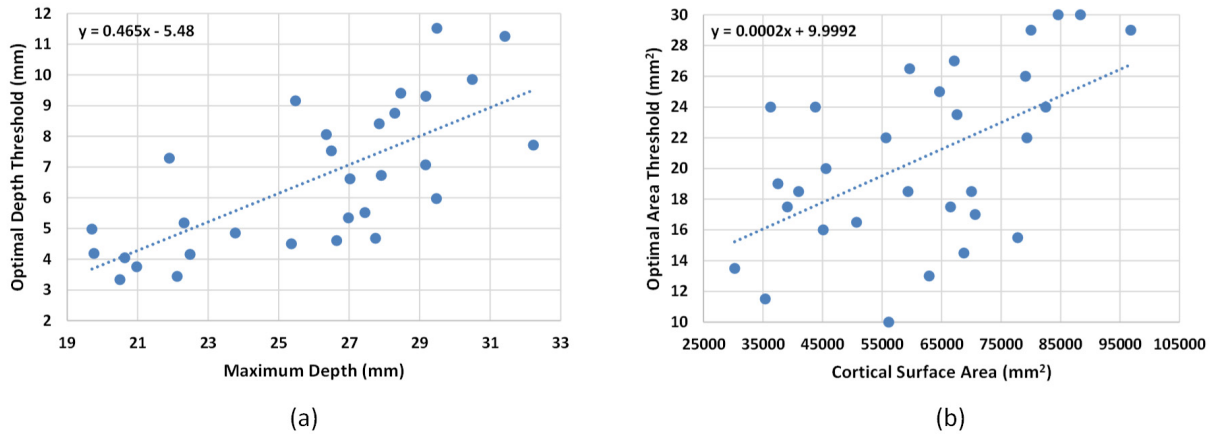


Figure 4.3. Relationships between optimal parameters and cortical surface metrics. (a) The relationship between the maximum depth of the cortical surface and the optimal depth threshold T . (b) The relationship between the cortical surface area and the optimal area threshold A .

The optimal parameters of depth threshold T and area threshold A are quite different across cortical surfaces. Intuitively, the correlation between T and the mean/maximum depth of the whole cortical surface was explored, so was the correlation between A and the total cortical

surface area. The experiment shows that the optimal depth threshold has an approximately linear relationship with the maximum depth, and the optimal area threshold has an approximately linear relationship with the total cortical surface area, as shown in **Figure 4.3**. For better illustrating the influences of these parameters, **Figures 4.4, 4.5, 4.6, 4.7** provide examples of sulcal pits extraction using different settings of parameters.

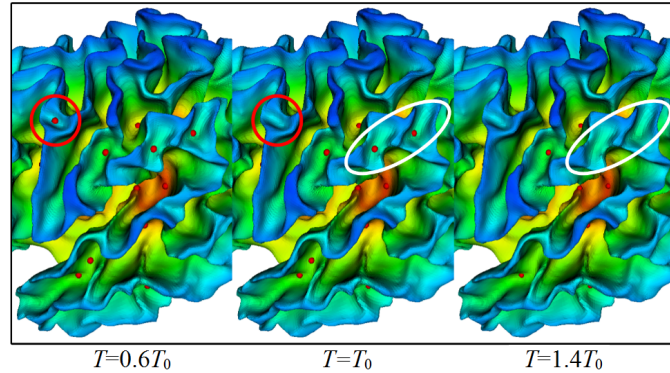


Figure 4.4. Sulcal pits extraction using different depth thresholds. T_0 denotes the “optimal” depth threshold.

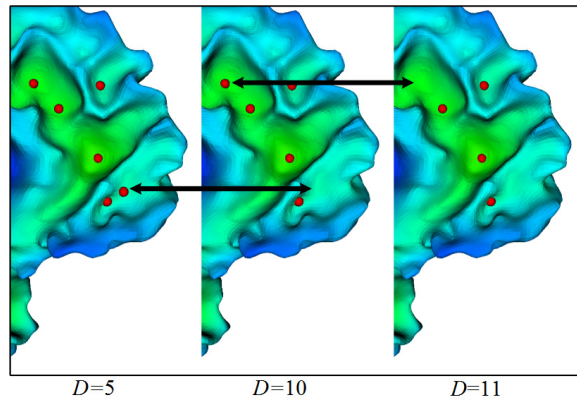


Figure 4.5. Sulcal pits extraction using different distance thresholds (rings).

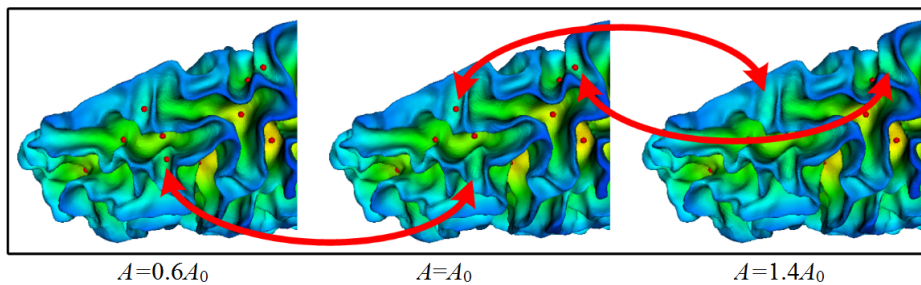


Figure 4.6. Sulcal pits extraction using different area thresholds. A_0 denotes the optimal area threshold.

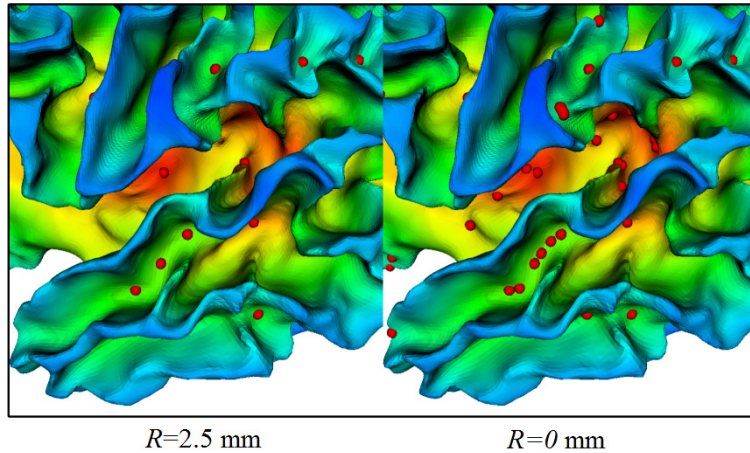


Figure 4.7. Sulcal pits extraction using different ridge height thresholds.

4.1.2 Spatial Distribution and Longitudinal Development

To investigate the spatial distribution and longitudinal development of sulcal pits in infant brains, the sulcal pits extraction method was applied to a longitudinal dataset of 73 infants. Each infant has three reconstructed cortical surfaces respectively at birth, 1 year and 2 years of age. **Figure 4.8** shows the sulcal pits extraction results on the cortical surfaces of the left hemisphere of a representative infant. The sulcal pits are represented by white balls, and the cortical surfaces are color-coded by the sulcal depth. Due to the convoluted cortical folding, it is difficult to observe the sulcal pits in some deep sulci. For better inspection, as shown in the second, fourth and sixth rows of **Figure 4.8**, sulcal pits are mapped onto the partially inflated cortical surfaces. As we can see, the cortical surface grows dynamically while the sulcal depth increases considerably, especially in the first year of life; however, the spatial distribution of deep sulcal pits in major sulci is temporally relatively consistent. For example, as can be observed in the third and fourth rows, there consistently exist two sulcal pits in the central sulcus, and their relative positions were nearly unchanged in the first 2 years, although the central sulcus became much longer and deeper.

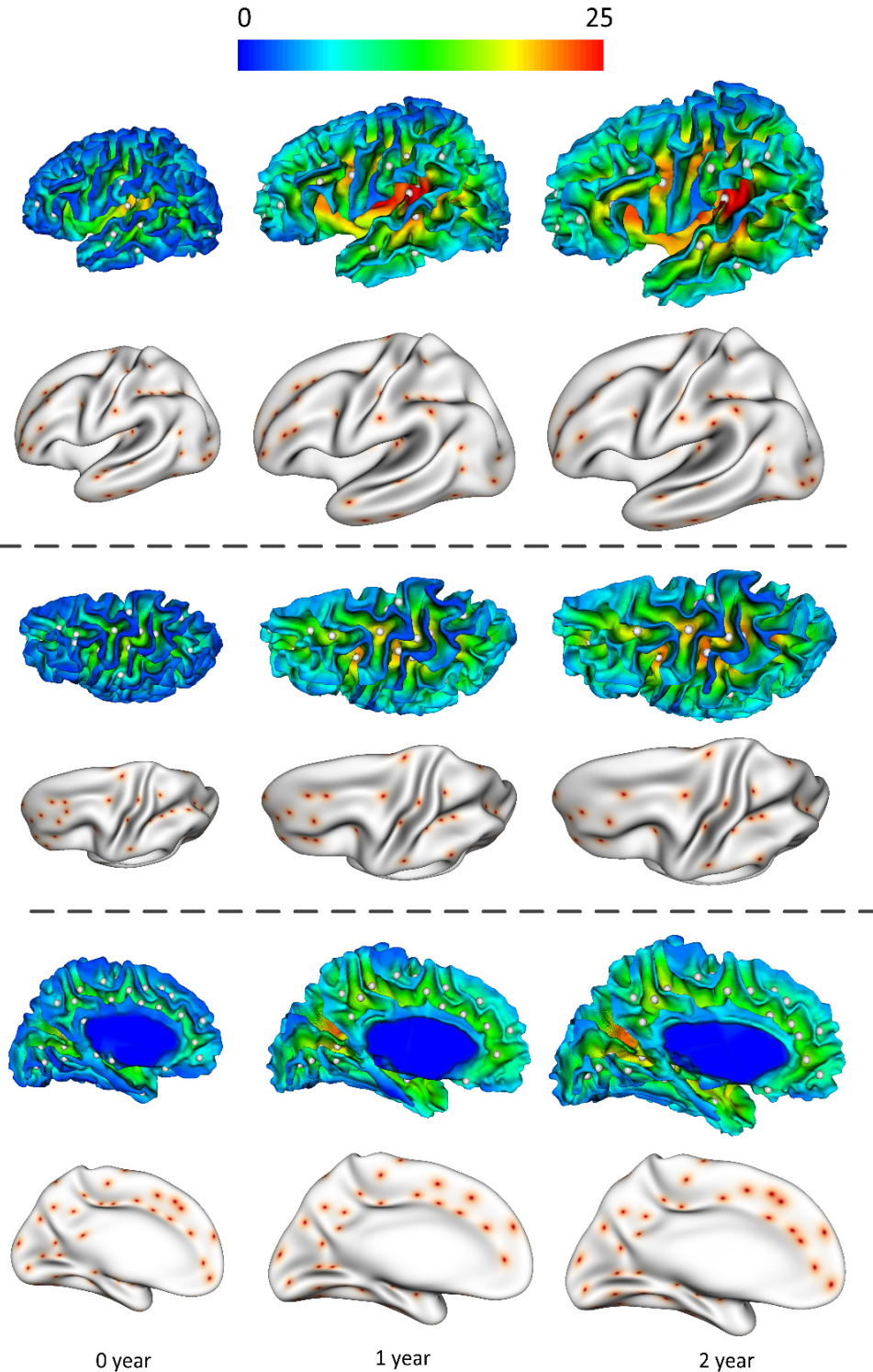


Figure 4.8. Sulcal pits extraction results on the left hemisphere of a representative infant at 0, 1 and 2 years of age. In the first, third and fifth rows, the sulcal pits represented by white points are overlaid on the cortical surfaces that are color-coded by the sulcal depth (mm). For better visual inspection, in the second, fourth and sixth rows, the sulcal pits represented by red points are mapped onto the partially inflated cortical surfaces.

