Spatial Regression Analysis of Diffusion Tensor Imaging for Longitudinal Studies

by

Bilan Liu

Submitted in Partial Fulfillment of the
Requirements for the Degree
Doctor of Philosophy

Supervised by
Professor Jianhui Zhong

Department of Electrical and Computer Engineering
Arts, Sciences and Engineering
Edmund A. Hajim School of Engineering and Applied Sciences

University of Rochester
Rochester, New York

2016
This dissertation is gratefully dedicated to my beloved parents

Changfu Liu and Gui’e Zhu
Biographical Sketch

The author was born in Wuhan, Hubei Province, P.R. China. She attended Harbin Institute of Technology (Harbin, P.R. China) from 2002 to 2006, and graduated with a Bachelor of Engineering degree in Electronic and Information Engineering. She attended Harbin Institute of Technology Shenzhen Graduate School (Shenzhen, P.R. China) from 2006 to 2008, and received a Master of Science degree in Communication Engineering. She came to University of Rochester in the fall of 2008, and began graduate studies in the Department of Electrical and Computer Engineering. She received her Master of Science degree in Electrical and Computer Engineering in 2010. She started her doctoral studies in Electrical and Computer Engineering at the University of Rochester in 2010, and has received a research assistantship since then. She pursued her research in the areas of diffusion tensor imaging and nonparametric statistical algorithm design for lesion detection in whole-brain analysis under the supervision of Dr. Jianhui Zhong since 2010.

The following publications were a result of work conducted during doctoral study:


Acknowledgements

First and foremost, I would like to express my special appreciation and thanks to my advisor, Prof. Jianhui Zhong, who has been very supportive, both academically and emotionally, through the rough road to finish this thesis. It is thanks to him that I had the opportunity to continue my PhD study in the area of Magnetic Resonance Imaging at the University of Rochester. He helped me come up with the thesis topic and guided me through the whole process with encouragement. He also gave me moral support, as well as the freedom I needed to move on during the most difficult time when writing this thesis.

I would like to give my heartfelt, special thanks to Prof. Xing Qiu and Dr. Tong Zhu. We worked together on several projects during my PhD study. They have been more like friends to me than mentors. They taught me the way of scientific thinking and gave me good advice and collaboration. Their enthusiasm for research was contagious and motivational for me. I very much appreciated their patience, genuine caring and concern, and faith in me during the dissertation process. They remained supportive and provided insight and direction all the way through my PhD study. I cannot thank them enough.

I am really grateful to my defense committee members, Prof. Mark Bocko, Prof. Marvin Doyley, and Prof. Hongyu Miao. Their insightful advice, comments and personal cheering are greatly appreciated. I would also like to thank Prof. Chen Ding for being my thesis committee chair.

My thesis could not have been finished without the help of many co-authors: Dr. Sven Ekholm, Dr. Giovanni Schifitto, Dr. Wei Tian, and Dr. Rui Hu. I would like to thank
them for their academic support and input. My gratitude is also extended to my mentor and dear friend Denise Gettman, who helped me immensely with my communication and presentation skills; graduate program coordinator Michele Foster, who is always helpful regarding all the paperwork necessary for graduate study, PhD dissertation and oral defense; and many friends at the Rochester Center for Brain Imaging (RCBI): Pat Web, who provided MRI scanner operation and safety training; Judy Ripton, who took care of the daily administrative activities; and many more. This list is undoubtedly long. I consider myself extremely lucky to have all these supportive people surround me. Thank you all for your help.

My time at the University of Rochester was made special and enjoyable largely due to the many friends I have here, who became an important part of my life. I would like to thank them for all the encouragement and support they gave me, they kept things light and me smiling. My time in Rochester was also enriched by my dear friends from the Ridgeland Community Church. I am grateful for the time spent with roommates and many friends here, which became invaluable memories that I will cherish.

Last but not least, I would like to thank my parents, for all the sacrifices that they’ve made on my behalf. They raised me with a love of science and supported me in all my pursuits. Their unfailing love and encouragement sustained me through all the years that I have been studying far away from home.
Abstract

Diffusion Tensor Imaging (DTI) enables accurate description of the degree and direction of water dynamics in biological tissues, which provides detailed information about white matter microstructure, and has been widely used in the field of neuroscience and medicine. However, DTI is susceptible to numerous detrimental artifacts that may impair the reliability and validity of the obtained data. Image quality of DTI is therefore critical. In Chapter 3, the effectiveness of three popular Quality Control (QC) tools including DTI studio, DTIprep and TORTOISE are quantitatively compared. Both synthetic and in vivo human brain data were used to quantify adverse effects of major DTI artifacts to tensor calculation as well as the effectiveness of different QC tools in identifying and correcting these artifacts. The technical basis of each tool was studied; the different functions and I/O formats that three QC tools provide were also discussed.

Subject-specific longitudinal DTI study is vital in the investigation of pathological changes of lesions, disease evolution and treatment evaluation, which remains an important area of current research. Spatial REgression Analysis of Diffusion tensor imaging (SPREAD) is a nonparametric statistical method that has been developed in the Zhong lab, which combines spatial regression and permutation techniques to achieve effective detection of localized longitudinal changes within the whole brain at individual-level without a priori hypotheses. In Chapter 4, we propose an improved SPREAD (iSPREAD) method. A three-dimensional nonlinear anisotropic diffusion filtering method is incorporated in iSPREAD to eliminate the potential shortcomings caused by the Gaussian kernel used in the SPREAD method. Results from both simulated and in
vivo human brain data demonstrated that iSPREAD identifies subject-specific longitudinal changes in the brain with substantially improved sensitivity, accuracy, and enhanced statistical power.

As an extension of iSPREAD, we present a general statistical method that facilitates analysis of serial DTI studies for testing regionally specific changes in Chapter 5. Two types of voxel-level test statistics were estimated and used to test against the null hypothesis among groups of DTI data across time. The proposed statistical framework is shown to be accurate and can be applied to a broad spectrum of longitudinal studies to help detect localized changes at individual-level with carefully designed test statistics.

Summary and Future Work are presented in Chapter 6.
Contributors and Funding Sources

This work was supported by a dissertation committee consisting of Prof. Jianhui Zhong (advisor), Prof. Mark Bocko of the Department of Electrical and Computer Engineering, Prof. Marvin Doyley of the Department of Electrical and Computer Engineering, and Prof. Hongyu Miao of the Department of Biostatistics. The following chapters of this dissertation proposal were jointly produced, and were funded by multiple sources. My participation and contributions to the research as well as funding sources are as follows.

I am the primary author of all the chapters. For Chapter 3, I collaborated with Dr. Tong Zhu and Dr. Jianhui Zhong. Dr. Tong Zhu and Dr. Jianhui Zhong gave me many useful advices during the process. This Chapter is based on the manuscript entitled “Comparison of quality control software tools for diffusion tensor imaging” by Bilan Liu, Tong Zhu and Jianhui Zhong, that appeared in Magnetic Resonance Imaging (2014), Copyright ©2014, Elsevier Inc. Reprinted with permission of Elsevier, Inc.

For Chapter 4, I collaborated with Dr. Tong Zhu, Dr. Wei Tian, Dr. Rui Hu, Dr. Sven Ekholm, Dr. Giovanni Schifitto, Dr. Xing Qiu, and Dr. Jianhui Zhong. This chapter is based on the manuscript entitled “Improved spatial regression analysis of diffusion tensor imaging for lesion detection during longitudinal progression of neurodegenerative disease in individual subjects”. The paper was appeared in Physics in Medicine and Biology (2016). Copyright © 2016 Institute of Physics and Engineering in Medicine. Reprinted with permission of Institute of Physics and Engineering in Medicine. This work was supported by NIH grants RO1 MH099921, NYSTEM Consortia.
For Chapter 5, I collaborated with Dr. Tong Zhu, Dr. Wei Tian, Dr. Rui Hu, Dr. Sven Ekholm, Dr. Giovanni Schifitto, Dr. Xing Qiu, and Dr. Jianhui Zhong. This chapter is based on the manuscript entitled “Spatial regression analysis of serial DTI for subject-specific longitudinal changes of neurodegenerative disease”. The paper was appeared in Neuroimage: Clinical (2016). Copyright ©2016, Elsevier Inc. Reprinted with permission of Elsevier, Inc.. This work was supported by NIH grants RO1 MH099921, NIH/NIDA 1 R01 MH097587, NYSTEM Consortia.
# Table of Contents

Biographical Sketch ........................................................................................................ iii

Acknowledgements ........................................................................................................ v

Abstract ........................................................................................................................ vii

Contributors and Funding Sources ............................................................................. ix

List of Tables .................................................................................................................. xvi

List of Figures ................................................................................................................ xvii

Chapter 1 Introduction ..................................................................................................... 1

1.1 Introduction and motivation .................................................................................... 1

1.2 Dissertation organization ......................................................................................... 4

Chapter 2 Introduction to MRI ....................................................................................... 5

2.1 Principles of MRI ..................................................................................................... 6

2.1.1 Physics of MR .................................................................................................... 6

2.2 MR data acquisition and image reconstruction .................................................... 10

2.3 Diffusion tensor imaging (DTI) ............................................................................. 12

2.3.1 Basics of diffusion ........................................................................................... 13

2.3.2 Diffusion tensor imaging mathematics ............................................................. 13

2.3.2.1 Tensor and diffusion ellipsoid .................................................................. 13

2.3.2.2 Tensor-derived parameters: quantification and visualization ................. 17
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.3.3 Fiber tractography</td>
<td>19</td>
</tr>
<tr>
<td>2.3.4 DTI artifacts</td>
<td>20</td>
</tr>
<tr>
<td>2.3.5 Main applications and current challenges of DTI</td>
<td>22</td>
</tr>
<tr>
<td>Chapter 3 Comparison of quality control software tools for DTI</td>
<td>24</td>
</tr>
<tr>
<td>3.1 Introduction</td>
<td>24</td>
</tr>
<tr>
<td>3.2 Material and methods</td>
<td>25</td>
</tr>
<tr>
<td>3.2.1 Introduction to three quality control tools</td>
<td>25</td>
</tr>
<tr>
<td>3.2.2 Two major DTI artifacts</td>
<td>27</td>
</tr>
<tr>
<td>3.2.2.1 Within-volume bulk motion artifacts</td>
<td>28</td>
</tr>
<tr>
<td>3.2.2.2 Localized cardiac artifacts</td>
<td>28</td>
</tr>
<tr>
<td>3.2.3 Image data for quantitative comparisons of three QC tools</td>
<td>29</td>
</tr>
<tr>
<td>3.2.3.1 Human DTI data</td>
<td>29</td>
</tr>
<tr>
<td>3.2.3.2 Monte Carlo simulation</td>
<td>29</td>
</tr>
<tr>
<td>3.2.4 Statistical analysis</td>
<td>32</td>
</tr>
<tr>
<td>3.2.4.1 Monte Carlo simulation data</td>
<td>32</td>
</tr>
<tr>
<td>3.2.4.2 In vivo human brain data</td>
<td>33</td>
</tr>
<tr>
<td>3.3 Results</td>
<td>34</td>
</tr>
<tr>
<td>3.3.1 Comparison of QC tool results for simulation data</td>
<td>36</td>
</tr>
<tr>
<td>3.3.1.1 Effects of SNR and number of directions</td>
<td>36</td>
</tr>
</tbody>
</table>
3.3.2 Comparison of QC tool results in vivo human brain data .................................. 43

3.4 Discussion ............................................................................................................. 46

3.5 Conclusion .......................................................................................................... 51

Chapter 4 Improved spatial regression analysis of diffusion tensor imaging for lesion detection during longitudinal progression of multiple sclerosis in individual subjects .... 53

4.1 Introduction .......................................................................................................... 53

4.2 Methods .............................................................................................................. 56

4.2.1 Overview of iSPREAD .................................................................................. 56

4.2.1.1 Justification of permutation invariance .................................................. 58

4.2.2 Overview of anisotropic diffusion (AD) filter .............................................. 59

4.2.3 Monte Carlo simulations .............................................................................. 62

4.2.4 In vivo human brain Data ............................................................................. 63

4.2.4.1 Subjects and image acquisition .............................................................. 63

4.2.4.2 Steps of analysis ..................................................................................... 65

4.3 Results .................................................................................................................. 67

4.3.1 Monte Carlo simulations .............................................................................. 67

4.3.2 Healthy volunteers ........................................................................................ 70

4.3.3 MS patients .................................................................................................... 71

4.3.3.1 MS patient 1 ........................................................................................... 73
Chapter 5 Spatial regression analysis of serial DTI for subject-specific longitudinal changes of neurodegenerative disease ................................................................. 83

5.1 Introduction ........................................................................................................... 83

5.2 Material and methods ............................................................................................ 88

5.2.1 Overview of iSPREAD for serial DTI analysis ................................................. 88

5.2.2 Statistical model for test statistics ...................................................................... 89

5.2.2.1 Model-free voxel-wise test statistic .............................................................. 89

5.2.2.2 Voxel-wise statistics based on general linear model ..................................... 90

5.2.3 Monte Carlo simulation ...................................................................................... 94

5.2.4 Human in vivo brain data .................................................................................. 97

5.2.4.1 Subjects and image acquisition .................................................................... 97

5.2.4.2 Analysis ....................................................................................................... 98

5.3 Results .................................................................................................................. 99

5.3.1 Monte Carlo simulation ..................................................................................... 99

5.3.2 Human in vivo data ......................................................................................... 101
List of Tables

Table 2.1 T1, T2 relaxation times of some normal tissue types at 3T and 1.5T (measured at 37°) (33) ................................................................. 9

Table 3.1 Summary of main characteristics of the three Quality Control packages .......... 25

Table 3.2 Overall performances of different QC tools .................................................. 34

Table 3.3 A brief summary of featured algorithms used by different software packages ... 46

Table 4.1 Patient clinical characteristics and diagnoses ................................................. 71

Table 4.2 FA results for iSPREAD vs. SPREAD .......................................................... 72

Table 5.1 TPR_L and FPR_NLWM calculated using each test statistic for MS patient 1 .... 102
Table 5.2 TPR_L and FPR_NLWM calculated using each test statistic for MS patient 2 .... 104
Table 5.3 TPR_L and FPR_NLWM calculated using each test statistic on the first slice .... 106
Table 5.4 TPR_L and FPR_NLWM calculated using each test statistic on the second slice .... 107
List of Figures

Figure 2.1 (a) Nuclear Spin. (b) Nucleus behaves like a tiny magnet with a north and south pole. ................................................................. 7

Figure 2.2 (a) Spins randomly oriented in the absence of static magnetic field. (b) Spins oriented in the direction of the main static magnetic field. .............................. 7

Figure 2.3 Rotating frame behavior. Illustration of the rotating spins a) longitudinal magnetization in equilibrium state. b) 90 degree RF pulse flips the magnetization to the transverse plane. c) Dephasing in the transverse plane and relaxation of longitudinal magnetization begins. d) Decay of the transverse magnetization and the increase of longitudinal magnetization which will eventually recover back to equilibrium state. ......................... 8

Figure 2.4 Illustration of diffusion ellipsoid .................................................................................. 14

Figure 2.5 Diffusion ellipsoids and tensors for isotropic unrestricted diffusion, isotropic restricted diffusion, and anisotropic restricted ........................................... 16

Figure 2.6. Example maps of MD (A), FA (B), and color FA (C) ................................................. 18

Figure 2.7 Schematic diagram of the linear propagation approach, which was dubbed FACT (fiber assignment by continuous tracking). Arrows indicated fiber orientations at each pixel. Red lines are FACT trajectories. .............................. 20

Figure 2.8 Human brain data affected by motion artifact and cardiac artifact. (A) Typical DWI affected by motion artifact, artifact-free data (left) vs. data corrupted by motion artifact (right). (B) Typical DWI affected by cardiac artifact, artifact-free data (left) vs. data corrupted by cardiac artifact (right) ......................... 22
Figure 3.1 Biases and SDs of measured FA/Trace for datasets corrupted with cardiac artifact as a function of SNRs with different gradient schemes (SNR=25, 50, 125, Number of directions=12, 30, 60) using Eqs. (3.1) – (3.2) (unit for Bias and SD of Trace is mm$^2$/s). Corrupted data points had an intensity value set to 50% of the original signal intensity. FA/Trace values were measured in WM. Simulated $FA_{GS} = 0.7$, which is representative of the typical FA value of brain WM, and $Trace_{GS} = 2100 \times 10^{-6}$mm$^2$/s, which is representative of trace of brain parenchyma. DW images were all with $b_{value} = 1000s/mm^2$. Results from TORTOISE, DTIprep and DTIstudio were shown as blue dots, green triangles and red squares, respectively.................

Figure 3.2 Biases of measured FA/Trace ($Bias_{FA}$, $Bias_{Trace}$), SDs of FA/Trace ($\sigma_{FA}$, $\sigma_{Trace}$) in WM (Fig. 3.2 A) and GM (Fig. 3.2 B) of simulation data as a function of corrupted ratio (SNR=50, Number of directions=12, 30, 60) using Eq. (3.1) – (3.2). Corrupted gradients were either evenly (blue squares) or unevenly (green triangles) distributed over the unit sphere (unit for Bias and SD of Trace is mm$^2$/s). The ratio of corrupted gradients in each dataset was set to 8%, 17% and 33%, with the outliers’ intensity value set to 10% of the original signal intensity (Signal Drop=90%). Simulated $FA_{GS} = 0.7$, $Trace_{GS} = 2100 \times 10^{-6}$mm$^2$/s in WM and $FA_{GS} = 0.2$, $Trace_{GS} = 2100 \times 10^{-6}$mm$^2$/s in GM..........................

Figure 3.3 Human brain data affected by motion artifact and cardiac artifact. (A) Typical DWI affected by motion artifact, artifact-free data (left) vs. data corrupted by
motion artifact (right). (B) Typical DWI affected by cardiac artifact, artifact-free data (left) vs. data corrupted by cardiac artifact (middle); binary mask output by DTIstudio showing the location of the artifact (right). (C) Region of interests (ROIs) used for human brain data analysis. Three WM ROIs were: the body of the corpus callosum (BCC; left); the genu of the corpus callosum (GCC) and the splenium of the corpus callosum (SCC; middle). The fourth ROI was in the middle cerebellar peduncles (MCP; right). ROIs were superimposed on the sampled FA data calculated from an artifact-free dataset, using DTIstudio as an example.......................... 44

Figure 3.4 Measured accuracy and precision in human brain data based on 6 selected ROIs. Unit for $\sigma$ (Trace) is mm$^2$/s. 6 ROIs were: BCC from subject 1; BCC from subject 2; GCC and SCC from subject 2, the last two were in the cortex of the cerebellum for subject 1 and 3. Accuracy and Precision level of FA/Trace were calculated using Eq. (3.5) - (3.6). ........................................ 44

Figure 4.1 Flow diagram of iSPREAD ........................................................................................................... 56

Figure 4.2. ROC analyses for the two-group comparisons using iSPREAD (red crosses) and SPREAD (blue dots) from Monte Carlo simulated data under different combinations of effect size ($es = 0.2, 0.4$) and sample size ($n = 3, 5$) of each group. The disease area was simulated as either a $5 \times 5 \times 3$ cubic region (Fig. 4.2 (a)) or a $3 \times 10 \times 3$ cuboid region (Fig. 4.2 (b)). Sensitivity and Specificity were calculated using Eq. (4.9). Results for FA analysis are shown here as an example. The nonlinear anisotropic diffusion filtering was
used for data smoothing in the iSPREAD method and a Gaussian kernel of fixed FWHM= $2 \times 2 \times 2$ voxels were applied for estimation of spatial regression for SPREAD.

Figure 4.3 Location of false positive voxels detected (with zoom in images on the right) using iSPREAD analysis in longitudinal DTI data of one healthy volunteer. False positive of FA (red dots) and MD (blue dots) for iSPREAD are mostly occurred due to image mis-registration error or BET residuals, either at tissue boundaries or brain boundaries. Figures are displayed according to radiological convention.

Figure 4.4 Comparison of significant voxels detected for FA (marked by magenta dots) and MD (marked by cyan dots) using SPREAD (Row E and G) and iSPREAD (Row D and F) analysis of longitudinal data in monitoring disease progression for patient 1. This patient had an active lesion around the atrium of the left lateral ventricle visible on the post contrast T1 images at 6-month (Row A) after baseline, which shows as hyperintensity on FLAIR image (Row B). The images are magnified in the disease area showing the detected significant voxels on the right column.

Figure 4.5 Comparison of significant voxels detected for FA (marked by magenta dots) and MD (marked by cyan dots) using iSPREAD (Row B & C. second column) and SPREAD (Row B & C. third volume) analysis of longitudinal data during disease progression for patient 2. This patient had a lesion in the periventricular white matter near the temporal parietal area, which shows
enhancement on the post contrast T1 image (Row A) 24 months after the baseline. ................................................................. 75

Figure 4.6 Comparison of significant voxels detected for FA (marked by magenta dots) and MD (marked by cyan dots) using iSPREAD (Row B & C. first column) and SPREAD (Row B & C. second column) analysis of longitudinal data in monitoring disease progression for patient 3. This patient had a lesion shown as a ring-enhancing lesion on T1-weighted post contrast image at baseline (Row A). ........................................................................................................ 77

Figure 5.1 Flow diagram of iSPREAD for serial DTI analysis ................................. 88

Figure 5.2 Illustration of linear and quadratic simulated models (Effect size on $\lambda_1$ vs. Timepoints)........................................................................................................... 96

Figure 5.3 ROC analysis for group comparisons using iSPREAD with five proposed test statistics to identify localized changes from Monte Carlo simulated data. Disease progression models were simulated by either (a, b) a simple linear model ($es_1 = 0$, $es_2 = 0.03$, $es_3 = 0.05$, $es_4 = 0.08$, $es_5 = 0.15$ ) or (c, d) a quadratic model ($es_1 = 0$, $es_2 = 0.5$, $es_3 = 0.2$, $es_4 = 0.1$, $es_5 = 0.05$). Sample sizes of $n = 3$ and $n = 5$ were considered. The disease area was simulated as a $5 \times 5 \times 3$ cubic region at the center of splenium corpus callosum. Sensitivity and Specificity were calculated using Eq.(5.13). Results for FA analysis are shown here as an example........................................... 99

Figure 5.4 This patient had an active lesion (red arrows) in the posterior limb of the right internal capsule visible on the post contrast T1 images at baseline and
resolved in the follow-up scans. The degree of hyperintensity within the lesion lessened after 3 month and 4 month compared to the baseline scan. 101

Figure 5.5 Comparison of significant voxels detected for FA (magenta dots) using iSPREAD with five proposed test statistics (Fig.5.5 (a)). Mean FA values calculated from gold standard ROIs at different time points were drawn in Fig.5.5 (b), together with a fitted simple linear model and a quadratic regression model for mean FAs over time.................................................. 102

Figure 5.6 This patient had an active lesion around the atrium of the left lateral ventricle visible on the post contrast T1 images at 6 month (red arrows) from baseline and resolved late. ................................................................. 104

Figure 5.7 Significant voxels detected for FA (marked by magenta dots) using iSPREAD with five proposed test statistics. TPRL and FPRNLWM calculated using each test statistic by Eq. (5.14) are listed in Table 5.2................................. 104

Figure 5.8 This patient had two new gadolinium enhancing lesions visible in this slice on the post contrast T1 images at baseline and resolved in the follow-up 3-month scan and 6-month scan. One subtle peripheral enhancing lesion (blue arrows) was within the medial aspect of the occipital horn of left lateral ventricle in the left posterior temporal area. Another one lesion (red arrows) was located in the right subcortical temporal area (near the occipital horn of right lateral ventricle). Both lesions demonstrate hyperintensity on FLAIR image at the baseline, with both the size of the lesions and degree of FLAIR
hyperintensity gradually lessened in the follow-up scans, which suggested a stable state.

Figure 5.9 Significant voxels detected for FA (marked by magenta dots) using iSPREAD with three proposed test statistics. TPRL and FPRNLWM calculated using each test statistic by Eq. (5.14) are listed in Table 3.

Figure 5.10 This patient had one new gadolinium enhancing lesions visible in this slice on the post contrast T1 images at baseline (first column, marked by red arrow) and resolved in the follow-up scans. The enhancing lesion was located in the right splenium of the corpus callosum extending to the right major forceps area and shown as hyperintensity on the FLAIR image at baseline, and gradually lessened both in size and intensity in the follow-up scans.

Figure 5.11 Significant voxels detected for FA (marked by magenta dots) using iSPREAD with three proposed test statistics. TPRL and FPRNLWM calculated using each test statistic by Eq. (5.14) are listed in Table 5.4.
Chapter 1 Introduction

1.1 Introduction and motivation

Diffusion Tensor Imaging (DTI) (1), a magnetic resonance imaging modality that enables mapping of water molecular random translational motion, has become one of the most important non-invasive imaging techniques for probing in vivo tissue structures at microscopic scale (2,3). It has been widely used in the fields of neuroscience and medicine for investigating brain development and aging (4), detecting abnormalities in normal-appearing white matter (WM) due to disease (5), identifying subtle changes in normal-appearing WM, as well as monitoring pathologic severity of lesions and disease evolution (6).