To examine the spatial distribution of sulcal pits across individuals at each age, all the sulcal pits from 73 infants were mapped onto the age-matched inflated surface atlas. Each subject may have some particular sulcal pits in the relatively minor sulcal regions, and such sulcal pits turn out to be the outliers after being mapped to the atlas. These outliers should be removed, because they are statistically meaningless. As most of major sulcal pits across subjects are spatially close on the surface atlas, whether a sulcal pit is a major one or an outlier can be identified by checking the number of sulcal pits around it. Specifically, an intuitive graph building algorithm was performed. Considering each sulcal pit as a node, for each sulcal pit, its two-ring neighborhood was checked. If there was another sulcal pit in the two-ring neighborhood, they were connected by an edge. After all sulcal pits had been checked, the spatially-close sulcal pits in different cortical regions were connected as different graphs. Then, the number of node in each graph was counted. If the node number was no more than 7 ($\approx 10\%$ of 73 subjects), all nodes in the graph were identified as outliers and then removed. **Figure 4.9** shows the distribution of sulcal pits after removing outliers.

As shown in **Figure 4.9**, the sulcal pits of the infant population are consistently concentrated in some specified regions from 0 to 2 years of age. This pattern is particularly pronounced in major sulci, such as the central sulcus, precentral sulcus, postcentral sulcus, superior temporal sulcus, and parieto-occipital sulcus. For example, in the central sulcus, two distinct clusters of sulcal pits can be identified on both left and right hemispheres at birth. During the dynamic cortex development from 0 to 2 years of age, the relative positions of concentration regions almost keep unchanged. This result suggests the existence of spatially-consistent distributions of sulcal pits in major sulci across subjects at term birth, and also suggests that these spatial distribution patterns are relatively stable during the dynamic cortex development in the first

2 years of life. In addition, the number of sulcal pits is also relatively stable. The average changes of individual's amount of sulcal pit are 3.67 ± 0.13 in the first year and 1.30 ± 0.67 in the second year. Such small changes are likely caused by the development of minor sulci, or even noises in the surface processing pipeline.

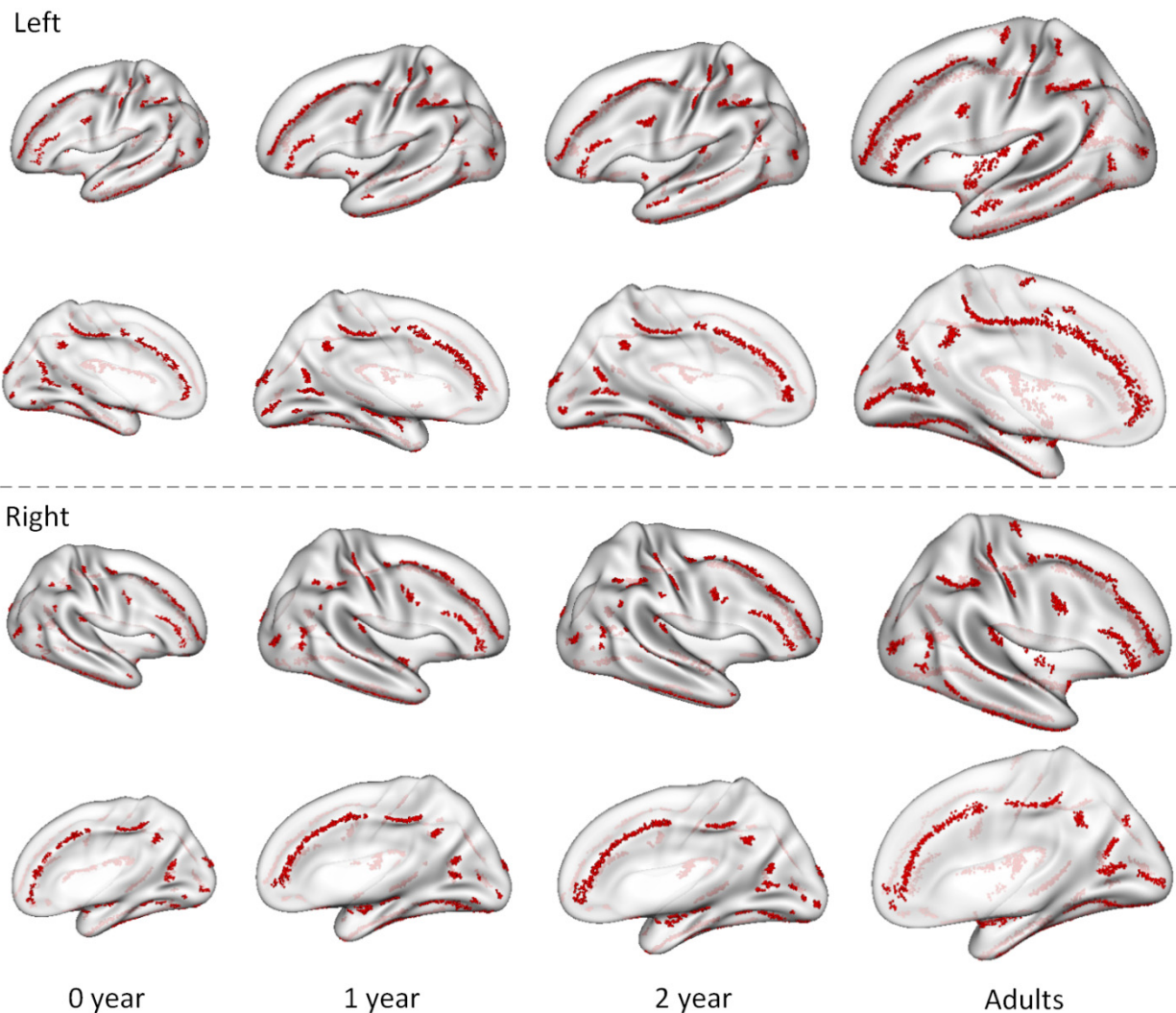


Figure 4.9. Spatial distributions of sulcal pits on both left and right hemispheres from 73 infants at 0, 1, 2 years of age and 64 young adults. All the sulcal pits (red points) were mapped onto the age-matched, partially-inflated cortical surface atlases.

To closely compare sulcal pits distributions in infants with those in adults, the sulcal pits extraction method was also applied to the cortical surfaces of 64 young adults. As shown in the

fourth column in **Figure 4.9**, the spatial distributions of sulcal pits in major sulci of adults were highly consistent with those of infants, although more sulcal pits in minor sulci existed in adults.

So far, the experiment results have demonstrated that the spatial consistency of sulcal pits in adult brains does still hold for the rapid growing infant brains. Therefore, it is safe to utilize sulcal pits as landmarks to study the sulcal patterns of infants as the previous studies did for the adult brains (Im, et al., 2011b; Im, et al., 2013b). Much more detailed studies and analyses on the spatial distribution and longitudinal development of infant sulcal pits have also been carried out, but they are not included in this dissertation, as they are more neuroscience related. For anyone interested in that work, please refer to the paper (Meng, et al., 2014).

4.2 Discovery of Major Sulcal Patterns in Neonates

This section introduces how to use sulcal pits to discover the major sulcal patterns from a large-scale dataset of neonatal brains. It is worth noting that neonatal brains are more ideal candidates than adult brains for studying major sulcal patterns. That is because the primary sulcal patterns are genetically influenced and have been established at term birth, and also neonates have minimal exposure to the complicated postnatal environment, which may potentially influence the shape of the cortex.

The overview of the proposed method for discovering major sulcal patterns is illustrated in **Figure 4.10**. First, sulcal pits are extracted using the method introduced in **Section 4.1.1**. Second, a graph-based shape representor, namely sulcal graph, is constructed for a certain region in each cortical surface. Third, the similarity between any pair of sulcal graphs is measured from six complementary aspects. Fourth, to capture both the common information and the complementary information, the six metrics are adaptively combined together. Finally, sulcal graphs are clustered into different groups based on the combined similarity matrix. To introduce the proposed method

in detail, the rest of this section is organized as follows. **Section 4.2.1** provides the definition of sulcal graph and presents how to quantitatively measure the similarity between two sulcal graphs; **Section 4.2.2** presents the adaptive fusion of the similarity metrics measured in multiple ways; **Section 4.2.3** presents the clustering algorithm that discovers the major sulcal patterns. Finally, **Section 4.2.4** shows the experiments and discusses the results.

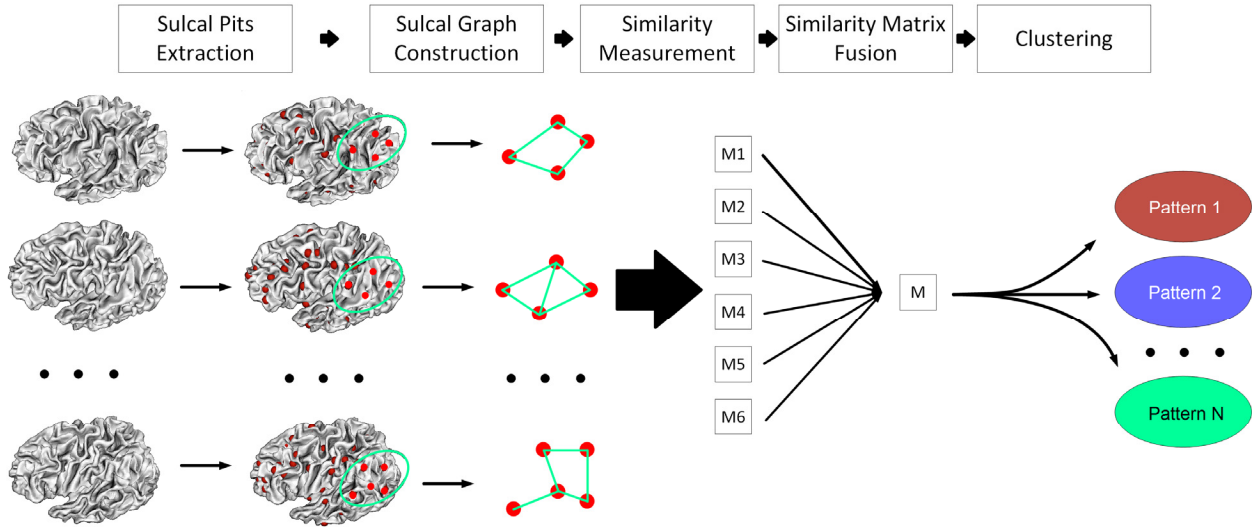


Figure 4.10. Overview of the proposed method for discovering major sulcal patterns.

4.2.1 Sulcal Graph and Similarity Measurement

Sulcal graph is used to characterize the sulcal folding patterns in each individual. A sulcal graph is built based on the sulcal pits and their spatial neighboring relationships. Specifically, each sulcal pit is defined as a node in the sulcal graph, and two nodes are linked by an edge if their corresponding sulcal basins are spatially connected on the cortical surface.

To compare the similarity of two sulcal graphs, their differences are measured using multiple metrics from the spatial, geometrical and topological points of view. Specifically, six distinct metrics are computed according to the sulcal pit position, sulcal pit depth, sulcal basin area, sulcal basin boundary, sulcal pit local connection, and ridge point depth. Each of them is detailed as follows.

(1) Sulcal Pit Position. Given each sulcal pit i in sulcal graph P , its corresponding sulcal pit is found in sulcal graph Q . Because all cortical surfaces have been aligned onto a spherical surface atlas, its closest sulcal pit in Q can be approximately treated as the corresponding sulcal pit of i . The position difference between the sulcal pit i in P and its corresponding sulcal pit in Q , noted as d_i , is measured using their geodesic distance on the spherical surface. Similarly, the position difference between a sulcal pit $j \in Q$ and its corresponding sulcal pit in P is computed as d_j . Then the difference between sulcal pit positions of P and Q is computed as:

$$D(P, Q) = \frac{1}{2} \left(\frac{1}{V_P} \sum_{i \in P} d_i + \frac{1}{V_Q} \sum_{j \in Q} d_j \right) \quad (4.1)$$

where V_P and V_Q are respectively the numbers of sulcal pits in P and Q .

(2) Sulcal Pit Depth. For each subject, the sulcal depth map is normalized by dividing by the maximum depth value of the cortical surface, to reduce the effect of the brain size variation. The depth difference between each sulcal pit $i \in P$ and its corresponding sulcal in Q is denoted as h_i . Similarly, the depth difference between each sulcal pit $j \in Q$ and its corresponding sulcal pit in P is denoted as h_j . Then, the difference between P and Q in terms of sulcal pit depth is defined as:

$$H(P, Q) = \frac{1}{2} \left(\frac{1}{V_P} \sum_{i \in P} |h_i| + \frac{1}{V_Q} \sum_{j \in Q} |h_j| \right) \quad (4.2)$$

(3) Ridge Point Depth. Ridge points are the locations where two sulcal basins meet. As suggested by (Im, et al., 2011b), the depth of the ridge point is an important indicator for distinguishing sulcal patterns. The difference between the average ridge point depths of sulcal graphs P and Q is computed as:

$$R(P, Q) = \left| \frac{1}{E_P} \sum_{e \in P} r_e - \frac{1}{E_Q} \sum_{e \in Q} r_e \right| \quad (4.3)$$

where E_P and E_Q are respectively the numbers of edges in P and Q , e is the edge connecting two sulcal pits, and r_e is the normalized sulcal depth of ridge point in the edge e .

(4) Sulcal Basin Area. To reduce the effect of surface size variation across subjects, the area of each basin is normalized by being divided by the area of the whole cortical surface. Let s_i denote the area difference between the basins of sulcal pit $j \in P$ and its corresponding sulcal pit in Q , and s_j denote the area difference between the basins of sulcal pit $j \in Q$ and its corresponding sulcal pit in P . The area difference between sulcal basins of graphs P and Q is defined as:

$$S(P, Q) = \frac{1}{2} \left(\frac{1}{V_P} \sum_{i \in P} |s_i| + \frac{1}{V_Q} \sum_{j \in Q} |s_j| \right) \quad (4.4)$$

(5) Sulcal Basin Boundary. A vertex is treated as a boundary vertex of a sulcal basin if any of its neighboring vertices belongs to a different basin. Given two corresponding sulcal pits $i \in P$ and $i' \in Q$, the sets of their sulcal basin boundary vertices are respectively denoted as B_i and $B_{i'}$. For any boundary vertex $a \in B_i$, its closest vertex a' is found from $B_{i'}$; and similarly for any boundary vertex $b' \in B_{i'}$, its closest vertex b was found from B_i . Then, the difference between the sulcal basin boundaries of sulcal pits $i \in P$ and $i' \in Q$ is defined as:

$$b_{i,i'} = \frac{1}{2} \left(\frac{1}{N_{B_i}} \sum_{a \in B_i, a' \in B_{i'}} \text{dis}(a, a') + \frac{1}{N_{B_{i'}}} \sum_{b' \in B_{i'}, b \in B_i} \text{dis}(b', b) \right) \quad (4.5)$$

where N_{B_i} and $N_{B_{i'}}$ are respectively the numbers of vertices in B_i and $B_{i'}$; and the function $\text{dis}(\cdot)$ computes the geodesic distance between two vertices on the spherical surface atlas. The difference between the sulcal basin boundaries of pit $j \in Q$ and its corresponding pit $j' \in P$ is defined similarly. Then, the difference between sulcal basin boundaries of graphs P and Q is the average boundary difference over all corresponding sulcal pit pairs, which is formulated as:

$$B(P, Q) = \frac{1}{2} \left(\frac{1}{V_P} \sum_{i \in P} b_{i,i'} + \frac{1}{V_Q} \sum_{j \in Q} b_{j,j'} \right) \quad (4.6)$$

(6) Sulcal Pit Local Connection. The difference of sulcal pit local connection measures how well the distance between two neighboring sulcal pits in a graph is preserved after mapping them to another graph. For a sulcal pit $i \in P$, assume $k \in P$ is one of its spatially connected sulcal pits, and their corresponding sulcal pits in graph Q are respectively i' and k' . The changes of local connection after mapping sulcal pit $i \in P$ to graph Q are measured by:

$$c_i = \frac{1}{N_{G_i}} \sum_{k \in G_i} |\text{dis}(i, k) - \text{dis}(i', k')| \quad (4.7)$$

where G_i is the set of all sulcal pits that connect to i , and N_{G_i} is the number of sulcal pits in G_i . The change of local connection when mapping a pit $j \in Q$ to graph P is measured similarly. Thus, the difference between local connections of two graphs P and Q is defined as:

$$C(P, Q) = \frac{1}{2} \left(\frac{1}{V_P} \sum_{i \in P} C_i + \frac{1}{V_Q} \sum_{j \in Q} C_j \right) \quad (4.8)$$

4.2.2 Fusion of Sulcal Graph Similarities

Given N sulcal graphs from all subjects, any two of them are compared using the above six metrics, so a $N \times N$ difference matrix is constructed for each metric. Each of the metric measures the inter-individual difference of sulcal graphs from different points of view and provides complementary information to each other. To leverage all these information, a similarity network fusion (SNF) method (Wang, et al., 2014a) is employed to adaptively integrate all six metrics together. SNF can both keep the shared information and capture the complementary information from different measurements. To perform SNF, each difference matrix is normalized by its maximum element, and then transformed into a similarity matrix as:

$$W_M(x, y) = \exp\left(-\frac{M^2(x, y)}{\mu\left(\frac{\Phi_x + \Phi_y + M(x, y)}{3}\right)}\right) \quad (4.9)$$

where μ is a scaling parameter; M could be anyone of the above six matrices; Φ_x and Φ_y are respectively the average values of the smallest K elements in the x -th row and y -th row of M . Finally, six similarity matrices are nonlinearly fused together as a single similarity matrix by using SNF with t iterations. The parameters are set as $\mu = 0.8$, $K = 30$, and $t = 20$ as suggested in (Wang, et al., 2014a).

4.2.3 Sulcal Pattern Clustering

To cluster sulcal graphs into different groups based on the fused similarity matrix, the Affinity Propagation Clustering (APC) algorithm (Frey and Dueck, 2007) is chosen to be used. There are some other clustering algorithms such as K-Means or Spectral Clustering algorithm (von Luxburg, 2007). These methods require users to explicitly choose a cluster number, but there is no prior knowledge about the number of sulcal patterns. Because APC could automatically determine the number of clusters based on the natural characteristics of data, it is more suitable here over the other clustering algorithms. However, since sulcal folding patterns are extremely variable across individuals, too many clusters would be identified after performing APC, making it difficult to observe the most important major patterns. Therefore, a hierarchical APC framework is proposed to further group the clusters. Specifically, after running APC, the exemplars of all clusters are treated as a new dataset to perform another APC. With the exemplars merging into new clusters, their corresponding old clusters are also merged. In this way, less clusters are generated. However, as the old clusters merge, the previous exemplars may be no longer representative for the new clusters. Thus, a new exemplar is selected for each new cluster by maximizing average similarity to all the other samples in the cluster. These steps can be repeated until the cluster number reduces to an expected level (≤ 5).

4.2.4 Experiments and Results

Sulcal pits were extracted from 677 neonatal cortical surfaces. To demonstrate the validity of the proposed method, three representative cortical regions (i.e., central sulcus, superior temporal sulcus, and cingulate sulcus) were selected for exploring the major sulcal patterns. For each cortical region, a 677×677 similarity matrix was computed using SNF and all subjects were then clustered into different groups by the hierarchical APC. To visually inspect the major sulcal patterns, an average cortical surface was constructed for each cluster based on the top 10% representative cortical surfaces, which were most similar to the exemplar in each cluster. All sulcal pits in each cluster were further mapped onto the average surfaces.

Results

For the central sulcus, three distinct folding patterns were identified, as shown in **Figure 4.11**. In the pattern (a), two sulcal pits concentration areas can be observed, indicating two sulcal basins in the central sulcus. This pattern is further confirmed by six representative examples of individual subjects (in the second to seventh columns). In the pattern (b), three distinct sulcal pits concentration areas can be observed, with one extra area (basin 3) located in the most inferior portion of the central sulcus, compared to the pattern (a). In the pattern (c), three distinct sulcal pits concentration areas can be observed as in the pattern (b), but they are more concentrated. This is also confirmed by six representative examples of (c). Moreover, compared to the pattern (b), the sulcal basin 2 is very short, while the sulcal basin 3 is very long in the pattern (c). Such phenomenon is likely related to the “hand knob shift” in a study of the central sulcus shape in adults (Sun, et al., 2012). Previously, different studies reported either two (Meng, et al., 2014) or three (Im, et al., 2010) sulcal basins in the central sulcus. Herein, we can see that both two-basin and three-basin patterns are the major patterns of sulcal folding.

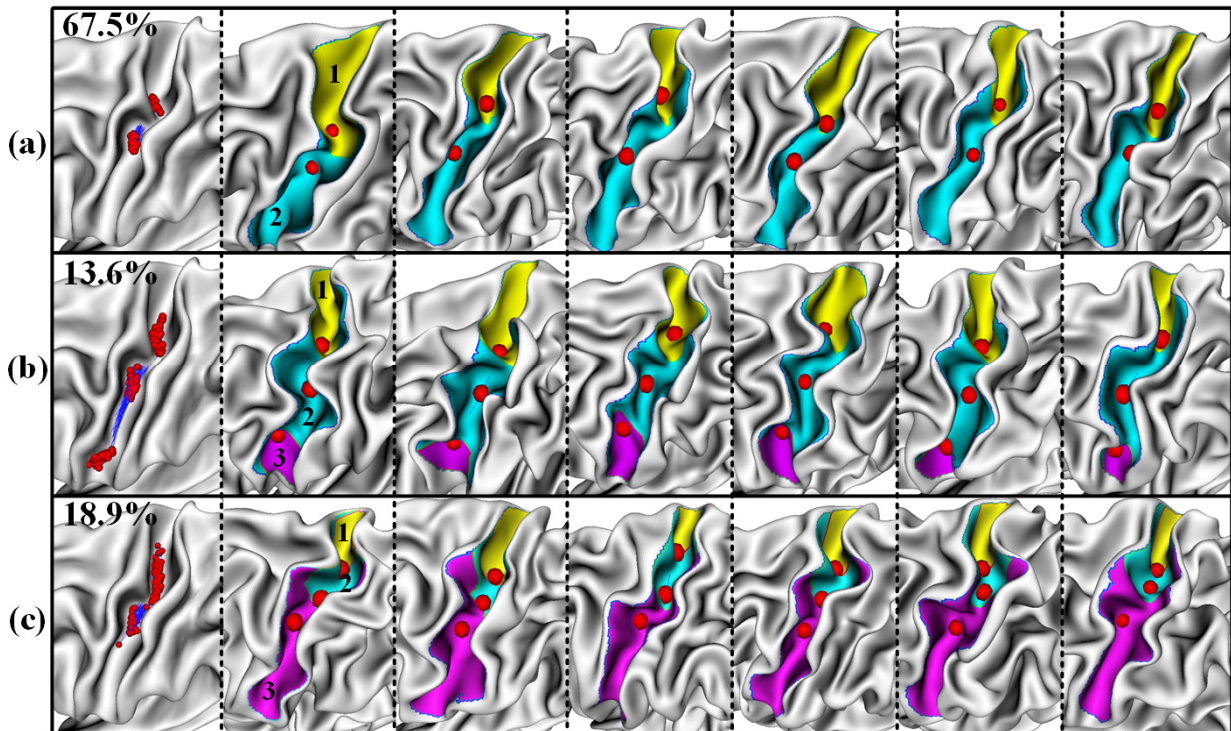


Figure 4.11. Sulcal patterns in the central sulcus. The first column shows three discovered sulcal folding patterns, with all sulcal pits (red spheres) mapped onto the average surface of each cluster. For each pattern, the second to seventh columns show six representative examples of individual subjects. Different sulcal basins are marked with different colors. The percentage of each pattern is shown at the top-left corner.

For the superior temporal sulcus (STS), three distinct folding patterns were identified, as shown in **Figure 4.12**. In the pattern (a), the distribution of sulcal pits in the posterior portion of STS is more diffused and bended, compared to the patterns (b) and (c), indicating the differences in the folding shape of STS. This is supported by a previous cortical folding study in adults, which reported a Y-shaped STS for some brains while a single long STS for other brains (Sun, et al., 2009). In the pattern (b), compared to (a) and (c), an extra concentration region of sulcal pits is exhibited near the temporal pole, which is again confirmed by six representative examples from individual subjects, showing small sulcal basins near the temporal pole. In the pattern (c), the sulcal basin in the anterior portion of STS is very long and straight, extending to the temporal pole.