In spite of its great utilities, DTI is subject to various artifacts and pitfalls, similar to many other MRI techniques but further accentuated by its inherent, low signal-to-noise ratio and high sensitivity to motions of different scales (7). Among those artifacts, eddy current-induced artifacts caused by rapidly switched strong field gradient pulses and motion artifacts caused by patient motion are two major sources of artifacts (7). Those artifacts, together with other artifacts caused by susceptibility variations and vibration due to gradient switching, strongly deteriorate the quality of DTI and subsequently affect their diagnostic accuracy. Consequently, quality control (QC) remains a major and recurring topic in the field of DTI. Due to its importance, the QC problem was first discussed and three popular QC tools were compared quantitatively and qualitatively in Chapter 3.

Group-level analysis, which includes both region-of-interest (ROI)-based (8) and voxel-based (9,10) statistical tests, are often used for statistical comparisons of longitudinal study of DTI-
derived parameters, such as mean diffusivity (MD) and fractional anisotropy (FA). However, group-level analysis is most suitable when pathological changes are located in similar anatomical regions among subjects. It often fails when the longitudinal changes/effects are diverse and highly specific for individual subjects, where the inter-subject variations exceed the between-group variations. In recent years, there has been increasing interest in investigation of the subject-specific changes within the brain without prior information regarding the spatial distribution of the pathology. Consequently, voxel-based methods (11-13) have gained much favor during recent years as an important alternative to ROI analysis in detecting localized changes within the brain and most suitable when changes/effects are diverse and highly specific for individual subjects (14,15).

Both parametric (16) and non-parametric statistical analysis methods (17,18) have been used to help identify the regionally specific changes such as differences due to activation in fMRI (19), neuroanatomical difference in structure MRI data (20) and pathophysiology in longitudinal studies (21,22). Since the non-normally distributed residuals of DTI parameters render valid statistical inferences with a parametric approach problematic (23), nonparametric resampling methods such as bootstrap and permutation have gained much favor in recent years. Compared to the parametric statistical methods (e.g. Statistical Parametric Mapping (16)), nonparametric methods, which include both descriptive and inferential statistics, have the advantage of achieving sufficient statistical power with minimal assumptions about the data being investigated. Bootstrap (17,24-28) has been used to quantify uncertainties of DTI-derived parameters through both simulation and real human DTI data. However, it is a descriptive statistical method and the p-value it generates is only an approximation. The permutation test, first introduced to image analysis by Holmes et al (18), is able to devise a data-driven null distribution of data with only
minimum assumptions and produce exact or almost exact \( p \)-values (19). This method gives researchers more freedom in terms of choosing suitable summary statistics and has been widely used in the field of functional MRI (fMRI) for detection of active brain regions. Any sensible test statistic that summarizes the local effect can be used, regardless of a known theoretical distribution under null hypothesis.

Quantitative measurement of localized longitudinal changes in brain abnormalities at individual level is an important area of current research. In recent years, several permutation-based methods such as PERVADE (15) and SPREAD (14) have been developed for voxel-wise whole brain longitudinal studies of local DTI changes or lesion evolution. SPatial REgression Analysis of Diffusion tensor imaging (SPREAD) (14) is a non-parametric permutation-based statistical framework that combines spatial regression and resampling techniques to achieve effective detection of localized longitudinal diffusion changes within the whole brain at individual-level without a priori hypotheses. However, the Gaussian kernel it used has intrinsic limitations such as boundary blurring and dislocation, which limits its sensitivity, especially towards detecting lesions of irregular shapes.

The main goal of this thesis in Chapter 4 and Chapter 5 is to develop an improved spatial regression analysis method (iSPREAD) using anisotropic filtering (Chapter 4) for lesion detection during longitudinal progression of neurodegenerative disease in individual subjects, and the proposed method iSPREAD was further extended to a general statistical framework that facilitates analysis of serial DTI studies (with more than two time points) for testing regionally specific changes (Chapter 5).
1.2 Dissertation organization

The organization of this dissertation proposal is as follows. In Chapter 3, the effectiveness of three popular QC tools including DTI studio (Johns Hopkins University), DTIprep (University of North Carolina at Chapel Hill, University of Iowa and University of Utah) and TORTOISE (National Institute of Health) are quantitatively and qualitatively compared. In Chapter 4, we present an improved spatial regression analysis of diffusion tensor imaging for lesion detection during longitudinal progression of neurodegenerative disease in individual subjects (iSPREAD). In Chapter 5, we further extend the method of iSPREAD to a general statistical framework that facilitates analysis of serial Diffusion Tensor Imaging (DTI) studies (with more than two time points) for testing regionally specific changes in longitudinal studies. This statistical framework can be applied to a broad spectrum of longitudinal studies as well as investigate localized effects within the whole brain (e.g. experimental/drug effect) at individual level.
Chapter 2 Introduction to MRI

Magnetic Resonance Imaging (MRI) is an important noninvasive imaging modality that has become an essential diagnostic imaging tool and has gained wide popularity in clinics. It uses imaging of the proton, and the most frequently used nucleus is hydrogen, due to its abundance in tissues containing water. The distribution of the hydrogen nuclei and parameters relating to their motion in water and lipids can be investigated and displayed using MRI. The different tissue characteristics $T_1$, $T_2$ and proton density provide variation between tissues, which permits acquisition of images with exquisite contrast between soft tissues, as well as normal and abnormal morphology and pathology. Therefore, MRI is able to probe the internal structure of the body without any potential hazard, and gives superior soft-tissue contrast.

The phenomenon of MR was first discovered by Felix Bloch and Edward Purcell (29); they were awarded jointly for the Nobel Prize in Physics in 1952 for their “development of new methods for nuclear magnetic precision measurements and discoveries in connection therewith”, which is the scientific principle behind MRI. In 1973, Lauterbur made the development of MRI possible by introducing a linearly varying magnetic field for spatial localization. The first MR scan of the human body was achieved in 1977 by Damadian and coworkers, Minkoff and Goldsmith (30). Around 1980, the first commercial MRI scanner was introduced. Since then, MRI has been widely used in a variety of clinical applications, including clinical neurology, cardiology, cancer, as well as help study brain functions and cancer growth (31). Moreover, various technologies have been developed and improvements have been made towards shorter scan time and finer image resolution.
2.1 Principles of MRI

2.1.1 Physics of MR

Nuclei with an odd number of protons or neutrons possess a net charge and have angular momentum, which behaves as magnetic dipoles. MR describes the phenomenon of the nuclei of certain atoms, when placed in a magnetic field, interacting with the following three types of magnetic fields: 1) main static field $B_0$, 2) radiofrequency field $B_1$, and 3) linear gradient fields $G$.

The most widely used nuclei in MRI is the hydrogen nucleus (or proton), due to its abundance in virtually all biological tissues. These nuclei absorb and emit RF signal of a specific frequency to generate the MR signal.

The proton can be visualized as a freely suspended bar magnet spinning about its magnetic axis. When a group of protons are placed in the main static magnetic field $B_0$, they tend to align in the direction of the applied $B_0$. This produces a net magnetization in the $B_0$ direction, which defines the z axis or longitudinal direction. Moreover, the nuclei respond to the magnetic couple like a gyroscope and rotate about the magnetic field direction at the same frequency. This movement is known as precession and the precession frequency is known as the Larmor frequency defined as follows:

$$\omega = \gamma B,$$

or

$$f = \frac{\gamma}{2\pi} B,$$

where $f$ is the resonant frequency, $\gamma$ is the gyromagnetic ration, and $B$ is the applied magnetic field. For $^1H$, $\frac{\gamma}{2\pi} = 42.58MHz/Tesla$. Most magnetic field strengths used for whole-body imaging ranges from 0.1T to 3T. Very high field-strength ranges from 7~14T have been used for animal MRI system and human studies in research setting, which results in higher SNR (32).
Figure 2.1 (a) Nuclear Spin. (b) Nucleus behaves like a tiny magnet with a north and south pole.

Figure 2.2 (a) Spins randomly oriented in the absence of static magnetic field. (b) Spins oriented in the direction of the main static magnetic field.
Figure 2.3 Rotating frame behavior. Illustration of the rotating spins a) longitudinal magnetization in equilibrium state. b) 90 degree RF pulse flips the magnetization to the transverse plane. c) Dephasing in the transverse plane and relaxation of longitudinal magnetization begins. d) Decay of the transverse magnetization and the increase of longitudinal magnetization which will eventually recover back to equilibrium state.

A pulse of oscillating Radiofrequency (RF) tuned to the Larmor frequency of the spin is applied in the $xy$ (transverse) plane to excite the spins out of equilibrium, which induce a strong interaction or resonant effect. The energy of RF pulse induces nutation of magnetization vector as it tips away from $z$-axis. The net magnetization along the $z$ axis is deviated through an angle which depends upon the strength and duration of the pulse of the RF magnetic field. The so-
called 90° and 180° pulses are used to rotate the magnetization in the z direction through 90°
and 180°, respectively. After the excitation, the magnetization will eventually return to its
equilibrium state along z. Two relaxation time constants, $T_1$ and $T_2$, characterize the return to
equilibrium. $T_1$ (spin-lattice) characterizes the return of the magnetization vector along z axis,
while $T_2$ (spin-spin) characterizes the decay of transverse component of the magnetization vector.
$T_1$, $T_2$, and proton density are important MR parameters, which vary from tissue to tissue.
Tissues with a short $T_1$ will show as bright on $T_1$-weighted images, examples including fat, sub-
acute hemorrhage, protein-rich fluid, slowly flowing blood, etc. Tissues with a long $T_2$ will show
as bright on $T_2$-weighted images, examples including increased water, as in edema, tumor,
infarction, inflammation, etc. $T_1$, $T_2$ of some normal tissue types is given in Table. 1 according to
previous literature (33). Since $T_1$, $T_2$ are intrinsic properties of tissue and varies from different
tissue types, an important advantage of MR imaging is the high contrast between various soft
tissues in the body it provides.

Table 2.1 $T_1$, $T_2$ relaxation times of some normal tissue types at 3T and 1.5T (measured at $37^\circ$) (33)

<table>
<thead>
<tr>
<th>Tissue</th>
<th>$T_1$ (ms)</th>
<th>$T_2$ (ms)</th>
<th>$T_1$ (ms)</th>
<th>$T_2$ (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>White matter</td>
<td>1084 ± 45</td>
<td>69 ± 3</td>
<td>884 ± 50</td>
<td>72 ± 4</td>
</tr>
<tr>
<td>Gray matter</td>
<td>1820 ± 114</td>
<td>99 ± 7</td>
<td>1124 ± 50</td>
<td>95 ± 8</td>
</tr>
<tr>
<td>Skeletal Muscle</td>
<td>1412 ± 13</td>
<td>50 ± 4</td>
<td>1008 ± 20</td>
<td>44 ± 6</td>
</tr>
<tr>
<td>Liver</td>
<td>812 ± 64</td>
<td>42 ± 3</td>
<td>576 ± 30</td>
<td>46 ± 6</td>
</tr>
<tr>
<td>Optical nerve</td>
<td>1083 ± 39</td>
<td>78 ± 5</td>
<td>815 ± 30</td>
<td>77 ± 9</td>
</tr>
<tr>
<td>Spinal cord</td>
<td>993 ± 47</td>
<td>78 ± 2</td>
<td>745 ± 37</td>
<td>74 ± 6</td>
</tr>
<tr>
<td>Blood</td>
<td>1932 ± 85</td>
<td>275 ± 50</td>
<td>1441 ± 120</td>
<td>290 ± 30</td>
</tr>
</tbody>
</table>

In MR imaging, applying only the main static field $B_0$ is not enough to distinguish the signals
generated from different spatial locations, because all spins precess at the same resonance
frequency defined by \( \omega = \gamma B_0 \). The spatial localization is obtained by using linear gradient magnetic fields such as \( G_x, G_y, G_z \). By using the linear gradient magnetic fields in addition to the main static field, the applied field becomes

\[
B = (B_0 + G_x x + G_y y + G_z z) z ,
\]

which is an inhomogeneous field that still points in the \( z \)-direction but the field strength varies with the \( x, y, z \) direction. In such case, the resonance frequency of the nuclei is also varies with the spatial location, which makes the spatial encoding possible.

The time dependence of the magnetization vector \( M(t) \) in the presence of an applied magnetic field is governed by Bloch equation as follows:

\[
\frac{dM}{dt} = M \times \gamma B - \frac{M_x i + M_y j}{T_2} - \frac{(M_z - M_0) k}{T_1} ,
\]

where \( i, j, k \) are unit vectors in the \( x, y, z \) directions respectively, \( M_0 \) is the equilibrium magnetization arising from the main field \( B_0 \) and \( B \) includes various magnetic field applied (34).