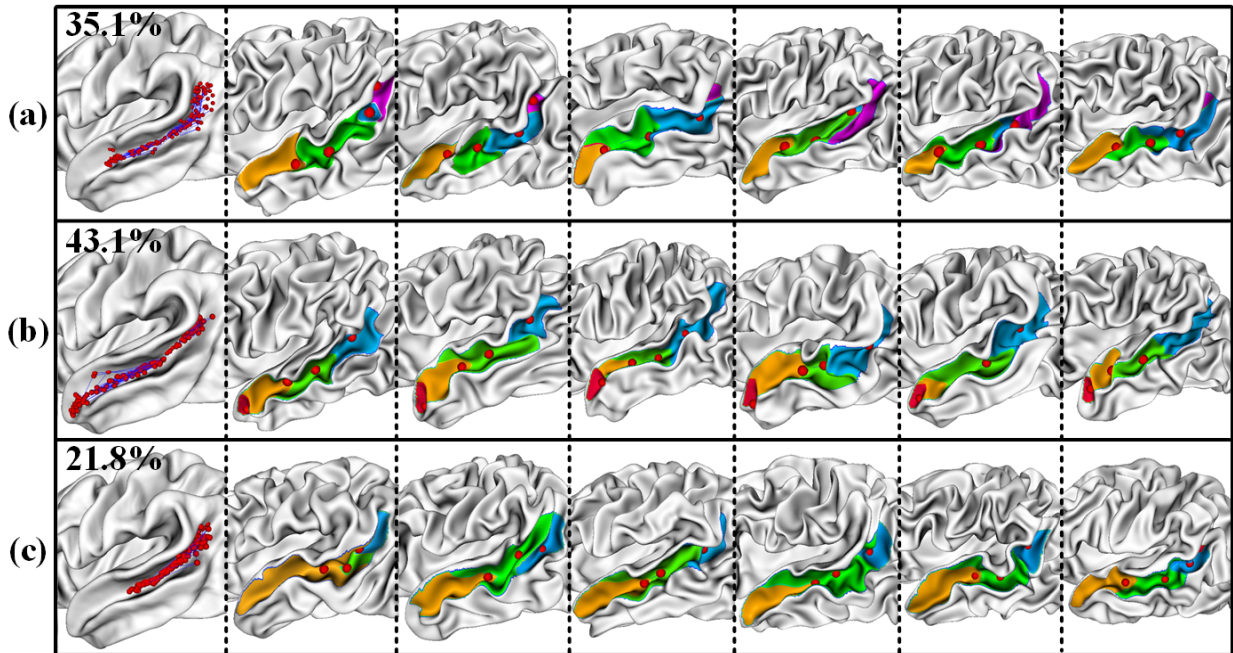


Figure 4.12. Sulcal patterns in the superior temporal sulcus. The first column shows three discovered sulcal folding patterns, with all sulcal pits (red spheres) mapped onto the average surface of each cluster. For each pattern, the second to seventh columns show six representative examples of individual subjects. Different sulcal basins are marked with different colors.

For the cingulate sulcus, four distinct major folding patterns were identified, as shown in **Figure 4.13**. In the pattern (a), a single long cingulate sulcus is clearly shown, while in the pattern (b) two long parallel sulci are observed. This is consistent with the previous cortical folding pattern study in adults (Sun, et al., 2009), which reported observation of two cingulate sulci in some brains. A study of autopsy specimen brains also reported that 24% left hemispheres had double parallel cingulate sulcus (Ono, et al., 1990). In the pattern (c), the cingulate sulcus is interrupted in the anterior region; in contrast, in the pattern (d), the cingulate sulcus is interrupted in the posterior region. This two types of interruption were also reported in (Ono, et al., 1990). In pattern (c) and pattern (d), some parallel sulci can also be observed, but they are much shorter than those in pattern (b).

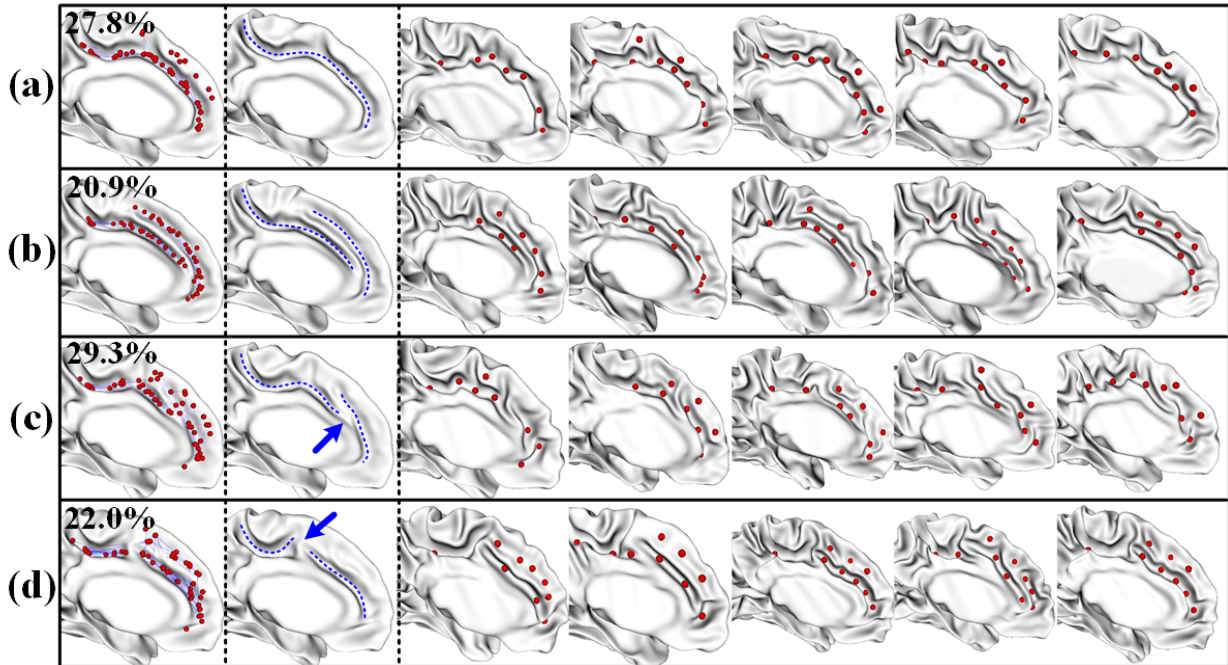


Figure 4.13. Sulcal patterns in the cingulate sulcus. The first column shows four discovered folding patterns, with all sulcal pits (red spheres) mapped onto the average surface of each cluster. The second column shows the schematic drawing of the sulcal curves (blue dashes) on the average surface of each cluster. For each pattern, the third to seventh columns show five representative examples of individual subjects. The percentage of each pattern is shown at the top-left corner.

Differences with Previous Studies

The above discoveries of major patterns in the central sulcus, superior temporal sulcus, and cingulate sulcus in neonates exhibit a lot of consistencies with the previous studies of adults. But the proposed method also reveals new patterns.

To the best of my knowledge, this is the first time to explicitly categorize the patterns of the central sulcus into three classes. One previous study of central sulcus patterns reports that the central sulcus morphology is handedness-related (Sun, et al., 2012), and one or double “hand-knob” may appear in different positions of the central sulcus. However, this study does not categorize the central sulcus folding patterns. Even in another study of building the dictionary of brain folding patterns (Sun, et al., 2009), the sulcal patterns are still not categorized for the central sulcus. Both of these studies use the same method to measure the similarity between sulcal patterns. In particular,

this method scales and rigidly aligns the local cortical surface of the central sulcus of different brains, and computes the distance between two aligned local surfaces as their similarity. If this similarity measurement is used to cluster the central sulcal patterns, the results would be quite different from the results reported above. For example, as shown in **Figure 4.14**, using the method in (Sun, et al., 2009), patterns in (a) and (b) will be grouped into the same cluster, and also pattern (c) and (d) will be grouped into another cluster. This is because the distance measurement treats every vertex in the central sulcus equally and reflects the similarity between the entire shapes of the central sulci. However, the proposed sulcal graph-based method will group patterns (c) and (d) into a cluster, because each of them has two sulcal basins (or sulcal pits); while patterns (a) and (d) will be separated, because each of them has three sulcal basins (or sulcal pits) and the boundary location between the pink and blue basins is quite different.

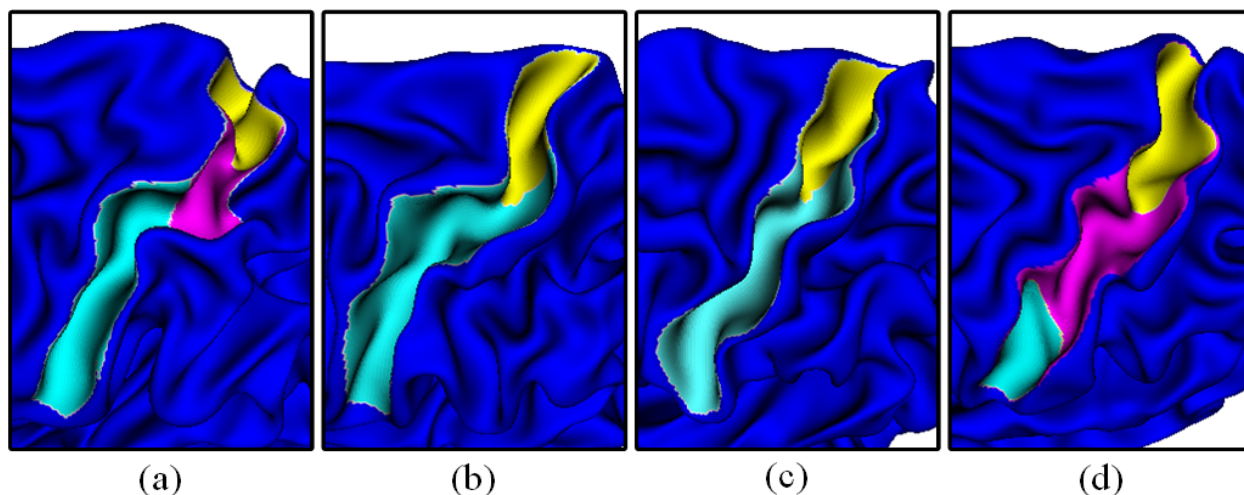


Figure 4.14. Examples of different central sulcal patterns. By comparing the sulcal shapes, sulcus (a) and sulcus (b) are similar with their curved shapes, while sulcus (c) and sulcus (d) are similar with their relatively straight shapes. By comparing the sulcal patterns, sulcus (b) and sulcus (c) are similar as they both have 2 sulcal basins, while sulcus (a) and sulcus (d) are different in the sizes of the sulcal basins, though they both have 3 sulcal basins.

In the superior temporal sulcus, the proposed sulcal graph-based method reports a new discovered major pattern, which exhibits a sulcal basin at the temporal pole. On the contrary, the method in (Sun, et al., 2009) is not able to discover this pattern, because of the same reason as discussed above.

Number of Clusters

The hierarchical APC algorithm utilized in the proposed method automatically decides the number of patterns. In order to validate the clustering method and as well as the number of clusters, another advanced clustering algorithm, namely spectral clustering, is also applied to the comprehensive similarity matrix to group the sulcal graphs into different number of clusters. **Figure 4.15** shows an example of comparison results in the cingulate sulcus. The experiments show that if the cluster number is set as same as the hierarchical APC reported, spectral clustering algorithm could produce very similar results. If the cluster number is set differently, some reported patterns are difficult to interpret or summarize. For example, in **Figure 4.15**, if the cluster number is set as 3, two major patterns in A-1 and A-2 can still be observed in B-1 and B-2, but the pattern in B-3 looks like a mixed combination of A-1 and A-3. If the cluster number is set as 5, all patterns from A-1 to A-4 can be observed in the results from D-1 to D-5. However, there is no significant difference between D-4 and D-5. The same phenomenon can also be seen from E-5 to E-6 if the cluster number is set as 6. This experiment suggests that the hierarchical clustering algorithm works well in exploring the major sulcal patterns. However, it should be pointed out the possible existence of other major patterns, which may be not discovered due to the possible limitations of the sulcal graph-based shape descriptor.

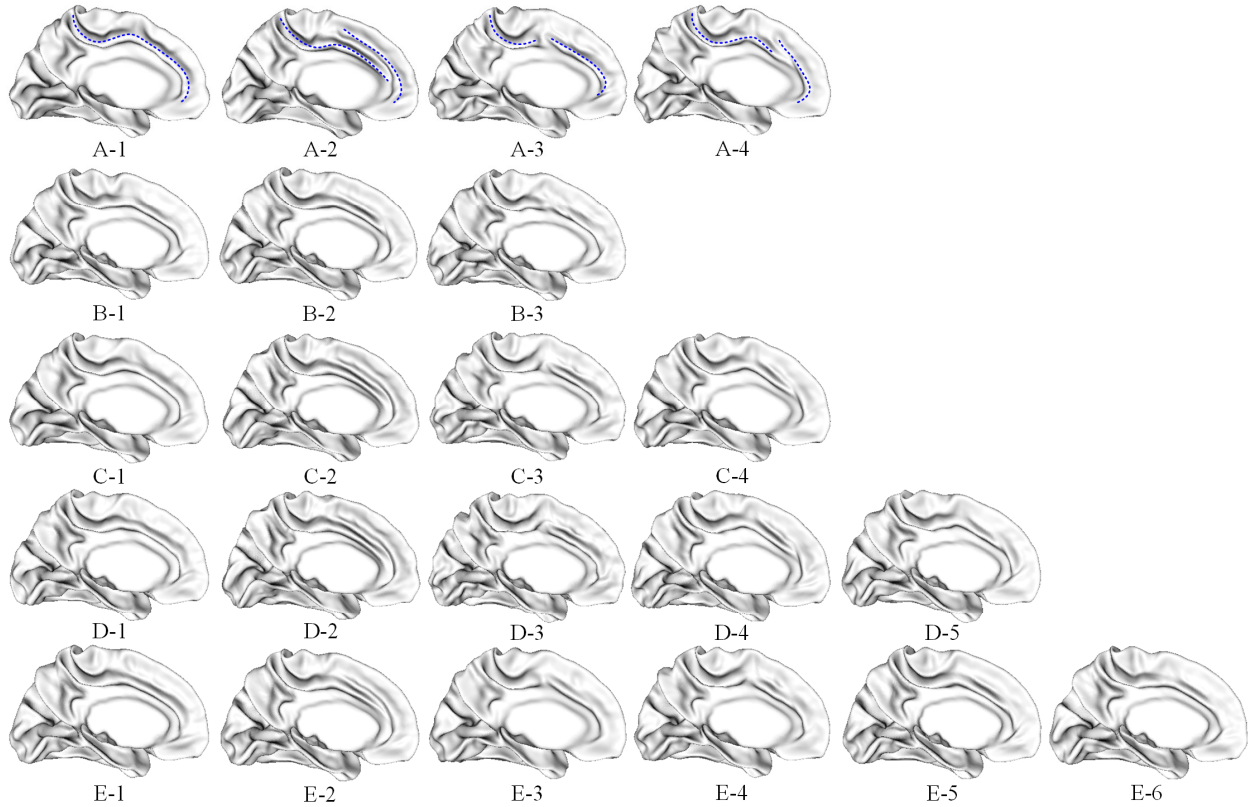


Figure 4.15. Different numbers of clusters in the cingulate sulcus. The first row is the results of the proposed method. The last four rows show the results using spectral clustering algorithm to generate 3, 4, 5, and 6 different patterns.

Reliability

To evaluate the reliability of the proposed method, the experiments were repeated 40 times with each time randomly excluding 5% ~ 10% subjects. For the most of the time, the proposed method reports the same number of major patterns as in **Figures 4.11, 4.12, and 4.13**, but sometimes one more or less pattern is reported. **Table 4.1** provides the ratio of the proposed method reporting different number of patterns. From the table, we can have two important observations. First, the proposed method tends to discover 3, 3, and 4 major sulcal patterns respectively in the central sulcus, superior temporal sulcus, and cingulate sulcus. Second, the proposed method is more stable in the central sulcus, where the sulcal pattern is relatively simple, but is less stable in the superior temporal sulcus and cingulate sulcus, where the sulcal patterns are

more complex and variable. Note that, for experiments that report the same number of patterns, their corresponding major patterns, which are observed by averaging top 10% representative surfaces, are also very similar. **Table 4.2** further provides the discovery rate of each reported pattern in **Figures 4.11, 4.12, and 4.13**. These results suggest that reported major patterns are the significant ones, as they can be repetitively discovered in cross validation. It also indicates that the proposed method is reliable for detecting these major sulcal patterns. It is worth noting that, if spectral clustering algorithm is used instead of hierarchical affinity propagation in the last step of the proposed method and also the certain number of clusters is given, the discovery rates for all major patterns could reach 100%. This may suggest that 1) the sulcal graph-based pattern similarity measurement is sufficiently stable; and 2) although hierarchical APC is able to automatically determine the number of patterns, it is not as stable as spectral clustering algorithm. Thus, in order to discover major sulcal patterns, the proposed method could be performed multiple times in a manner of cross validation, and then the majority of results can be selected as the final results.

Table 4.1. Ratios of different reported number of patterns.

Number of patterns	2	3	4
Central Sulcus	5%	92.5%	2.5%
Superior Temporal Sulcus	5%	82.5%	12.5%
Cingulate Sulcus	2.5%	10%	87.5%

Table 4.2. Pattern discovery rate.

Pattern Index	(a)	(b)	(c)	(d)
Central Sulcus	100%	95%	100%	-
Superior Temporal Sulcus	97.5%	100%	97.5%	-
Cingulate Sulcus	100%	100%	87.5%	87.5%

4.3 Utilization of Sulcal Patterns to Help Predict Cortical Attributes Development

4.3.1 Motivation

Many cortical morphological attributes are, to some degree, correlated to cortical folding patterns. Intuitively, if cortical folding patterns in a specific location of two cortical surfaces are similar, the corresponding local cortical attribute maps could be also similar. That is why the Haar-like features computed based on the sulcal depth map are included as inputs in DARF for predicting the development of cortical thickness map. However, Haar-like features are the randomly selected mean values of neighborhoods or regional differences between two neighborhoods, which have limited ability to describe the local cortical folding patterns. Sulcal graph, which is used to discover major sulcal patterns, on the other hand, captures more meaningful and higher level information about the characteristics of cortical folding. Therefore, if the features of sulcal graph could be effectively integrated into DARF, the accuracy of predicting the development of cortical attribute maps could be boosted.

4.3.2 Method

In order to make use of sulcal graphs to help predict cortical attribute development, the following sulcal pattern related feature maps are computed and added into the feature vectors when using DARF: (1) basin area map, (2) local pit density map, (3) pit degree map, (4) pit depth map, and (5) pit distance map. Specifically, each vertex on the surface is associated with a “home sulcal basin” and a “home sulcal pit”. The “home sulcal basin” is the sulcal basin where the given vertex lies, and the “home sulcal pit” is the sulcal pit in the “home sulcal basin”. Each sulcal basin has one and only one sulcal pit. For a given vertex, its corresponding value in the basin area map is the area of its “home sulcal basin”; its corresponding value in the local pit density map is the number of sulcal pits in its neighborhood; its corresponding value in the pit degree map is the

degree of its “home sulcal pit” in the sulcal graph representation; its corresponding value in the pit depth map is the depth of its “home sulcal pit”, and its corresponding value in the distance map is the Euclidian distance from the given vertex to its “home sulcal pit”. Because sulcal basins have clear boundaries, the computed feature maps may have sharp value changes on the sulcal basin boundaries, namely the boundary effect. Such boundary effect is numerically sensitive and also varies largely in different surfaces, thus it brings a lot of “uncertainty” to the DARF model and consequently lowers the prediction accuracy. To reduce the boundary effect, the sulcal pattern related feature maps are smoothed. **Figure 4.16** shows the smoothed feature maps of a randomly selected subject.

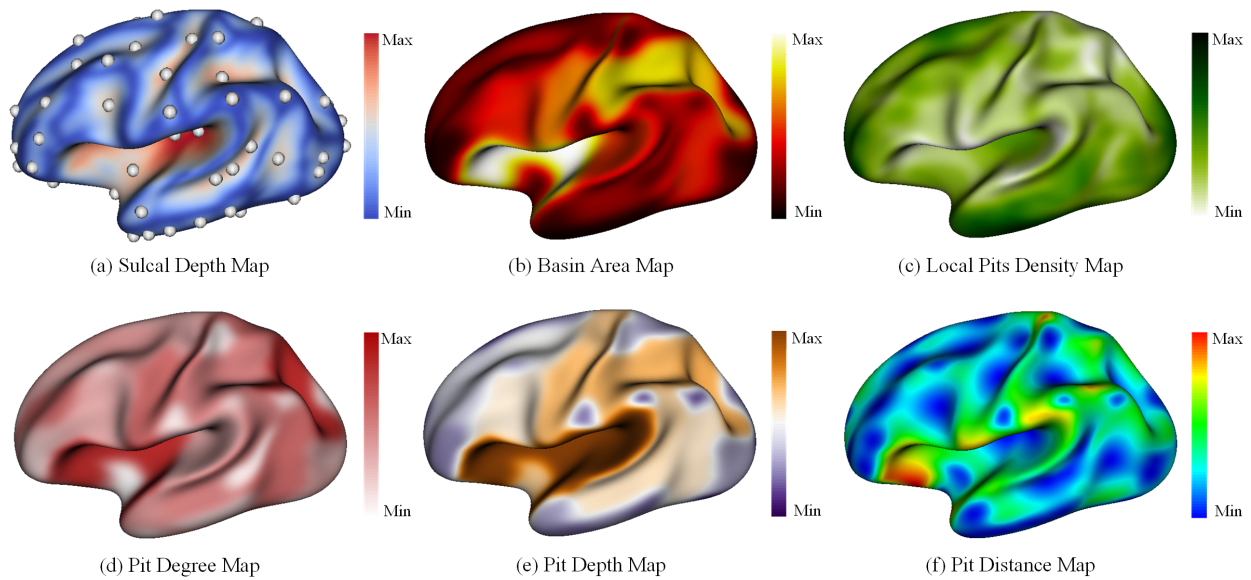


Figure 4.16. Sulcal pattern related feature maps. The maps are displayed on the inflated cortical surface atlas. In (a), sulcal pits are drawn as white balls.








4.3.3 Experiments and Results

To test how the sulcal pattern related features can help DARF, the experiment of predicting cortical thickness map was carried out again as in **Section 3.1.4**. In the experiment, the prediction target was the 9-month-old cortical thickness map, and the inputs were the feature maps at birth,

including cortical thickness map, sulcal depth map, and all above sulcal pattern related feature maps. The experiments were repeated using leave-one-out cross validation with 23 subjects.

The prediction result was compared with the previous one, which was predicted without using sulcal pattern related feature maps. Overall comparison shows that the improvement measured for the whole cortical thickness map is subtle, i.e., reducing MSE from 0.34 (mm) to 0.33 (mm) and reducing MRE from 12.4% to 12.1%. However, ROI-based comparison reveals that the improvement is significant in many regions in the high-order association cortex with more variable folding patterns. **Table 4.3** reports the quantitative error estimations in 7 ROIs, where the improvement is significant (with $p < 0.05$ in a pairwise T-test).

Table 4.3. Quantitative comparison of error estimations in 7 ROIs.