### 2.2 MR data acquisition and image reconstruction

The Bloch equation in Eq. (2.3) can be written as matrix form as follows:

\[
\frac{dM}{dt} = \begin{pmatrix}
-1/T_2 & \gamma B_0 & 0 \\
-\gamma B_0 & -1/T_2 & 0 \\
0 & 0 & -1/T_1
\end{pmatrix} M + \begin{pmatrix}
0 \\
0 \\
M_0
\end{pmatrix}
\]

The solution to this equation is

\[
M(t) = \begin{pmatrix}
M_x(t) \\
M_y(t) \\
M_z(t)
\end{pmatrix} = \begin{pmatrix}
e^{-t/T_2} & 0 & 0 \\
0 & e^{-t/T_2} & 0 \\
0 & 0 & e^{-t/T_1}
\end{pmatrix} R_x(\omega_0 t) M^0 + \begin{pmatrix}
0 \\
0 \\
M_z(1 - e^{-t/T_1})
\end{pmatrix}
\]

(2.5)
The solution to the time-varying gradient fields

\[ B(r, t) = [B_0 + G(t) \cdot r]k \]  

is

\[ M(r, t) = M^a(r) e^{-\frac{t}{T_2}} e^{-i\omega_0 t} \exp(-i y \int_0^t G(\tau) \cdot r d\tau) \]  

(2.7)

If we ignore the relaxation term \( \exp(-t/T_2) \), and only deal with 2D (planar) imaging methods, then only the linear gradients \( G_x(t) \) and \( G_y(t) \) are considered, the resultant received time signal from becomes as follows:

\[ s(t) = \int_x \int_y m(x, y)e^{-i2\pi[k_x(t)x+k_y(t)y]}dx dy, \]  

(2.8)

where

\[ k_x(t) = \frac{\gamma}{2\pi} \int_0^t G_x(\tau) d\tau \]

\[ k_y(t) = \frac{\gamma}{2\pi} \int_0^t G_y(\tau) d\tau \]

Eq. (2.7) is usually referred to as the signal equation.

If we refer to the Fourier transform of \( m(x, y) \) as \( \mathcal{M}(k_x, k_y) \), the signal can then be expressed as:

\[ s(t) = \mathcal{M}(k_x(t), k_y(t)) \]  

(2.9)

\[ s(t) = \mathcal{M}\left(\frac{\gamma}{2\pi} \int_0^t G_x(\tau) d\tau, \frac{\gamma}{2\pi} \int_0^t G_y(\tau) d\tau\right) \]  

(2.10)

both \( k_x, k_y \) are in units of spatial-frequency, typically cycles/mm or cycles/cm. Therefore, the received signal equals the value of the 2D Fourier transform of \( m(x, y) \) at some spatial...
frequency. The 2D Fourier transform of the MR image measured is called the “k-space” (35,36), where \( k \) represents the spatial-frequency variable. In order to reconstruct \( m(x, y) \), the trajectories of the given by \( s(t) \) should cover a sufficient part of \( k \)-space.

In practice, the \( k \)-space is often sampled with sampling periods of \( \Delta k_x \) and \( \Delta k_y \) on a Cartesian grid, the sampling periods determines the effective Field-of-View (FOV) in the image, the sampling periods and effective FOV are related by the following equation according to the properties of the Fourier transform:

\[
\text{FOV}_x = \frac{1}{\Delta k_x} \quad \quad \text{FOV}_y = \frac{1}{\Delta k_y}
\]  

(2.11)

### 2.3 Diffusion tensor imaging (DTI)

Diffusion tensor magnetic resonance imaging (DTMRI or DTI) is a noninvasive method for elucidating the diffusion process of water molecules inside biological tissues (37,38). Such diffusion information provides exquisite details on tissue microstructure, as well as changes in those tissues with physiological or pathological states, which makes DTI a promising tool for monitoring brain development and abnormalities (37). Since it’s a very sensitive biomarker for investigation of the structural integrity of the brain, DTI has been widely used in a broad spectrum of clinical applications as well as provides great potential in the field of neuroscience and medicine. Various researches using DTI for investigation regarding WM changes caused by brain development and aging (39,40), detecting abnormalities in normal-appearing WM due to disease (1), as well as identifying pathologic specificity (7). The current development of diffusion MRI, which includes advancements in fast imaging technology, image analysis, etc., makes diffusion MRI the most promising method to study the human brain in vivo.
2.3.1 Basics of diffusion

The term “diffusion” refers to the random motion of molecules (also called Brownian motion) in a fluid, which results from the thermal energy carried by these molecules. The diffusion coefficient, $D$ is used for measuring the relative amount of diffusion, which depends on the size (mass) of the molecules, the temperature as well as nature of the medium. The water molecule is the most studied in DTI, even though some other metabolites are also studied. Usually, the diffusion of water molecules inside biological tissues is different in different directions or anisotropic, due to the cell membranes, fibers or other biological molecules. Such displacement distribution provides unique information about the internal structure and geometric organization of neural tissues on a microscopic scale noninvasively. DTI is a technique that measures such features, as well as changes in those features with physiological or pathological states. The unique microstructural information DTI provides is what makes the DTI so alluring and attracts so many researchers in recent years.

2.3.2 Diffusion tensor imaging mathematics

2.3.2.1 Tensor and diffusion ellipsoid

The diffusion process can be represented visually as an ellipsoid in 3D that reflects the root mean square displacement in each direction; Figure 2.1 gives an illustration of such a diffusion ellipsoid. Mathematically, this ellipsoid can be represented as a tensor, which is a symmetric $3 \times 3$ matrix
Diffusion tensor imaging generally assumes a monoexponential signal decay as the diffusion-weighting “$b$ factor” (characterized by the gradient pulses) increase.

$$S_i = S_0 \exp(-bD),$$

(2.12)

where $S_0$ is the signal intensity without diffusion sensitizing gradients. The diffusion coefficient $D$ is used to measure the degree of diffusion. However, in biological systems, the diffusion process cannot be fully represents by a single value $D$ but the apparent diffusion coefficient ($ADC$), which is also noted as $D$. In 3D anisotropic diffusion, the degree of diffusion is often different in different directions, so does the measured $ADC$. In such case, anisotropic diffusion is modeled as a diffusion ellipsoid that shows the root mean squared displacement in each direction at a particular time. Such diffusion ellipsoid can be represented mathematically by a symmetric $3 \times 3$ tensor as in Eq (2.13).
\[ D = \begin{bmatrix}
D_{xx} & D_{xy} & D_{xz} \\
D_{yx} & D_{yy} & D_{yz} \\
D_{zx} & D_{zy} & D_{zz}
\end{bmatrix} \] (2.13)

where \( D_{ij} = D_{ji} \), with \( i, j = x, y, z \).

The eigenvalues \((\lambda_1, \lambda_2, \lambda_3)\) and corresponding eigenvectors \((\hat{e}_1, \hat{e}_2, \hat{e}_3)\) of the diffusion tensor \( D \) can be obtained by diagonalization of the diffusion tensor.

\[ \Lambda = \begin{bmatrix}
\lambda_1 & 0 & 0 \\
0 & \lambda_2 & 0 \\
0 & 0 & \lambda_3
\end{bmatrix} = R \cdot D \cdot R^T \] (2.14)

where \( R \) is the rotation matrix with columns being the eigenvectors of the tensor. \( R^T \) is the matrix transpose of \( R \).

The eigenvectors describe the directions and degrees of diffusion in each axis of principle diffusion. As mentioned before, the diffusion tensor can be visualized as a tensor ellipsoid, with eigenvectors define the directions of the principle axes, and the eigenvalues define the radii of the diffusion tensor ellipsoid. When the eigenvalues are nearly equal (i.e. \( \lambda_1 \sim \lambda_2 \sim \lambda_3 \)), the diffusion is considered isotropic and the diffusion ellipsoid is a sphere; when eigenvalues are different from each other (i.e. \( \lambda_1 > \lambda_2 > \lambda_3 \)), the diffusion is considered anisotropic and can be visualized as an elongated ellipsoid. Fig. 2.2 gives illustrations of the diffusion ellipsoid and tensors for isotropic unrestricted diffusion, isotropic restricted diffusion, and anisotropic restricted. (Figure is derived from (41)). Since the magnitude of the eigenvalue can be affected by changes in the tissue microstructure, it is a very sensitive biomarker and provides information for changes in the states in pathology (i.e. tumor) and physiology (i.e. aging) (42).
Since there are six independent unknown parameters in the tensor, the diffusion tensor elements can be calculated from a minimum of six non-collinear, non-coplanar encoding measurements.

\[
\ln \left( \frac{S(b)}{S(b=0)} \right) = -\sum_{i=1}^{3} \sum_{j=1}^{3} b_{ij} D_{ij} = -(b_{xx} D_{xx} + b_{yy} D_{yy} + b_{zz} D_{zz} + 2b_{xy} D_{xy} + 2b_{xz} D_{xz} + 2b_{yz} D_{yz}) = -\text{Trace}(bD),
\]

(2.15)

where \(S(b)\) and \(S(b = 0)\) are the diffusion weighted signals and non-diffusion weighted signals respectively. \(b_{ij}\) is a component of the symmetric b-matrix, \(b\). The tensor \(D\) is then calculated using obtained DWI and its corresponding b-matrix using multivariate linear regression (43). However, the Eq. (2.15) only applies in the range of b-values up to 2000\(s/mm^2\), for b-values.
great exceeding 2000 $s/mm^2$, DWI single intensity no longer conforms to monoexponential decay as described above (44).

### 2.3.2.2 Tensor-derived parameters: quantification and visualization

Since it is difficult to interpret the diffusion tensor image data with a $3 \times 3$ diffusion matrix at each voxel, tensor-derived parameters are calculated based on DTI data to help simplify the data and distill the image information into simpler scalar maps. The most widely used two tensor-derived parameters are Trace and anisotropy of the diffusion tensor. The Trace of the tensor, which is defined as the sum of the diagonal elements of the tensor $D$, is rotationally invariant and a measure of the degree of diffusion. Mean diffusivity (MD), which is the average of eigenvalues ($Trace/3$), is often used instead of $Trace$. Fractional anisotropy (FA) is used in many literatures as invariant measure of anisotropy (45). Other tensor-derived parameters including axial diffusivity (AD) and radial diffusivity (RD) are also widely used and defined as follows:

$$MD = \frac{Tr}{3} = \frac{\lambda_1 + \lambda_2 + \lambda_3}{3},$$  \hspace{1cm} (2.16)

$$FA = \sqrt{\frac{(\lambda_1 - \bar{\lambda})^2 + (\lambda_2 - \bar{\lambda})^2 + (\lambda_3 - \bar{\lambda})^2}{2(\lambda_1^2 + \lambda_2^2 + \lambda_3^2)}},$$  \hspace{1cm} (2.17)

$$AD = \lambda_1,$$  \hspace{1cm} (2.18)

$$RD = \frac{\lambda_2 + \lambda_3}{2},$$  \hspace{1cm} (2.19)

FA is often used as a summary measure of microstructural integrity; it ranges from 0 for isotropic diffusion to 1 for completely anisotropic diffusion. Since it is highly sensitive to microstructural change, it has been widely used in a variety of clinical applications as well as helps identify pathological conditions. MD is an inverse measure of membrane density; which reflects the average magnitude of molecular displacement by diffusion. The MD values for both
GM and WM are very similar and higher for CSF. Both Voxel-based and ROI-based methods have been used for comparison and analysis of those parameters to help identify brain abnormalities in either longitudinal studies or group comparisons, which remain as a very active research area within the field of DTI.

Although those measures provide information about the degree of diffusion, they do not provide any information about tensor orientation or tensor shape, as tensors with different shapes and eigenvalues can have the same FA, MD values. Therefore, while FA and MD are highly sensitive but fairly nonspecific biomarkers of neuropathology and microstructural architecture, extra information can be obtained by eigenvectors, or the combination of different tensor-derived parameters. The tensor orientation is described by the major eigenvector direction, which is generally parallel to the direction for white matter tract. An RGB (red-green-blue) map is usually used for indicating the orientation of eigenvectors. The color encoding for the map as shown in Figure 2.6 C is: red for fibers crossing from left to right; green for fibers traversing in anterior-posterior direction; blue for fibers going from superior to inferior.

Figure 2.6. Example maps of MD (A), FA (B), and color FA (C)
2.3.3 Fiber tractography

Fiber tractography is a method used for identifying inter-voxel connectivity on the basis of the anisotropic diffusion of water within the human brain, especially in white matter (41,46). Tractography provides the only available noninvasive tool for identifying degrees of connectivity between different brain regions, which provides information for depicting human neuroanatomy as well as helping understanding brain function, brain pathology and disease effect in vivo (47).

White matter tracks can be estimated by starting at a specified voxel location (the “seed” point), from which the direction of propagation is estimated based on the direction of the major eigenvector ($v_1$). The tract is terminated when reach a predefined threshold, usually a low anisotropy region where no coherent tract orientation is presented within a pixel, or when the angle change between pixels is large (48). Tractography can be used to generate anatomically plausible estimates of white matter trajectories within the human brain, which provides a vivid visualization of WM trajectories in 3D and segmentation of specific brain regions. Figure 2.7 shows fiber tracking streamlines. (Figure is derived from (41))
Figure 2.7 Schematic diagram of the linear propagation approach, which was dubbed FACT (fiber assignment by continuous tracking). Arrows indicated fiber orientations at each pixel. Red lines are FACT trajectories.

2.3.4 DTI artifacts

Eddy-current induced artifact and motion-induced artifact are two major sources of DTI artifacts. Eddy current effect is induced when strong gradient pulses are switched rapidly. The consequence of such effect is the geometric distortion in the obtained images, which includes contraction, overall shift and shear of final images. Eddy currents can be either reduced by using bipolar diffusion-weighting schemes (49), which are now standard in MRI system, or by post image processing procedure such as eddy current correction, which is implemented in scientific software such as FSL (FMRIB’s Software Library, Oxford, UK).
Another major source of DTI artifact comes from motion, both in terms of bulk motion such as brain motion and localized motion such as cardiac pulsation. Motion leads to phase shift, which results in attenuation of the signal. The consequences of motion include large signal variation across image, usually manifest as signal dropout across the whole image or localized signal dropout. The potential cure for localized motion artifact includes cardiac gating by acquiring data only during the diastole of the cardiac cycle. This prevents the motion artifacts induced by movement of the heart throughout the cardiac cycle and movement of the lungs during the respiratory cycle. However, cardiac-gating increases the acquisition time and therefore is not prevalent in clinics. Image QC tools can be also used for precluding artifact contaminated images, as well as image post-processing software packages for image registration such as FSL and SPM (Wellcome Department of Cognitive Neurology, London, England).

Echo-planar imaging (EPI) (50), which is the most widely used DWI method, solves part of the problem by shortening the image acquisition time using a single shot typically around 100 msec, with which the motion artifact is largely suppressed. However, EPI usually provides image with coarse image resolution (~ 2 mm), and it also suffered from other types of artifacts such as susceptibility artifact caused by magnetic field inhomogeneities, which also cause geometric distortions. The geometric distortions in EPI caused by magnetic field inhomogeneity can be solved by obtaining a field map during the data acquisition. The field map is a measure of the field inhomogeneity and used for shifting the signal to its correct location and correcting the obtained images retrospectively. It can be estimated by measuring the phase difference between gradient echo image data at two different echo times and can be easily obtained within 2-3 minutes. Figure 1.5 gives an example of data suffered from motion artifact and cardiac-induced artifact.
Figure 2.8 Human brain data affected by motion artifact and cardiac artifact. (A) Typical DWI affected by motion artifact, artifact-free data (left) vs. data corrupted by motion artifact (right). (B) Typical DWI affected by cardiac artifact, artifact-free data (left) vs. data corrupted by cardiac artifact (right).

2.3.5 Main applications and current challenges of DTI

The early clinical application of DTI dated back to 1980s (51), almost at the same time when DTI came into existence. Due to its exceptional sensitivity to changes in white matter integrity, DTI is certainly a sensitive biomarker for detecting microscopic changes in tissue properties and has been applied to numerous research studies (42). The most successful DTI applications include brain maturation (52), diffuse demyelination (53), ischemic (including stroke) (54,55), edema (56), and traumatic brain injury(57). A detailed review about the most common clinical applications of DTI can be found in (1). DTI can also be used for predicting diseases at an early stage, while information is still not accessible with conventional MRI, as well as be used for disease treatment (37,58). Moreover, the use of multiple DTI-derived indexes, such as the combination of MD and FA, as well as the combination of DTI with other traditional MR modalities (e.g. T1, T2, perfusion, etc.) also help improve the diagnostic accuracy of complex disease such as multiple sclerosis (MS).
Since DTI is inherently a noise sensitive and artifact-prone imaging technique, it suffers from many sources of artifacts as we discussed in the previous section. Therefore, the image quality assurance and robust image analysis is of crucial importance in DTI. In the following chapters, we will first introduce and compare three existing quality control tools for DTI, aiming to address the potential problems of data quality control. An improved spatial regression analysis of diffusion tensor imaging for lesion detection during longitudinal progression of neurodegenerative disease in individual subjects is proposed and validated through both simulation data and human in vivo brain data in the following chapters.
Chapter 3 Comparison of quality control software tools for DTI

3.1 Introduction

For DTI, a data quality control tool is a tool that can be used for controlling the quality of the input image dataset and finding potential outliers, that is, images corrupted by different artifacts, and excluding them in the subsequent tensor calculation and image processing. Despite increasing clinical and research applications of DTI, quality control in DTI remains elusive and the search for suitable quality control tools continues. In recent years, several approaches have been developed for automatically detecting and correcting or excluding anomalous observations in DTI, using techniques such as iterative linear (59) and nonlinear least squares (60) fitting for tensor calculation. Each method has its own tradeoffs to achieve the balance of efficiency and accuracy. However, to date, there have been few direct comparisons among them and the practical impact of these tradeoffs is not well characterized. In this Chapter, we evaluated three popular QC tools for DTI, DTI studio (Latest x64 bit version of DTI Studio) (61), DTIprep (DTIPrep 1.2.2_linux64, released in Dec 2013) (62) and TORTOISE (V2.0.1, released in Dec 2013) (63).

Among different DTI artifacts, some are caused by hardware imperfections and can be corrected using image co-registration, one such example being eddy current-induced artifact. Popular scientific software packages, such as FSL (FMRIB, Oxford, UK) and SPM (Wellcome Department of Cognitive Neurology, London, UK), already provide effective corrections towards this type of artifacts. Therefore, our study focuses mainly on two major artifacts that are
beyond correction by such means: motion artifact due to patient movement and brain pulsation artifacts caused by cardiac activity. Both synthetic and human in vivo data were used to quantify not only the adverse effects of these artifacts to tensor calculation but also the effectiveness of these three QC tools in correcting and compensating for these artifacts. The specific objective of this study is to investigate different tools for identification of the data quality issues and the selection of the tools for addressing those issues.

In order to obtain this goal, we studied how different artifacts and the ratio and distribution of different artifacts impact the accuracy and precision of fractional anisotropy (FA), and mean diffusivity (MD) for the three software packages. By using the same datasets, we were able to crosscheck the performance of these tools in the presence of typical artifacts in DTI. In addition, we discussed the technique each of the tools was based on, and analyzed the pros and cons for each briefly. We also showed the potential for combining these tools together to build an effective data processing pipeline.

### 3.2 Material and methods

#### 3.2.1 Introduction to three quality control tools

<table>
<thead>
<tr>
<th></th>
<th>TORTOISE</th>
<th>DTIprep</th>
<th>DTIstudio</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Operating System (OS)</strong></td>
<td>Mac/Linux</td>
<td>Mac/Linux</td>
<td>Windows</td>
</tr>
<tr>
<td><strong>GUI or Command Line</strong></td>
<td>GUI/Command line</td>
<td>GUI/Command line</td>
<td>GUI</td>
</tr>
<tr>
<td>Allow Batch scripting</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>-----------------------</td>
<td>------</td>
<td>------</td>
<td>-----</td>
</tr>
<tr>
<td>Input data format</td>
<td>Philips Par/Rec, Dicom, Siemens Mosaic Dicom, NIFTI, Bruker data format</td>
<td>DICOM /NRRD.</td>
<td>Philips Par/Rec, Dicom, NIFTI, NRRD, Raw image, Analyze, etc.</td>
</tr>
<tr>
<td>Output data format</td>
<td>NIFTI, NRRD, VTK, etc.</td>
<td>NRRD</td>
<td>Analyze, raw data, NRRD</td>
</tr>
</tbody>
</table>

Among all the available software packages, DTI studio, DTIprep and TORTOISE were chosen due to their popularity and availability in the DTI research field. A brief summary of computer configuration requirements for these packages is provided in Table 3.1. Among the three tools studied, DTIprep specializes in and emphasizes quality control of DTI. DTIstudio and TORTOISE were initially designed to provide a comprehensive DTI data processing routine, within which quality control is an essential step of the entire DTI process workflow. We focused mainly on the functionalities of QC of these tools in our comparison. Other functions may be mentioned here, but were not the focus of our comparison.

DTIstudio is a comprehensive DTI processing software package developed at Johns Hopkins University (JHU) that provides image registration, tensor calculation, fiber tracking, etc. An efficient and easy-to-use function for outlier detection and correction for DTI datasets is an important part of this package. The software implements multiple analytical tools, including linear regression, nonlinear regression, slice-wise outlier detection and pixel-wise outlier
detection (i.e. Yue’s method (64)), in order to compensate or correct outliers in DTI. Yue’s method provides a precise pixel-wise detection of image outliers, which has been proven to be very effective in the presence of some localized artifacts (e.g. cardiac artifact). This package runs under Windows OS and provides a user friendly GUI, which makes it easy to use even for those with little computer knowledge.

TORTOISE is another DTI processing package developed by the National Institute of Health (NIH) Pediatric Neuroimaging Diffusion Tensor MRI group. As an integral part of its DTI processing workflow, TORTOISE detects image outliers and excludes them during tensor calculation using either linear or nonlinear methods such as RESTORE (59), and informed RESTORE (iRESTORE) (60). In addition, the newest version of TORTOISE contains another module called bup_bdown_TORTOISE that provides correction for susceptibility-induced distortions in EPI-DTI data.

DTIprep is a software package specially designed for DTI quality control purposes by the University of North Carolina, Chapel Hill, University of Iowa and University of Utah. It can be fully automated under the command line mode. The built-in algorithm for outlier detection is simple but effective; quality control results can be reported and saved automatically. The pipeline for DTI Quality Control includes data format conversion, image information check, eddy current/ head motion correction, and outlier check.

3.2.2 Two major DTI artifacts

In this study, we focused on quantitative comparisons of three major QC tools towards compensating and correcting two main DTI artifacts.
3.2.2.1 Within-volume bulk motion artifacts

Motion artifact due to patient head motion is among the most common artifacts in DTI, especially when scanning non-cooperative patients such as infants and young children. Motion can happen between and within single-shot acquisition of DWI volumes. Although motion artifacts between DWIs of different directions can be effectively corrected through advanced image reregistration, corrupted image signals due to motion within the single-shot EPI acquisition of single DWI volume are beyond repair (65). Typical effects for within volume motion artifacts include ghosting and large signal variations across the image (7).

3.2.2.2 Localized cardiac artifacts

Besides bulk body motion-related artifacts, regional physiological motion such as non-linear deformation of brain tissue due to cardiac pulsations can also lead to a significant amount of phase shift of MRI signal and, therefore, signal loss (38). Pulsatile brain motion occurs as a result of arterial expansion, which usually follows systole (66). Cardiac gating has not been prevalent in clinics due to its significantly longer acquisition time. For non-gated data, about 20% of the images will be collected during systole and are likely to display severe artifacts, but an additional fraction of the images may also be affected to a lesser degree (67). Compared to bulk motion artifacts that usually spread through a whole slice, cardiac artifacts mainly induce regional signal dropouts around ventricles and within the cerebellum.
3.2.3 Image data for quantitative comparisons of three QC tools

3.2.3.1 Human DTI data

Three DTI data sets contaminated with common DTI artifacts, including the two types of artifacts mentioned above, were used in this study. All DTI data were acquired with a single-shot dual-echo spin-echo EPI sequence with parallel imaging techniques, and each consists of two repeated scans of the same subject. Among these repeated scans, one was contaminated with motion or cardiac artifacts or both, and the other was artifact-free through careful visual inspection. Three DTI datasets were acquired on a 1.5T GE scanner (General Electric Medical Systems, Milwaukee, WI) with following imaging parameters: 30 non-collinear directions with \( b = 1000 \text{ s/mm}^2 \) and five repeats of minimally weighted images \( (b_0 = 0 \text{ s/mm}^2) \), \( TR/TE = 10000ms/66.3ms \), voxel resolution of \( 1.9 \times 1.9 \times 2.5 \text{ mm} \).

3.2.3.2 Monte Carlo simulation

Adverse effects of DTI-specific artifacts to the accuracy and precision of tensor-derived parameters are often complicated by several confounding factors. In order to achieve comprehensive evaluation of the effectiveness of artifact correction among the three QC tools, Monte Carlo simulations were performed using various combinations of the following five conditions:

1. Motion artifacts: Within-volume bulk body motion artifacts were simulated with DWI intensity decreased by 90% within a slice. Localized cardiac pulsation artifacts were simulated with DWI intensity decreased by 10%, 25% or 50% near the area prone to cardiac artifact.