ROI	Name	MSE (no pattern features)	MSE (with pattern features)	MRE (no pattern features)	MRE (with pattern features)
	Supramarginal Gyrus	0.382±0.045	0.344±0.041	13.7±1.8	12.3±1.7
	Pars Triangularis	0.412±0.07	0.383±0.11	13.6±3.1	12.8±3.3
	Inferior Temporal Gyrus	0.376±0.054	0.356±0.049	13.3±2	12.7±1.9
	Superior Frontal Gyrus	0.405±0.072	0.389±0.073	13.1±1.6	12.8±1.6
	Precuneus Cortex	0.352±0.053	0.341±0.057	13.5±2	13.2±2.1
	Rostral Anterior Cingulate Cortex	0.507±0.272	0.483±0.244	14.1±5.1	13.4±4.9
	Caudal Anterior-Cingulate Cortex	0.386±0.148	0.365±0.142	13.7±4.1	13.1±4.0

Moreover, the contribution of each of the five sulcal pattern related feature maps was investigated. In particular, the above experiment was repeated with each time including or excluding only one of the five feature maps. The experimental results show that including pit distance map alone could improve the prediction accuracy slightly, but the improvement is not as

significant as including all 5 feature maps. The experimental results also show that including each of the other four feature maps separately cannot achieve any noticeable improvement. This may suggest that such sulcal pattern related feature maps provide complementary information, and should be used together to help DARF predict the development of cortical thickness map. Additionally, the experimental results by excluding one of the five feature maps shows that discarding pit depth map didn't reduce the prediction accuracy, which indicates that the pit depth map doesn't help in predicting the development of cortical thickness map. One explanation might be that the smoothed pit depth map is very similar or correlated to the sulcal depth map, therefore the pit depth map doesn't provide new valuable information if the sulcal depth map has already been used.

4.4 Summary

Cortical folding pattern is an important attribute of the cerebral cortex. It links to cognitive functions and neurological disorders. This chapter presented several studies and techniques for studying cortical folding patterns. **Section 4.1** introduced the watershed algorithm for sulcal pits extraction. As the brain grows rapidly in early development stage, the original watershed algorithm developed for the adult brains cannot be directly applied to the infant cortical surfaces. Thus **Section 4.1** introduced how the watershed algorithm could be adapted for extracting sulcal pits from the rapid growing infant cortical surfaces. Moreover, the spatial and temporal consistency of the distribution of sulcal pits was validated, thus sulcal pits could be utilized for comparing sulcal patterns in both infants and adults. **Section 4.2** proposed the method for discovering major sulcal patterns in the specific cortical regions by using sulcal pits from a large-scale dataset of neonates. The proposed method first built a sulcal graph of a certain cortical region for each subject based on the local sulcal pits distribution, and then the similarity between any two sulcal graphs were

measured in six different ways, thus forming six similarity matrices. To capture both the common and complementary information in these six similarity matrices, they were further fused together using SNF. Finally, the fused similarity matrix was fed to the hierarchical APC, to categorize the cortical regions into different classes of major sulcal patterns. The proposed method was validated on a dataset with 677 neonatal brains, and meaningful major sulcal patterns were revealed from the central sulcus, superior temporal sulcus, and cingulate sulcus. Motivated by the relevance of cortical morphological attributes to the sulcal patterns, **Section 4.3** further investigated whether sulcal pattern information could help better predict the development of cortical thickness map. In particular, five sulcal pattern related features maps were computed and fed into DARF for the prediction. The results indicated that the sulcal pattern information could improve the prediction accuracy in many cortical regions with highly variable folding patterns.

The above proposed work also has some weaknesses. First, the hierarchical APC, which is used for grouping sulcal patterns, is not very stable. Repetitive experiments show that sometimes hierarchical APC reports different numbers of clusters. The instability is more serious for the cortical regions with more complex structures. This suggests that to use the proposed method to detect major sulcal patterns, multiple experiments with each time using a different subset of the whole dataset are needed, and the reliable conclusion can be obtained from the result with the highest repetitive rates. Second, the proposed method has been validated in only three primary cortical sulcal regions, and whether it also works well in other cortical regions is not clear. Future work may include testing the proposed method in more cortical regions, or even for the entire cortical surface. However, due to structural complexity of the entire cortical surface as well as huge diversity across subjects, categorizing sulcal patterns for the whole brain is a more challenging work. Third, only five sulcal pattern related feature maps have been investigated for

helping predict the development of cortical thickness map. However, the sulcal pattern can be measured and encoded in many other different ways, such as using ridge point depth or sulcal basin boundary. Hence, future work may also include computing more meaningful feature maps to better encode sulcal pattern information for the prediction of cortical attributes development.

5 SUMMARY AND FUTURE WORK

5.1 Summary

This section rephrases the contribution claims that are made in **Chapter 1**. After each contribution claim, how that claim is addressed is briefly summarized. Finally, the thesis statement is revisited.

The contributions of this dissertation are as follows:

(1) A novel prediction model, Dynamically-Assembled Regression Forest (DARF), is proposed for accurately estimating the early development of cortical attribute maps from birth to 1 year of age. By dynamically grouping and sharing local decision trees, DARF is able to produce accurate and spatially smooth prediction results and is also computationally efficient.

The Dynamically-Assembled Regression Forest (DARF) was presented in **Section 3.1.1**. Different from the conventional regression forest model, in which a fix number of decision trees are specifically trained for a forest, in the implementation of DARF a single decision tree is trained at each vertex and then shared by many local forests. In particular, in the training stage a single decision tree is trained at each vertex using its neighboring vertices as training samples. In the testing stage, the cortical attribute value of each vertex is predicted using a vertex-specific forest, which is formed by grouping the nearby decision trees around a given vertex. The experiments in **Section 3.1.4** show that DARF is able to accurately predict the normal development of cortical thickness map during the first year of life.

(2) *A surface-based feature computation method is proposed for extracting Haar-like features from spherical surface. By encoding rich neighboring information of local cortical attribute maps, surface-based Haar-like features make important contributions in training accurate DARF models for the prediction of developmental cortical attributes.*

Surface-based Haar-like features were introduced in **Section 3.1.2**. Haar-like features are the randomly generated features, which could either be a mean value of a local cortical attribute map or a difference between the mean values of two local cortical attribute maps. The conventional way of computing Haar-like features in the 2D or 3D image space cannot be directly applied to the surface manifold, due to the lack of Cartesian coordinate system in the surface manifold. The proposed method projects each local cortical region to a corresponding tangential plane, and computes local Haar-like features in this tangential plane. In this way, the generated local Haar-like features are comparable within each local cortical region.

(3) *A novel missing data estimation strategy, which consists of “pairwise estimation” and “joint refinement”, is proposed particularly for recovering the missing cortical attribute maps in incomplete longitudinal datasets.*

Section 3.2.1 presented this missing data estimation strategy in detail. In the missing data estimation, the core problem is how to effectively use the available data. Typically, there is a conflict between using more time points and using more training subjects. The proposed method solves this conflict. In particular, the stage of “pairwise estimation” focuses on using more training subjects to produce the initial estimations of missing data. This enables the stage of “joint refinement” to use as many training subjects and time points as possible. The experiments in **Section 3.2.2** illustrate the effectiveness of the proposed strategy. Moreover, the experiments also show that the “pairwise estimation” stage prepares a good initialization for the “joint refinement”

stage, and the “joint refinement” stage significantly improves the initial estimation in the “pairwise estimation” stage.

(4) *Extensive experiments and comparisons show that the proposed methods can accurately estimate the development of cortical thickness map and missing data, and also show that DARF outperforms four existing regression methods in this task.*

The experiments and comparisons were presented in **Section 3.1.4** and **Section 3.2.2**. The methods were tested on a longitudinal dataset of 31 infants \times 5 time points, with some missing data. I first tested DARF with 15 subjects with complete data at all 5 time points for predicting the development of cortical thickness map, and then validated the missing data estimation strategy using the whole incomplete dataset. Moreover, I also compared DARF with four existing regression methods, including the mixed effect model (MEM), polynomial regression (PR), conventional regression forest (CRF), and sparse linear regression (SLR). The experiments show that 1) DARF is able to accurately predict the development of cortical thickness map; 2) DARF outperforms all other four regression methods in terms of accuracy and output smoothness; 3) The proposed missing data estimation strategy can make a better use of the existing data to recover the missing cortical thickness maps effectively.

(5) *A watershed algorithm, which is originally tuned for adult brains, is adapted for extracting sulcal pits from the cortical surfaces of infants at different ages.*

This work was presented in **Section 4.1.1**. The sulcal pits extraction algorithm is controlled by four parameters: depth threshold, area threshold, distance threshold, and ridge height threshold. These four parameters were previously tuned for the adult brains with similar brain sizes, and thus are not applicable to the developing infant brains with variable brain sizes. To find the appropriate parameters for the infants at any age, a grid searching is performed first, and then the factors that

influence parameter selections are explored. It is found that the ridge height threshold is independent of the brain size and can be set as 2.5mm. The distance threshold can be set as 10 (rings). Since all cortical surfaces are resampled using a uniform triangular mesh, this parameter could automatically fit the infant cortical surfaces of different sizes. It is also found that the depth threshold can be set by following a linear relationship with the maximum sulcal depth, and the area threshold can be set by following a linear relationship with the total cortical surface area.

(6) The spatial and temporal consistency of the distribution of sulcal pits in the rapid developing infant brains is validated.

Section 4.1.2 presented the study of spatial distribution and longitudinal development of infant sulcal pits. For this study, the sulcal pits were extracted from the cortical surfaces of 73 infants, respectively, at birth, 1-year-old, and 2-year-old, and also from the cortical surfaces of 64 young adults. The spatial distributions of sulcal pits were compared across both subjects and ages. The experiments show that the spatial distribution of sulcal pits in infant brains is as consistent as that in adult brains. This suggests that the infant sulcal pits can be used as stable landmarks for the study of sulcal patterns.

(7) A new framework is proposed for automatically discovering the major sulcal patterns in local cortical regions in neonatal cortical surfaces.

The proposed framework was introduced in **Section 4.2**. In this framework, sulcal graphs are built for characterizing the sulcal patterns using the deep sulcal pits, due to their spatial consistency. In order to comprehensively compare the sulcal patterns, the similarity between two sulcal graphs are estimated from six different points of view and then adaptively fused together. The sulcal patterns are categorized using a hierarchical APC method, which could automatically determine the number of categories. For a better visual inspection of the major sulcal patterns of

each category, an average cortical surface is computed using the most representative subjects for each major pattern. The validation experiments were performed on a large-scale dataset of 677 neonatal brains, and three primary cortical regions were selected for the tests, as introduced in **Section 4.2.4**. The experiments show that 1) the proposed framework is able to reveal both the major sulcal patterns reported in previous adult studies and the newly discovered major sulcal patterns; 2) to obtain reliable results, the proposed method would better be executed multiple times with different subsets, and then the majority of the results could be used as the reliable patterns.

(8) Whether and how sulcal pattern information could help DARF for better predicting the development of cortical attribute maps are investigated.

The investigation was presented in **Section 4.3**. In this investigation, five sulcal pattern related feature maps (basin area, local pit density, pit degree, pit depth, and pit distance) were computed and fed into the feature vectors used in DARF to predict the development of cortical thickness map. The experiment results indicate that, while the overall benefit from sulcal pattern related feature maps is subtle, these features could make apparent improvement in predicting the development of cortical thickness in some cortical regions with highly-variable folding patterns.

***Thesis:** Dynamically-Assembled Regression Forest (DARF) is able to accurately estimate the early development of cortical attribute maps from birth to 1 year of age. Sulcal pits, which have relatively stable spatial distributions across ages and individuals, can be utilized for discovering the major sulcal patterns. Sulcal pattern information can further improve the performance of DARF for estimating cortical attribute maps.*

Estimating the rapid development of vertex-wise cortical attributes is a challenging task. To effectively address this task, this dissertation presented Dynamically-Assembled Regression Forest (DARF). Different from the conventional regression forests, DARF adopts a decision tree

sharing technique. In particular, a single decision tree is trained at each vertex using its neighboring vertices as training samples, and the local neighboring trees are grouped as a vertex-specific forest in the testing phase. This vertex-specific forest captures more regionally detailed information than the conventional regression forest trained for the whole brain, thus able to predict the cortical attributes more precisely. Moreover, since nearby forests share a large portion of same decision trees, the predicted cortical attribute map is spatially smooth, which is difficult to achieve using the ROI-based regression forests. By integrating surface-based Haar-like features into DARF, the relevance of the prediction target and the regional context can be explored and used to benefit the prediction. Additionally, DARF can also be used for estimating the missing cortical attributes in the incomplete longitudinal datasets. The experiments and comparisons show that DARF is an effective technique in the prediction/estimation of cortical attribute maps, and it also outperforms four popular existing methods.

To discover the major sulcal patterns, this dissertation adapts the method of sulcal pits extraction for adult brains to infant brains, and validates the spatial and temporal consistency of sulcal pits in the infant cortical surfaces at different ages. Then, the sulcal pits are used as landmarks to form local sulcal graphs, which are the abstract descriptors of sulcal patterns. To discover the major sulcal patterns, a sulcal graph-based clustering framework is applied to a large dataset with 677 neonatal cortical surfaces. The results show that multiple new and previously reported major sulcal patterns have been successfully found via the proposed method. Motivated by the observed correlation between the sulcal patterns and the cortical thickness map, this dissertation further encodes the sulcal pattern information into multiple feature maps, and uses them to help predict the development of cortical thickness map. Experiments confirm that the sulcal pattern information is helpful in improving the prediction accuracy in some cortical regions.

5.2 Future Work

This dissertation focuses on developing methods for two purposes: 1) estimating the development of cortical attribute maps, and 2) discovering the major cortical folding patterns. For each of them, a few future directions that may be interesting to explore are briefly discussed below.

For the estimation of cortical attribute maps, several potential future works are listed as follows.

Abnormal Brains. The current research is focused on modeling the normal development of cortical thickness map of healthy infants; but, due to the lack of data, the proposed method has not been tested on the infants with neurodevelopmental disorders. Since it has been reported that cortical thickness is correlated to many diseases (e.g., williams syndrome, attention-deficit/hyperactivity disorder, autism, and bipolar disorder), modeling the abnormal cortical thickness map development has important clinical meanings. On one hand, building a cortical thickness development model for a certain disease could help better understand the development and effects of that disease. On the other hand, a prediction model trained with a dataset that mixes both normal and abnormal subjects is likely to be useful for estimating the risk of a new subject having brain diseases, for the purpose of early diagnosis.

Multiple Cortical Attributes. The methods proposed in this dissertation have been tested for estimating only the development of cortical thickness maps. However, these methods are generally designed for estimating the development of any cortical morphological attribute (e.g., sulcal depth, local gyrification index, and myelin content). Because the development of different cortical attributes may be correlated, the proposed method could be extended to estimate the development of multiple correlated cortical attribute maps jointly. Since the correlation in these maps could help each other in the joint estimation, the estimation accuracy might be improved

compared to the current case of independent estimation. Additionally, applying the proposed missing data estimation strategy to multiple cortical attributes could help establish a “complete” multimodal longitudinal dataset, which could further benefit other studies.

Feature Learning. Currently, in our prediction model, the regional context information comes partially from Haar-like features and partially from the sulcal pattern related features. Haar-like features are randomly generated, and thus, to capture real useful context information, a large number of Haar-like features are needed. This is not very efficient. Sulcal pattern related features are manually designed to encode the sulcal pattern information. This manual design process is heuristic, and thus is neither efficient. To better explore regional context information, the convolutional neural network, which is widely used to learn useful features in image analysis, might be adaptively modified to learn more useful regional context features for cortical surfaces-based analysis.

For the discovery of major folding patterns, several potential future works are listed as follows.

Sulcal Pattern Descriptors. This dissertation adopts sulcal pits as landmarks and builds sulcal graphs for describing sulcal patterns, as the sulcal pits lying in the deep sulcal basins are not sensitive to inter-subject variance in shallower cortical regions. However, the current sulcal graph-based measurements have two limitations in representing sulcal patterns. First, current measurements only capture the sulcal depth of the sulcal pits, and how the sulcal depth changes in each sulcal basin is not taken into account. To address this limitation, one possible solution might be comparing the local sulcal depth maps of the corresponding sulcal basins when computing the similarity between two sulcal graphs. However, one must be very careful of not involving the sulcal depth map in the shallower cortical regions; otherwise, due to the large inter-subject

variation in those shallower cortical regions, the similarity measurement will be too noisy to detect the actual major sulcal patterns. The second limitation is that sulcal graphs cannot effectively represent the sulcal pattern changes along the sulcal banks. To address this limitation, more sophisticated sulcal pattern descriptors are needed.

Major Sulcal Patterns in More Cortical Regions. In this dissertation, major sulcal patterns are explored in three primary cortical regions, including the central sulcus, superior temporal sulcus, and cingulate sulcus. For establishing a complete dictionary of major sulcal patterns, the proposed method should be also applied to all other cortical regions. Moreover, for high-level abstraction, mining the major sulcal patterns of the entire cortical surface is also expected.

Major Gyral Patterns. Sulcus and gyrus are the two important structures of the cerebral cortex. In this dissertation, only the major patterns of local sulci are explored. The major patterns of local gyri are also important but have not been explored yet. However, the proposed sulcal graph-based method is not applicable to the discovery of major gyral patterns, because sulcal pits exist only in deep sulcal basins. Fortunately, the framework proposed in this dissertation for clustering similar patterns can still be used. Therefore, the only problem that is need to solve in mining the major gyral patterns is to design an effective shape descriptor, which can *not only* capture major gyral patterns but also be robust to superficial gyral patterns. One possible solution could be first using spherical wavelet technique to decompose the gyral folding metrics into different levels and then to measure the similarities/differences between gyri only in the major levels.

REFERENCES

- Bernal-Rusiel, J.L., Greve, D.N., Reuter, M., Fischl, B., Sabuncu, M.R., Alzheimer's Disease Neuroimaging, I. (2013) Statistical analysis of longitudinal neuroimage data with Linear Mixed Effects models. *NeuroImage*, 66:249-60.
- Besl, P.J., McKay, N.D. (1992) A Method for Registration of 3-D Shapes. *IEEE transactions on pattern analysis and machine intelligence*, 14:239-256.
- Breiman, L. (2001) Random forests. *Mach Learn*, 45:5-32.
- Brun, L., Auzias, G., Viellard, M., Villeneuve, N., Girard, N., Poinso, F., Da Fonseca, D., Deruelle, C. (2016) Localized Misfolding Within Broca's Area as a Distinctive Feature of Autistic Disorder. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 1:160-168.
- Budday, S., Raybaud, C., Kuhl, E. (2014) A mechanical model predicts morphological abnormalities in the developing human brain. *Scientific reports*, 4:5644.
- Cachia, A., Borst, G., Tissier, C., Fisher, C., Plaze, M., Gay, O., Riviere, D., Gogtay, N., Giedd, J., Mangin, J.F., Houde, O., Raznahan, A. (2016) Longitudinal stability of the folding pattern of the anterior cingulate cortex during development. *Developmental cognitive neuroscience*, 19:122-7.
- Cai, J.F., Candes, E.J., Shen, Z.W. (2010) A Singular Value Thresholding Algorithm for Matrix Completion. *Siam J Optimiz*, 20:1956-1982.
- Candes, E.J., Recht, B. (2009) Exact Matrix Completion via Convex Optimization. *Found Comput Math*, 9:717-772.
- Candes, E.J., Tao, T. (2010) The Power of Convex Relaxation: Near-Optimal Matrix Completion. *Ieee T Inform Theory*, 56:2053-2080.
- Ching, W.K., Li, L., Tsing, N.K., Tai, C.W., Ng, T.W., Wong, A.S., Cheng, K.W. (2010) A weighted local least squares imputation method for missing value estimation in microarray gene expression data. *International journal of data mining and bioinformatics*, 4:331-47.
- Coulon, O., Fonov, V., Mangin, J.F., Collins, D.L. (2012) Atlas-Based Clustering of Sulcal Patterns - Application to the Left Inferior Frontal Sulcus. *2012 9th Ieee International Symposium on Biomedical Imaging (Isbi)*:426-429.
- Criminisi, A., Shotton, J., Konukoglu, E. (2012) Decision Forests: A Unified Framework for Classification, Regression, Density Estimation, Manifold Learning and Semi-Supervised Learning. *Foundations and Trends® in Computer Graphics and Vision*, 7:81-227.
- Dale, A.M., Fischl, B., Sereno, M.I. (1999) Cortical surface-based analysis. I. Segmentation and surface reconstruction. *NeuroImage*, 9:179-94.

- Ducharme, S., Albaugh, M.D., Nguyen, T.V., Hudziak, J.J., Mateos-Perez, J.M., Labbe, A., Evans, A.C., Karama, S., Brain Development Cooperative, G. (2015) Trajectories of cortical surface area and cortical volume maturation in normal brain development. *Data in brief*, 5:929-38.
- Ducharme, S., Albaugh, M.D., Nguyen, T.V., Hudziak, J.J., Mateos-Perez, J.M., Labbe, A., Evans, A.C., Karama, S., Brain Development Cooperative, G. (2016) Trajectories of cortical thickness maturation in normal brain development--The importance of quality control procedures. *NeuroImage*, 125:267-79.
- Evans, A.C., Brain Development Cooperative, G. (2006) The NIH MRI study of normal brain development. *NeuroImage*, 30:184-202.
- Faramarzi, E., Rajan, D., Christensen, M.P. (2013) 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*, 22:2101-14.
- Fischl, B. (2012) FreeSurfer. *NeuroImage*, 62:774-81.
- Frey, B.J., Dueck, D. (2007) Clustering by passing messages between data points. *Science*, 315:972-6.
- Han, X., Pham, D.L., Tosun, D., Rettmann, M.E., Xu, C.Y., Prince, J.L. (2004) CRUISE: Cortical reconstruction using implicit surface evolution. *NeuroImage*, 23:997-1012.
- Han, X., Xu, C.Y., Prince, J.L. (2003) A topology preserving level set method for geometric deformable models. *IEEE transactions on pattern analysis and machine intelligence*, 25:755-768.
- Hazlett, H.C., Gu, H., Munsell, B.C., Kim, S.H., Styner, M., Wolff, J.J., Elison, J.T., Swanson, M.R., Zhu, H., Botteron, K.N., Collins, D.L., Constantino, J.N., Dager, S.R., Estes, A.M., Evans, A.C., Fonov, V.S., Gerig, G., Kostopoulos, P., McKinstry, R.C., Pandey, J., Paterson, S., Pruett, J.R., Schultz, R.T., Shaw, D.W., Zwaigenbaum, L., Piven, J., Network, I., Clinical, S., Data Coordinating, C., Image Processing, C., Statistical, A. (2017) Early brain development in infants at high risk for autism spectrum disorder. *Nature*, 542:348-351.
- Im, K., Choi, Y.Y., Yang, J.J., Lee, K.H., Kim, S.I., Grant, P.E., Lee, J.M. (2011a) The relationship between the presence of sulcal pits and intelligence in human brains. *NeuroImage*, 55:1490-1496.
- Im, K., Jo, H.J., Mangin, J.F., Evans, A.C., Kim, S.I., Lee, J.M. (2010) Spatial distribution of deep sulcal landmarks and hemispherical asymmetry on the cortical surface. *Cerebral cortex*, 20:602-11.
- Im, K., Lee, J.M., Jeon, S., Kim, J.H., Seo, S.W., Na, D.L., Grant, P.E. (2013a) Reliable identification of deep sulcal pits: the effects of scan session, scanner, and surface extraction tool. *PloS one*, 8:e53678.