2. SNR: This is the signal-to-noise-ratio value for the baseline image \( (b = 0 \text{ s/mm}^2) \) with a single measurement, or when the number of excitations (NEX) is 1. Since both real and
imaginary parts of the signal is contaminated with uncorrelated Gaussian noise with zero-mean and equal variance, the magnitude images, which are calculated from the magnitude of real and imaginary parts of the signal, will follow a Rician distribution (68,69). Rician noise was simulated and three different levels of SNRs = 25, 50, 125 were evaluated.

A similar approach for MC simulation as previously reported (70) was used in this study. Briefly, at each SNR level, the effect of thermal noise was first generated using complex random numbers with their real and imaginary parts sampled independently from a Gaussian distribution function with zero mean and a standard deviation determined by the desired SNR level. The real parts of the complex noise signals were then added to the noise-free baseline signal \( S_0 \) and DW signals \( S_i \). The magnitude of the final complex data was then used to synthesize the noisy DTI datasets that were further used for calculations of the noisy tensors.

(3) \( N_{UDG} \), the number of unique diffusion-weighting gradient directions used in the DTI scan: Three different diffusion weighting (DW) schemes with number of diffusion directions = 12, 30, 60. In all instances, the ratio of the number of DWIs to the number of \( b_0 \) images (\( N_{DWI}: N_{b_0} \)) in each set was maintained at 6:1 to simulate an optimal DTI acquisition (71).

(4) \( Corrupt \% \): This is the percentage of DWI volumes contaminated with artifacts with respect to the total number of DWIs of the DTI dataset. For each selection of \( N_{UDG} \), three levels of corrupted datasets were considered with \( Corrupt\% = 8\%, 17\%, \) and 33\%.

(5) \( Evenness \): This describes whether corrupted gradients are evenly or unevenly distributed over the unit sphere span by diffusion-weighting gradient vectors or not. DTI requires DWIs acquired with at least six unique diffusion-weighted gradients that evenly sample the hemisphere. Our assumption was that the results will be affected more by the unevenly distributed artifacts.
rather than evenly distributed. To evaluate how the evenness of spatial distribution of the corrupted DWIs lead to error and bias in the final results, two extreme scenarios were considered: i) evenly distributed, where corrupted DWIs were uniformly distributed on the hemisphere; and ii) focally distributed, where corrupted DWIs all resided within a narrow range of spatial orientation (extremely uneven).

Previous studies (72) have shown that the minimum energy (ME) method provides the uniformity and regular sampling in a hemisphere and was therefore used in this study to choose evenly distributed subsets of diffusion weighting gradient direction from DTI data with a total number of diffusion weighting direction of $N_{UDG}$. ME is defined as the minimum total interaction (Coulombic) energy of $2N_e$ unit charges on the unit sphere (73). This algorithm adjusts the orientation of gradient vectors until the sum of interaction energy between every possible pair of charges on the unit sphere is minimized (71). The definition for ME optimization criterion is as follows:

$$E = \sum_{i=1}^{2N_e} \sum_{j>i}^{2N_e} \frac{1}{\|\hat{G}_i - \hat{G}_j\|}$$

(3.1)

where $N_e$ is number of diffusion gradient directions, and $\hat{G} = [G_x, G_y, G_z]$ is the unit gradient vector representing the diffusion-encoding direction.

In order to find subset gradients that focally distributed, we chose the gradients that had the minimum overall distance between each other. The distance between two vectors was calculated as Euclidean distance as follows:

$$D(\hat{G}_i, \hat{G}_j) = D(\hat{G}_j, \hat{G}_i) = \sqrt{(G_{xi} - G_{xj})^2 + (G_{yi} - G_{yj})^2 + (G_{zi} - G_{zj})^2}$$

(3.2)

An exhaustive search was used to help select optimal evenly/unevenly distributed subsets. A subset of N directions was first chosen at random and then optimized via different combination
of the subset of $N$ to either minimize the $ME$ equation described in Eq. (3.1) to find the evenly distributed subset, or to minimize Eq. (3.2) to find the focally distributed subset.

With different combinations of the abovementioned five factors, a total of 81 datasets were generated using the Monte Carlo simulation. The simulated data were based on an artifact-free FA slice at the anterior-posterior commissure plane from one of the human DTI datasets in this study. This FA map was first segmented into white matter (WM), grey matter (GM) and cerebrospinal fluid (CSF). To simplify the simulation, simulation settings for voxels of the same tissue type were chosen to be the same: $FA=0.7$, $\text{Trace}=2100 \times 10^{-6} \text{mm}^2/\text{s}$ ($\lambda_1>>\lambda_2=\lambda_3$) for WM, and $FA=0.2$, $\text{Trace}=2100 \times 10^{-6} \text{mm}^2/\text{s}$ ($\lambda_1=\lambda_2>>\lambda_3$) for GM.

### 3.2.4 Statistical analysis

#### 3.2.4.1 Monte Carlo simulation data

For simulation data, above-mentioned 81 datasets passed through QC procedure with three software packages, and diffusion tensors and tensor-derived quantities (FA/Trace) were then calculated from QCed datasets (i.e. datasets after quality control procedure) using three QC tools. Results were processed offline using MATLAB (The Mathworks, MA, USA). For each simulation dataset, two measures were computed over each dataset to quantify accuracy and precision of DTI derived parameters, using calculations for FA in the following examples, with the same analyses applied for Trace.

1. Biases of FA, $Bias(FA)$, as a measure for accuracy level:

$$Bias(FA) = \bar{FA} - FA_{GS},$$

(3.3)
where \( \overline{FA} \) is the mean value of all the QCed MC samples and \( FA_{GS} \) is 0.7 in WM and 0.2 in GM in MC simulation.

(2) Standard deviations of FA, \( \sigma (FA) \), as a measure for precision level:

\[
\sigma (FA) = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (FA_i - \overline{FA})^2},
\]

where \( FA_i \) is FA value from the \( i \)th sample point and \( n \) is the number of sample points.

### 3.2.4.2 In vivo human brain data

All in vivo human brain DTI datasets went through two preprocessing steps before statistical analysis. First, all datasets underwent the QC procedure using three different QC tools. Artifacts due to eddy-current and between-volume bulk body motion were simultaneously removed. Images corrupted by different artifacts were detected as outliers and subsequently excluded. In the second step, diffusion tensors were calculated from QCed datasets and tensor-derived quantities (FA/Trace) were calculated using three different QC tools. For each dataset, one non-DWI image from the artifact-free scan was selected as the reference for registration among different scans using FLIRT (FMRIB’s Linear Image Registration Tool, Oxford, UK) of FSL to remove intra-series motion. For FA/Trace of human in vivo brain data, voxel-wise quantitative measures were developed for analysis of measurement errors. Regions of Interest (ROIs) were drawn manually to include the areas that were affected by the artifacts.

(1) Measured Accuracy, \( Bias(FA) \) and \( Bias(Trace) \):

The accuracy level was measured as bias values between FA calculated using each QC tool from datasets with artifact and the corresponding values from the gold standard dataset. FA values calculated from artifact-free datasets were selected as the gold standard \( (FA_{GS}) \). For ROI analysis, FA bias was calculated as:
\[ Bias(FA) = \overline{FA} - \overline{FA}_{GS}, \]  

(3.5)

where $\overline{FA}$ and $\overline{FA}_{GS}$ are the mean FA values within an ROI from datasets with artifact and artifact-free respectively. Similar calculation was used for $Bias(Trace)$.

(2) Measured precision, $\sigma(FA)$ and $\sigma(Trace)$:

The precision level was measured as variability of FA in each ROI and calculated as in Eq. (3.6).

\[ \sigma(FA) = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (FA_i - \overline{FA})^2}, \]  

(3.6)

where $FA_i$ is FA value from the ith voxel, $\overline{FA}$ represents the mean FA value over the ROI and $n$ is the number of voxels within the ROI. Similar calculation was used for $\sigma(Trace)$.

### 3.3 Results

Table 3.2 Overall performances of different QC tools

<table>
<thead>
<tr>
<th></th>
<th>TORTOISE</th>
<th>DTIprep</th>
<th>DTIstudio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outlier Detection</td>
<td>Slice/Voxel based</td>
<td>Gradient-based</td>
<td>Slice/Voxel based</td>
</tr>
<tr>
<td>Stability</td>
<td>Less stable when $Corrup\ %$ is large</td>
<td>Stable when $corrupt\ % \leq 33%$ and $N_{UDG} \geq 6$</td>
<td>Less stable when $Corrup\ %$ is large</td>
</tr>
<tr>
<td>Efficiency</td>
<td>Moderate</td>
<td>Fast</td>
<td>Moderate</td>
</tr>
<tr>
<td>Ease-of-use</td>
<td>Moderate</td>
<td>Easy</td>
<td>Easiest</td>
</tr>
</tbody>
</table>
Pros | Comprehensive data preprocessing tools. | Fully automated and efficient. | Accurate local outlier detection and user-friendly GUI.

We evaluated the performance of three QC tools based on 81 simulated datasets. The overall performance of different QC tools is summarized in Table 3.2. Efficiency of QC tools was evaluated based on the relative time each used for QC procedure. Ease-of-use is a multi-faceted problem and both ease of learning, use and user attitudes were chosen to consider during evaluation. Admittedly, this evaluation is subjective. Stability of each tool was evaluated based on whether the FA/Trace results calculated from QCed data remained accurate within experimental precision under different corrupted ratios.
3.3.1 Comparison of QC tool results for simulation data

3.3.1.1 Effects of SNR and number of directions

3.3.1.1.1 Cardiac artifact

Figure 3.1 Biases and SDs of measured FA/Trace for datasets corrupted with cardiac artifact as a function of SNRs with different gradient schemes (SNR=25, 50, 125, Number of directions=12, 30, 60) using Eqs. (3.1) – (3.2) (unit for Bias and SD of Trace is $\text{mm}^2/\text{s}$). Corrupted data points had an intensity value set to 50% of the original signal intensity. FA/Trace values were measured in WM. Simulated $\text{FA}_{\text{GS}} = 0.7$, which is representative of the typical $\text{FA}$ value of brain WM, and $\text{Trace}_{\text{GS}} = 2100 \times 10^{-6}\text{mm}^2/\text{s}$, which is representative of trace of brain
parenchyma. DW images were all with $b_{value} = 1000s/mm^2$. Results from TORTOISE, DTIprep and DTIstudio were shown as blue dots, green triangles and red squares, respectively.

Cardiac pulsation artifact was simulated with the Monte Carlo simulation and results are shown in Fig. 3.1. In the course of our experiment, we used RESTORE/iRESTORE from TORTOISE, and pixel-wise detection (Yue’s method) from DTIstudio. The QC results depend largely on the parameters being chosen. The author stated in detail how different parameters were chosen and the guideline for choosing between RESTORE and iRESTORE in TORTOISE in (63). In our experiments, we used iRESTORE for datasets with low redundancy (Number of directions = 12, 30) and RESTORE with extra constraints when the number of directions was 60. For iRESTORE parameters, we chose Redundancy Coefficient (RC) based on the specific number of directions of imaging protocol, and the reference-sampling scheme was chosen to be the default six-direction scheme. As for DTIprep, we used the default protocol parameter for each imaging protocol with the exception of setting the bad Gradient Percentage Tolerance to a higher value (default=0.2) to avoid unexpected termination of the program. Tensors were calculated using GTRACT (74), which was also the recommended software for tensor calculation.

The results show that, in general, the bias from measured FA decreased as the numbers of gradient directions increased (from 12 to 60), and FA SD decreased as SNR ratios increased (from 25 to 125) as expected, and the performance of three QC tools became compatible with a large number of directions and at high SNR. DTIprep could detect localized artifact accurately, however, the whole gradient that contains artifact was deleted in the output QCed image since the outlier algorithm is gradient-based, which yielded a slightly large FA Bias. Similar trends were also found in changes in Trace as a function of number of directions and SNR. Since the
differences between biases were \(~0.1\) for FA and \(\sim10^{-5}\) \(mm^2/s\) for Trace, which were equivalent to or smaller than the intra-site variance components for FA and MD [21], the performance of the three QC tools were still similar within experimental precision. Similar but not exact trends were also found in GM FA/Trace (Data not shown here). Differences between different QC tools were more pronounced with small number of directions, which minimized with large number of directions. Since the corrupted ratio was low in this case, all three software could detect and compensate for the artifacts successfully.
3.3.1.1.2 Motion artifact and direction effects

(a) Measured Biases and SDs of FA (A. WM; B. GM, DIR=12, 30, 60, SNR=50)
(b) Biases and SDs of Trace (A. WM; B. GM, DIR=12, 30, 60, SNR=50)

Figure 3.2 Biases of measured FA/Trace ($\text{Bias}_{FA}, \text{Bias}_{\text{Trace}}$), SDs of FA/Trace ($\sigma_{FA}, \sigma_{\text{Trace}}$) in WM (Fig. 3.2 A) and GM (Fig. 3.2 B) of simulation data as a function of corrupted ratio (SNR=50, Number of directions=12, 30, 60) using Eq. (3.1) – (3.2). Corrupted gradients were either evenly (blue squares) or unevenly (green triangles) distributed over the unit sphere (unit for Bias and SD of Trace is mm²/s). The ratio of corrupted gradients in each dataset was set to 8%, 17% and 33%, with the outliers’ intensity value set to 10% of the original signal intensity.
Simulated FA_{GS} = 0.7, Trace_{GS} = 2100 \times 10^{-6}\text{mm}^2/\text{s} \text{ in WM and FA}_{GS} = 0.2, \text{ Trace}_{GS} = 2100 \times 10^{-6}\text{mm}^2/\text{s} \text{ in GM.}

The accuracy of single diffusion tensor measurements not only depends on the ratio of corrupted gradients within a dataset, but also depends upon the encoding scheme used. In order to investigate the effect of motion artifact as well as the distribution of the artifacts, three levels (8%, 17%, and 33%) of corrupted datasets were considered. For each level of corrupted datasets, simulations with corrupted gradients either evenly or unevenly distributed on a unit sphere were performed and investigated. All simulated datasets underwent QC procedure using three software packages; tensor and DT-derived metrics such as FA/Trace were calculated and representative results are shown in Figure 3.2.

Over the course of our experiments, we used iRESTORE/RESTORE from TORTOISE and the slice-wise detection method from DTIStudio for detection of simulated motion artifacts. Although a pixel-wise method, such as Yue’s method in DTIstudio, was a powerful QC tool in the presence of localized artifact, it became very slow and sometimes even stopped working when ratio of artifact in the dataset was relatively large.

In general, motion artifacts at some level affected the calculation of DW-derive parameters from three QC tools, and the orientation and distribution of the corrupted gradients had a great impact on the resulting FA/Trace bias and precision. Although there were exceptions, the general rule was that results calculated from evenly distributed corrupted gradients yielded smaller biases/SDs compared to the case when corrupted gradients were unevenly distributed. Differences between DTI derived quantities calculated from three QC tools were accentuated when the number of directions was small (i.e. Number of directions = 12) with unevenly
distributed corrupted gradients, and reduced with a large number of directions (i.e. Number of directions = 60) and when corrupted gradients were evenly distributed on the unit sphere.

Almost all three QC tools yielded satisfactory results in terms of FA/Trace calculation at the level of 8% corruption-- results showed minimum bias and SD, which reflected the good performance in finding and compensating for the outliers. Although results calculated for TORTOISE showed somewhat larger bias and SD, and a slightly larger FA SD from DTIprep when number of direction was small, the performance of three QC tools were still similar within the experimental precision. However, the loss in accuracy increased with increasing ratio of corrupted gradients for both DTIstudio and TORTOISE. As the corrupted ratio increased to 17%, the performance of both QC tools started to become unstable while the performance of DTIprep remained relatively stable. Resulting FA/Trace calculated from TORTOISE started to show some bias or large SD, which indicated erroneous calculation results. Although the results from DTIstudio remained stable when the corrupted gradients were evenly distributed, its results also suffered large bias when the corrupted gradients were unevenly distributed. This trend continued as the ratio of corrupted gradients increased to 33%, wherein FA/Trace results calculated from TORTOISE became totally incorrect, which showed either large bias and SD or large bias with low SD. Erroneous DT-derived parameter calculations were also observed from DTIstudio when number of directions was 12, as the number of directions increased to 30 and 60, results calculated from DTIstudio could still maintained stability when corrupted gradients were evenly distributed, while the results from unevenly corrupted gradients showed large bias or SD.

The instability of the performance of QC tools resulted in erroneous outlier detection, which sneaked into the calculation of tensors and finally caused biases and SDs in the FA/Trace. However, when the ratio of outliers increased to a certain threshold beyond detection and
compensation, the resulting FA/Trace map became totally incorrect, and was reflected as large bias and relatively small SD in the measured FA/Trace results. Similar but clearer trends were also found in FA/Trace results calculated from GM (Fig. 3.2 (B)). Although there was a slightly larger measured FA SD for results from DTIprep (compared to the other two QC tools) when number of directions was low (DIR=12), it still yielded the most robust results in the presence of global artifacts among the three QC tools.

Similar trends were also observed for measured FA/Trace values when SNR=25 and 125 (Data not shown here). While results from TORTOISE and DTIstudio started to show poor performance as the corrupted ratio increased to 17%, results from DTIprep remained relatively stable. Although there were still slight differences between results from DTIprep for evenly/unevenly-distributed corrupted gradients when the results were zoomed in, those differences were very small.

3.3.2 Comparison of QC tool results in vivo human brain data
Figure 3.3 Human brain data affected by motion artifact and cardiac artifact. (A) Typical DWI affected by motion artifact, artifact-free data (left) vs. data corrupted by motion artifact (right). (B) Typical DWI affected by cardiac artifact, artifact-free data (left) vs. data corrupted by cardiac artifact (middle); binary mask output by DTIstudio showing the location of the artifact (right). (C) Region of interests (ROIs) used for human brain data analysis. Three WM ROIs were: the body of the corpus callosum (BCC; left); the genu of the corpus callosum (GCC) and the splenium of the corpus callosum (SCC; middle). The fourth ROI was in the middle cerebellar peduncles (MCP; right). ROIs were superimposed on the sampled FA data calculated from an artifact-free dataset, using DTIstudio as an example.

![Figure 3.4 Accuracy and Precision in Human Brain Data](image)

Figure 3.4 Measured accuracy and precision in human brain data based on 6 selected ROIs. Unit for $\sigma$ (Trace) is mm$^2$/s. 6 ROIs were: BCC from subject 1; BCC from subject 2; GCC and
SCC from subject 2, the last two were in the cortex of the cerebellum for subject 1 and 3. Accuracy and Precision level of FA/Trace were calculated using Eq. (3.5) - (3.6).

A similar comparison was also performed on human brain data. In the course of our experiments on human brain data, we used pixel-wise detection (Yue’s method) from DTIstudio and iRESTORE from TORTOISE. Default quality control protocol parameters were used for each imaging protocol for DTIprep. Results from human brain data are shown in Fig. 3.4.

The previous results of simulation data indicated that when the ratio of the artifacts was low in the dataset, there tended to be no significant difference between the results using those three QC tools. The results from the human brain data further validated this. Since there was only one direction experience image corruption within the 30 diffusion directions, the portion of corrupted images was very slim and didn’t affect much of the tensor calculation; all three QC tools could detect and compensate the artifact as expected. Results showed comparable performance within the experimental precision of all tools in the presence of artifacts in human data.

Six ROIs were selected from 3 datasets to quantitatively analyze the voxel-wise variability of the results and differences between the three QC tools. Among those 6 ROIs, four were drawn manually in WM region contaminated by motion artifact and two were drawn in the middle cerebellar peduncles, as previous studies showed tensor-derived parameters affected by cardiac-induced artifact are most pronounced in the cerebellar peduncles, anterior portions of the cerebellum [22]. ROIs were superimposed on the FA/MD map of DTI datasets (Fig. 3.3 (C)). Results of human data evaluated from 6 selected ROIs are illustrated in Fig. 3.4. It is evident from Fig. 3.4 that the differences in resulting DT-derived parameters from three QC tools were small compared to the intra-session variability, which was ~0.1 for FA and ~0.05 × 10^{-3} mm^2/s for MD using the well-balanced DW scheme of 30 directions according to (75).
### 3.4 Discussion

Table 3.3 A brief summary of featured algorithms used by different software packages

<table>
<thead>
<tr>
<th>Software Package</th>
<th>Featured Algorithm</th>
<th>Formulas</th>
</tr>
</thead>
<tbody>
<tr>
<td>DTIprep</td>
<td>Normalized Correlation (NC) (62)</td>
<td>$NC(P, Q) = \frac{\sum_{i=1}^{N} (P_i \cdot Q_i)}{\sqrt{\sum_{i=1}^{N} P_i \cdot \sum_{i=1}^{N} Q_i}}$</td>
</tr>
<tr>
<td>TORTOISE</td>
<td>RESTORE (59)</td>
<td>$\sum w_i x_i^2$</td>
</tr>
<tr>
<td>DTIstudio</td>
<td>iRESTORE (60)</td>
<td>Based on RESTORE with prior information about physiological artifacts and two added constraints:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1) Condition number of b-matrix</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2) Redundancy coefficient (RC)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\sum w_i x_i^2$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Weighting term: $w_i = 1/(r_i^2 + C^2)$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$r_i$: fitting error for the $i^{th}$ data point.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C: median absolute deviation (MAD) of the residuals (fitting errors)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>{ $r_i$, $i=1,2,\ldots$ # of data points }</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$C^2$: MAD of all ISID values</td>
</tr>
<tr>
<td></td>
<td></td>
<td>id: Inter-Slice Intensity</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ISID</td>
</tr>
<tr>
<td></td>
<td></td>
<td>c: MAD of all ISID values</td>
</tr>
</tbody>
</table>

The results of the study suggest different tools made different tradeoffs. Although the tools showed comparable performance in the finding of outliers when the percentage of the outliers in the datasets was low, there were still differences among different tools in their tolerance to the percentage of outliers in the dataset. Those differences, when within the tolerance of the
percentage of outliers of each tool, were accentuated at low SNR, with a small number of diffusion gradients, and minimized at high SNR, with a large number of diffusion gradients.

As the simulation results showed (see Fig. 3.2), while both TORTOISE and DTIstudio started to show unstable performance as the ratio of corrupted gradients increased to 17% and 33%, the performance of DTIprep remained relatively stable and therefore was the most efficient and robust among the three. However, its gradient-based detecting algorithm may render itself problematic when the number of remaining gradients is too small (e.g. less than 6). The option “bad Gradient Percentage Tolerance” in the QC protocol is designed to stop the program from deleting too many bad gradients. In the case of a global artifact that affects the whole slice, both slice-wise methods provided by DTIstudio and the voxel-wise method provided by TOROTISE are not a counterpart for DTIprep, which provided efficient and accurate results. But given the consideration that it is very rare for most clinical data to have a data-corrupted ratio larger than 15%, all QC tools can give satisfactory results within the experimental precision for clinical studies.

In general, the ratio of corrupted gradients had significant impact on the calculation of DT-derived parameters when the number of directions was small (e.g. 12), while reduced when there was sufficient number of directions (e.g. 30, 60), which indicates that DW schemes with a greater number of directions are more immune to artifacts and data corruptions than schemes with a small number of directions. Therefore, when the scan time is limited, it is more optional for users to choose DW schemes with a higher number of directions, rather than more averages of the same direction, which was well approved in previous studies (72). The distribution of the corrupted gradients on the unit sphere also had a remarkable impact on the results of DT-derived parameters in general (see Fig. 3.2), which demonstrates that there is a need to either
quantitatively or visually test the imbalance of gradients before and after QC. Both TORTOISE and DTIprep take into account this and have included related functions in their software packages. TORTOISE adds Redundancy coefficient (RC) into its iRESTORE algorithms to determine whether the remaining gradients (after QC) are directionally imbalanced. DTIprep provides a visual tool for the users to see the 3D visualization of the gradients before and after the QC on the unit sphere.

The performance of each QC tool is, to a certain degree, related to the underlining algorithms. Table 3.3 gives a brief summary of featured algorithms used by different software packages: (i) normalized correlation method used by DTIprep. (ii) RESTORE/iRESTORE used by TORTOISE; and (iii) Yue’s method used by DTIstudio.

TORTOISE is a comprehensive data preprocessing and tensor calculation tool for DT images and its outlier rejection function is just one of its many useful functions. For tensor calculation, it recommends non-linear methods such as RESTORE and iRESTORE to detect outliers and calculate tensors. RESTORE is a robust tensor calculation method that uses the iterative reweighting process for pixel-based outlier rejection (59). iRESTORE improved the method RESTORE by using the prior information of physiological artifacts with two added constraints: 1) Condition number of b-matrix: to avoid ill- defined b-matrix; 2) Redundancy coefficient (RC): to determine directionally unbalanced (60). The prior information used in iRESTORE assumes that an artifact would usually cause signal dropouts rather than signal increases, which makes it very suitable for datasets corrupted by motion artifact or cardiac-pulsation induced artifact. Our results show that iRESTORE outperformed RESTORE as the percentage of outliers increased, especially for datasets with low redundancy (number of directions = 12), probably due to our assumption that the artifacts would lead to signal dropout within the slices, which meets the
prerequisite of the iRESTORE method. The many output data format options TORTOISE provides also facilitate pipeline with other popular image processing tools such as DTIstudio, FSL, SPM, etc. Although the configuration files are a little overwhelming for beginners, the GUI itself provides an easier alternative. Our tests of this software package show that, though it takes time to get familiar with it, it is a very useful and comprehensive data processing software package once the user becomes familiarized with it.