- Im, K., Pienaar, R., Lee, J.M., Seong, J.K., Choi, Y.Y., Lee, K.H., Grant, P.E. (2011b) Quantitative comparison and analysis of sulcal patterns using sulcal graph matching: a twin study. *NeuroImage*, 57:1077-86.
- Im, K., Pienaar, R., Paldino, M.J., Gaab, N., Galaburda, A.M., Grant, P.E. (2013b) Quantification and discrimination of abnormal sulcal patterns in polymicrogyria. *Cerebral cortex*, 23:3007-15.
- Im, K., Raschle, N.M., Smith, S.A., Ellen Grant, P., Gaab, N. (2016) Atypical Sulcal Pattern in Children with Developmental Dyslexia and At-Risk Kindergarteners. *Cerebral cortex*, 26:1138-1148.
- Kim, S.H., Lyu, I., Fonov, V.S., Vachet, C., Hazlett, H.C., Smith, R.G., Piven, J., Dager, S.R., McKinstry, R.C., Pruett, J.R., Jr., Evans, A.C., Collins, D.L., Botteron, K.N., Schultz, R.T., Gerig, G., Styner, M.A., Network, I. (2016) Development of cortical shape in the human brain from 6 to 24months of age via a novel measure of shape complexity. *NeuroImage*, 135:163-76.
- Koenderink, J.J., Vandoorn, A.J. (1992) Surface Shape and Curvature Scales. *Image Vision Comput*, 10:557-564.
- Le Goualher, G., Procyk, E., Collins, D.L., Venugopal, R., Barillot, C., Evans, A.C. (1999) Automated extraction and variability analysis of sulcal neuroanatomy. *IEEE transactions on medical imaging*, 18:206-17.
- Le Guen, Y., Auzias, G., Leroy, F., Noulhiane, M., Dehaene-Lambertz, G., Duchesnay, E., Mangin, J.F., Coulon, O., Frouin, V. (2017) Genetic Influence on the Sulcal Pits: On the Origin of the First Cortical Folds. *Cerebral cortex*:1-12.
- Leordeanu, M., Hebert, M. (2005) A spectral technique for correspondence problems using pairwise constraints. *Tenth Ieee International Conference on Computer Vision, Vols 1 and 2, Proceedings*:1482-1489.
- Li, G., Lin, W., Gilmore, J.H., Shen, D. (2015a) 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*, 35:9150-62.
- Li, G., Nie, J., Wang, L., Shi, F., Gilmore, J.H., Lin, W., Shen, D. (2014a) Measuring the dynamic longitudinal cortex development in infants by reconstruction of temporally consistent cortical surfaces. *NeuroImage*, 90:266-79.
- Li, G., Nie, J., Wang, L., Shi, F., Lin, W., Gilmore, J.H., Shen, D. (2013) Mapping region-specific longitudinal cortical surface expansion from birth to 2 years of age. *Cerebral cortex*, 23:2724-33.

- Li, G., Nie, J., Wang, L., Shi, F., Lyall, A.E., Lin, W., Gilmore, J.H., Shen, D. (2014b) Mapping longitudinal hemispheric structural asymmetries of the human cerebral cortex from birth to 2 years of age. *Cerebral cortex*, 24:1289-300.
- Li, G., Nie, J., Wu, G., Wang, Y., Shen, D., Alzheimer's Disease Neuroimaging, I. (2012) Consistent reconstruction of cortical surfaces from longitudinal brain MR images. *NeuroImage*, 59:3805-20.
- Li, G., Wang, L., Shi, F., Gilmore, J.H., Lin, W., Shen, D. (2015b) Construction of 4D high-definition cortical surface atlases of infants: Methods and applications. *Medical image analysis*, 25:22-36.
- Li, G., Wang, L., Shi, F., Lin, W., Shen, D. (2014c) Simultaneous and consistent labeling of longitudinal dynamic developing cortical surfaces in infants. *Medical image analysis*, 18:1274-89.
- Li, R., Zhang, W., Suk, H.I., Wang, L., Li, J., Shen, D., Ji, S. (2014d) 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*, 17:305-12.
- Liu, J., Musialski, P., Wonka, P., Ye, J.P. (2013) Tensor Completion for Estimating Missing Values in Visual Data. *IEEE transactions on pattern analysis and machine intelligence*, 35:208-220.
- Lohmann, G., von Cramon, D.Y., Colchester, A.C. (2008) Deep sulcal landmarks provide an organizing framework for human cortical folding. *Cerebral cortex*, 18:1415-20.
- Lohmann, G., von Cramon, D.Y., Steinmetz, H. (1999) Sulcal variability of twins. *Cerebral cortex*, 9:754-63.
- Lorensen, W.E., Cline, H.E. (1987) Marching cubes: A high resolution 3D surface construction algorithm. *SIGGRAPH Comput. Graph.*, 21:163-169.
- Lyall, A.E., Shi, F., Geng, X., Woolson, S., Li, G., Wang, L., Hamer, R.M., Shen, D., Gilmore, J.H. (2015) Dynamic Development of Regional Cortical Thickness and Surface Area in Early Childhood. *Cerebral cortex*, 25:2204-12.
- McKay, D.R., Kochunov, P., Cykowski, M.D., Kent, J.W., Laird, A.R., Lancaster, J.L., Blangero, J., Glahn, D.C., Fox, P.T. (2013) Sulcal Depth-Position Profile Is a Genetically Mediated Neuroscientific Trait: Description and Characterization in the Central Sulcus. *Journal of Neuroscience*, 33:15618-15625.
- Meng, Y., Li, G., Lin, W., Gilmore, J.H., Shen, D. (2014) Spatial distribution and longitudinal development of deep cortical sulcal landmarks in infants. *NeuroImage*, 100:206-18.
- Nie, J., Guo, L., Li, G., Faraco, C., Stephen Miller, L., Liu, T. (2010) A computational model of cerebral cortex folding. *Journal of theoretical biology*, 264:467-78.

- Nie, J., Li, G., Wang, L., Gilmore, J.H., Lin, W., Shen, D. (2012) A computational growth model for measuring dynamic cortical development in the first year of life. *Cerebral cortex*, 22:2272-84.
- Ono, M., Kubik, S., Abernathy, C.D. (1990) Atlas of the cerebral sulci. Stuttgart New York. G. Thieme Verlag; Thieme Medical Publishers. xiv, 218 p. p.
- Piao, X., Hill, R.S., Bodell, A., Chang, B.S., Basel-Vanagaite, L., Straussberg, R., Dobyns, W.B., Qasrawi, B., Winter, R.M., Innes, A.M., Voit, T., Ross, M.E., Michaud, J.L., Descarie, J.C., Barkovich, A.J., Walsh, C.A. (2004) G protein-coupled receptor-dependent development of human frontal cortex. *Science*, 303:2033-6.
- Rakic, P. (1988) Specification of cerebral cortical areas. *Science*, 241:170-6.
- Regis, J., Mangin, J.F., Ochiai, T., Frouin, V., Riviere, D., Cachia, A., Tamura, M., Samson, Y. (2005) "Sulcal root" generic model: a hypothesis to overcome the variability of the human cortex folding patterns. *Neurologia medico-chirurgica*, 45:1-17.
- Rekik, I., Li, G., Lin, W., Shen, D. (2015) Prediction of Longitudinal Development of Infant Cortical Surface Shape Using a 4D Current-Based Learning Framework. *Information processing in medical imaging : proceedings of the ... conference*, 24:576-87.
- Rekik, I., Li, G., Lin, W., Shen, D. (2016) Predicting infant cortical surface development using a 4D varifold-based learning framework and local topography-based shape morphing. *Medical image analysis*, 28:1-12.
- Remer, J., Croteau-Chonka, E., Dean, D.C., 3rd, D'Arpino, S., Dirks, H., Whiley, D., Deoni, S.C.L. (2017) Quantifying cortical development in typically developing toddlers and young children, 1-6 years of age. *NeuroImage*, 153:246-261.
- Schaer, M., Cuadra, M.B., Schmansky, N., Fischl, B., Thiran, J.P., Eliez, S. (2012) How to Measure Cortical Folding from MR Images: a Step-by-Step Tutorial to Compute Local Gyrfication Index. *Jove-J Vis Exp*.
- Schnack, H.G., van Haren, N.E., Brouwer, R.M., Evans, A., Durston, S., Boomsma, D.I., Kahn, R.S., Hulshoff Pol, H.E. (2015) Changes in thickness and surface area of the human cortex and their relationship with intelligence. *Cerebral cortex*, 25:1608-17.
- Shaw, P., Eckstrand, K., Sharp, W., Blumenthal, J., Lerch, J.P., Greenstein, D., Clasen, L., Evans, A., Giedd, J., Rapoport, J.L. (2007) Attention-deficit/hyperactivity disorder is characterized by a delay in cortical maturation. *Proceedings of the National Academy of Sciences of the United States of America*, 104:19649-54.
- Shi, F., Wang, L., Dai, Y., Gilmore, J.H., Lin, W., Shen, D. (2012) LABEL: pediatric brain extraction using learning-based meta-algorithm. *NeuroImage*, 62:1975-86.
- Shi, F., Yap, P.T., Wu, G., Jia, H., Gilmore, J.H., Lin, W., Shen, D. (2011) Infant brain atlases from neonates to 1- and 2-year-olds. *PloS one*, 6:e18746.

- Sled, J.G., Zijdenbos, A.P., Evans, A.C. (1998) A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE transactions on medical imaging*, 17:87-97.
- Smart, I.H., McSherry, G.M. (1986) Gyrus formation in the cerebral cortex in the ferret. I. Description of the external changes. *Journal of anatomy*, 146:141-52.
- Smith, S.M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., Behrens, T.E., Johansen-Berg, H., Bannister, P.R., De Luca, M., Drobnjak, I., Flitney, D.E., Niazy, R.K., Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J.M., Matthews, P.M. (2004) Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage*, 23 Suppl 1:S208-19.
- Sun, T., Hevner, R.F. (2014) Growth and folding of the mammalian cerebral cortex: from molecules to malformations. *Nature reviews. Neuroscience*, 15:217-32.
- Sun, Z.Y., Kloppel, S., Riviere, D., Perrot, M., Frackowiak, R., Siebner, H., Mangin, J.F. (2012) The effect of handedness on the shape of the central sulcus. *NeuroImage*, 60:332-9.
- Sun, Z.Y., Perrot, M., Tucholka, A., Riviere, D., Mangin, J.F. (2009) Constructing a dictionary of human brain folding patterns. *Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention*, 12:117-24.
- Sun, Z.Y., Riviere, D., Poupon, F., Regis, J., Mangin, J.F. (2007) Automatic inference of sulcus patterns using 3D moment invariants. *Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention*, 10:515-22.
- Tallinen, T., Chung, J.Y., Biggins, J.S., Mahadevan, L. (2014) Gyrfication from constrained cortical expansion. *Proceedings of the National Academy of Sciences of the United States of America*, 111:12667-72.
- Tenenbaum, J.B., de Silva, V., Langford, J.C. (2000) A global geometric framework for nonlinear dimensionality reduction. *Science*, 290:2319-+.
- Thung, K.H., Wee, C.Y., Yap, P.T., Shen, D., Alzheimer's Disease Neuroimaging, I. (2014) Neurodegenerative disease diagnosis using incomplete multi-modality data via matrix shrinkage and completion. *NeuroImage*, 91:386-400.
- Tibshirani, R. (1996) Regression shrinkage and selection via the Lasso. *J Roy Stat Soc B Met*, 58:267-288.
- Troyanskaya, O., Cantor, M., Sherlock, G., Brown, P., Hastie, T., Tibshirani, R., Botstein, D., Altman, R.B. (2001) Missing value estimation methods for DNA microarrays. *Bioinformatics*, 17:520-5.

- Tsiporkova, E., Boeva, V. (2007) Two-pass imputation algorithm for missing value estimation in gene expression time series. *Journal of bioinformatics and computational biology*, 5:1005-22.
- von Luxburg, U. (2007) A tutorial on spectral clustering. *Stat Comput*, 17:395-416.
- Wang, B., Mezlini, A.M., Demir, F., Fiume, M., Tu, Z., Brudno, M., Haibe-Kains, B., Goldenberg, A. (2014a) Similarity network fusion for aggregating data types on a genomic scale. *Nature methods*, 11:333-7.
- Wang, L., Shi, F., Li, G., Gao, Y., Lin, W., Gilmore, J.H., Shen, D. (2014b) Segmentation of neonatal brain MR images using patch-driven level sets. *NeuroImage*, 84:141-58.
- Wang, L., Shi, F., Yap, P.T., Lin, W., Gilmore, J.H., Shen, D. (2013) Longitudinally guided level sets for consistent tissue segmentation of neonates. *Human brain mapping*, 34:956-72.
- Wierenga, L.M., Langen, M., Oranje, B., Durston, S. (2014) Unique developmental trajectories of cortical thickness and surface area. *NeuroImage*, 87:120-6.
- Yeo, B.T., Sabuncu, M.R., Vercauteren, T., Ayache, N., Fischl, B., Golland, P. (2010) Spherical demons: fast diffeomorphic landmark-free surface registration. *IEEE transactions on medical imaging*, 29:650-68.
- Yuan, L., Wang, Y., Thompson, P.M., Narayan, V.A., Ye, J., Alzheimer's Disease Neuroimaging, I. (2012) Multi-source feature learning for joint analysis of incomplete multiple heterogeneous neuroimaging data. *NeuroImage*, 61:622-32.