DTIstudio is another data procession and fiber tracking tool for DTI. DTIstudio provides two tools to help with visual inspection: a subtraction tool and a theoretical tool for outlier rejection purposes, which use either subtraction of multiple repeated measurements or errors between theoretical value and measured value to help detect outliers. Its voxel-wise detection method (Yue’s method) is based on RESTORE, yet adds an additional weighting term to add sensitivity to the image intensity changes along the slice (z) direction. The results of our simulation show Yue’s method was accurate and effective when used to detect localized artifacts such as cardiac artifacts. The output binary mask gives the users accurate information of the location, as well as size, of the artifacts. Results from datasets affected by motion where the slice-wise rejection method was used yielded stable results in the presence of motion artifacts that affected the whole slice when the ratio of corrupted gradients was relatively low. The diversity of DTIstudio’s output data formats makes it compatible with some of the most popular MRI analysis tools such as FMRIB Software library (FSL) and Statistical Parametric Mapping (SPM). However, while DTIstudio provides various tools for visual inspection and outlier detection, it doesn’t allow for batch scripting, which render it less likely to be used for large amounts of datasets within a limited period of time.
DTIprep is the only tool among the three that is designed specifically for the purpose of DTI data quality control. It can run under MAC or LINUX OS either from command line or user-friendly GUI. Its whole pipeline for DTI quality control includes data format conversion, data information check, data quality check, eddy-current and motion artifact correction. Tensor calculation and preliminary fiber tracking can be calculated upon user’s choice using the GTRACT program. DTIprep uses Normalized Correlation (NC) to detect intensity-related artifacts. The assumption behind this method is that normalized correlations between neighboring slices at the same slice location across all gradients, either slice-wise or interlace-wise, are normally distributed (62). Therefore it is easy to detect the outliers when such distribution is disturbed. Once a dataset goes through all the procedures of the data preprocessing, the software will output the results for the whole procedure including the outlier parameters, which gradient is detected as the outlier and excluded, how many gradients are left, etc., as a text file, which makes the user aware of the data quality. The coefficients it uses to set the threshold at NC is very crucial for getting accurate QC results -- to avoid false alarming or missing outliers, which need to be adjusted only once before running datasets using the same protocol. Our results show that this tool can detect any intensity-related artifact, whether they be localized or global, efficiently and accurately. Therefore, it is the most suitable tool for going through large quantities of datasets within a limited time frame. If inspection time is limited, we recommend this package to be the first choice. Although the input/output format of DTIprep is confined to Dicom/NRRD, it is still compatible with software packages such as GRACT (part of Slicer), DTI process ToolKit (UNC) etc., which also makes a complete data processing line that includes preprocessing, tensor calculation and fiber tracking.
There are a few limitations that may affect the validity of this study. The most significant one is the scope of this study, both in terms of the types of artifacts we chose to test and the selection of QC tools, though we believe the most representative and common artifacts in DTI were chosen. The second shortcoming is that in our study, we did not actually identify every false positive or negative for each tool. This would have been very difficult to do, as for certain software package such as TORTOISE, the outlier rejection procedure is integrated into its tensor calculation and cannot output the results of its QC procedure. Therefore, we used the calculated tensor and DT-derived metrics from each software package as an alternative way to test the effectiveness of each QC tool. The third limitation is the algorithms we chose from the three QC tools for comparison. Although there is no gold standard for choosing among those algorithms, we tried to choose the featured algorithms from each QC tool that also fit with our practical needs. Several questions that warrant further investigation include the simulation of multiple tensors and the ways in which artifacts impact the principal eigenvector (PEV) and fiber tracking for different QC tools.

### 3.5 Conclusion

In this Chapter, we presented a quantification protocol for the evaluation of three QC tools (TORTOISE, DTIstudio, DTIprep) and compared quantitatively how major DTI artifacts, as well as the ratio and distribution of different artifacts, affected the output accuracy and precision of three popular QC tools. The results of our study show that none of the tools strictly subsumed another; each tool had its own advantages and disadvantages. The observed differences in the resulting DTI contrasts between three QC tools were shown to be small when the corrupted ratio was relatively low, which indicates that all QC tools were compatible and could detect and
compensate the outliers well under such conditions. Featured algorithms from each QC tools, as well as their compatibility with other popular tools, were also discussed. The results of this study may help DTI users to choose among different QC tools and to develop a pipeline for quality control of images in diffusion tensor MRI studies.
Chapter 4 Improved spatial regression analysis of diffusion tensor imaging for lesion detection during longitudinal progression of multiple sclerosis in individual subjects

4.1 Introduction

Diffusion tensor Imaging (DTI) enables description of the degree and direction of water movement in biological tissues, which in turn provides information about white matter (WM) microstructure (3). It has been widely used in neuroscience and medicine to investigate brain development, to identify subtle changes in white matter, and to monitor pathologic severity of lesions and disease evolution (6). Group-level analysis, which includes both ROI-based (8) and voxel-based (9,10) statistical tests, are often used for statistical comparisons in longitudinal study of DTI-derived parameters, such as mean diffusivity (MD) and fractional anisotropy (FA). However, group-level analysis is most suitable when pathological changes are located in similar anatomical regions among subjects. It often fails when the longitudinal changes/effects are diverse and highly specific for individual subjects, where the inter-subject variations (within-group) exceed the between-group variations. More recent works have focused on individual-level longitudinal analysis with DTI measurements (14,15).
Both parametric (16) and non-parametric statistical methods (17,18) have been used for subject-specific longitudinal studies. Since non-normally distributed residuals of DTI parameters make statistical inferences with a parametric approach problematic (23), nonparametric resampling methods such as bootstrap and permutation have gained much favor in recent years. Compared to the parametric methods (16), nonparametric methods, which include both descriptive and inferential statistics, have the advantage of being able to achieving sufficient statistical power with minimal assumptions about the data being investigated. Bootstrap (17,24,25) has been used to quantify uncertainties of DTI-derived parameters through both simulation and real human DTI data. However, these methods rely on bootstrapping the signs or labels of residuals from regression analysis, which do not necessarily share the same distribution even under the null hypothesis. Furthermore, while bootstrap tests are known to be asymptotically consistent, the $p$-values they generate are only approximations for finite sample (76). The permutation test, first introduced to image analysis by Holmes et al (18), is able to devise a data-driven null distribution of data with only minimum assumptions and produce exact or almost exact $p$-values (19). This method provides more freedom in terms of selecting suitable summary statistics and has been widely used in the field of functional MRI (fMRI). In recent years, several permutation-based methods such as PERVADE (15) and SPREAD have been developed for voxel-wise whole brain longitudinal studies of local DTI changes or lesion evolution.

The SPREAD (14) relies on permuting time and scan labels and spatial kernel regression. It makes use of intrinsic correlation between neighboring image voxels and
overcomes major limitations of existing non-parametric statistical methods in DTI analysis, most of which depend largely on the same DTI protocol at different time points and the achievable $p$-value is often limited by the availability of existing diffusion-weighting directions. SPREAD requires as little as only one scan per time point for a valid hypothesis test, which greatly reduces the granularity of permutation and has proved to be an effective method for monitoring lesion progression. Results from Monte Carlo simulation show that SPREAD consistently outperforms voxel-based morphometry (VBM) approach when the number of repeated scans per subject is small ($n < 5$). However, the Gaussian kernel used for spatial regression blurs and dislocates important image features such as edges, common drawbacks of non-adaptive linear filtering. These intrinsic limitations could introduce partial volume or voxel-averaging artifacts and consequently affect the accuracy of interpretation of the orientation of neuron fiber structures (77).

In this work, we improved the sensitivity and accuracy of SPREAD by incorporating nonlinear anisotropic filtering. This new method is dubbed improved SPREAD or iSPREAD. The nonlinear adaptive filtering, based on the nonlinear partial differential equation proposed by Perona and Malik (78), prefers intra-region smoothing over inter-region smoothing, which leads to nonlinear scale in image space. The method has been applied for image enhancement (79), edge detection (80), and image segmentation (18,81) to preserve regional boundaries as well as for noise reductions (82,83). The goal of the current work is to demonstrate that the detection sensitivity of SPREAD can be improved substantially by adapting nonlinear anisotropic filtering when monitoring subject-specific
longitudinal DTI changes. Our results suggest the method preserves the heterogeneous spatial correlation between neighboring voxels, which allows identifying local voxel-wise changes in the brain in single subject longitudinal studies.

4.2 Methods

4.2.1 Overview of iSPREAD

In SPREAD (14), the longitudinal tensor-derived parameter $D_{nsti}$, which is the $n^{th}$ subject’s $s^{th}$ scan at time $t$ measured for the $i^{th}$ voxel ($n = 1, 2, ..., N, s = 1, 2, ..., S, i = 1, 2, ..., I, t = t_1, t_2, ..., t_T$), is modeled as a continuous spatial function $\Phi_n(z_i)$, where $z_i$ is the 3D spatial coordinate at the $i^{th}$ voxel, superimposed with measurement errors $\varepsilon_{nsti}$. The disease effect is modeled as the longitudinal DTI change with rate $\beta(z_i)$ at the $i^{th}$ voxel (Eq. (4.1)).
\[ D_{\text{nsti}} = \Phi_n(z_i) + \beta(z_i) \odot f(z_i, t - t_1) + \epsilon_{\text{nsti}} \quad , \tag{4.1} \]

where \( \odot \) is the Hadamard product, \( f(z_i, t') \) is a continuous tensor-valued function of \( z_i \) and \( t' \) with \( f(z_i, 0) = 0 \). \( \beta(z_i) \) is a continuous, tensor-valued spatial function. The statistical theory behind SPREAD is that both the scan (\( s \) in Eq. (4.1)) and the time (\( t \) in Eq. (4.1)) are interchangeable without affecting the distribution functions of \( D_{\text{nsti}} \) under the following null hypothesis:

\[ H_{00}: \beta(z_i) = 0 \quad , \tag{4.2} \]

Given the exchangeability, the scan/time labels for the DTI-derived parameters such as FA or MD maps from each subject are randomly permuted at each voxel for \( N = 1000 \) times to generate a permutation distribution under the null hypothesis for each voxel. The original FA/MD images, together with the permutated images are smoothed in iSPREAD using the nonlinear anisotropic instead of Gaussian filtering for edge-preserving image enhancement as well as for maintaining spatial correlation between neighboring voxels, which will be discussed in detail in the next section. In the third step, a repeated measures comparison is performed for each subject between the baseline scan and scans at other time points after permutation and image smoothing to test whether there are subject-specific longitudinal changes of DTI-derived parameters between any two time points. The following voxel-wise test statistic was chosen to illustrate the temporal changes in FA at each voxel.

\[ \Delta FA_i = \frac{1}{\sqrt{2}} \sum_{n=1}^{N} | \overline{FA}_{n,1i} - \overline{FA}_{n,2i} | \quad , \tag{4.3} \]

where \( \overline{FA}_{n,ti} = \frac{1}{s} \sum_{s=1}^{s} FA_{nsti} \) is the scan average and \( \overline{FA}_{n,tl} = \frac{1}{t} \sum_{s=1}^{t} FA_{n,tl} \) is the scan
and time average for each subject. The Westfall-Young multiple testing procedure (84) is used to control the familywise error rate (FWER) in the last step. Voxels are identified as significantly changing voxels (e.g. lesion area) if the adjusted p-value is less than a predefined threshold (e.g. 0.05). A similar procedure can also be applied to the MD map. The flowchart of the method iSPREAD is shown in Fig.4.1.

4.2.1.1 Justification of permutation invariance

Statistical inference of iSPREAD depends on permuting time/scan labels, so it is important to show that the likelihood function of the permuted scans are the same as the original ones under the null hypothesis. According to Eq. (4.1) and (4.2), \( D_{nsti} = \Phi_{n}(z_{i}) + \epsilon_{nsti} \) under the null hypothesis, we denote the totality of \( D_{nsti} \) for all \( s,t,i \), as \( D_{n} \).

Although the nonlinear anisotropic filtering used in iSPREAD is a complex procedure, it can be abstracted as a nonlinear function of \( D_{n} \). In other words, \( \hat{D}_{n} = G(D_{n}) \), where \( \hat{D}_{n} \) is the collection of filtered diffusion tensors of the \( n \)-th subject and \( G \) is the nonlinear transformation defined by the filtering procedure. Let \( \pi \) be a permutation of \( t \) and \( \pi \) the induced transformation of \( D_{n} \), we only need to show that the joint-density function of \( G(\pi(D_{n})) \) is the same as that of \( G(D_{n}) \). Obviously, this reduces to showing the joint-density function of \( \epsilon \), the totality of random errors of subject \( n \), is invariant under time-permutation because the true signal, \( \Phi_{n}(z_{i}) \), is a constant for all time points/scans under the null hypothesis. The exchangeability is guaranteed if we assume \( D_{nsti} \) follows a typical random effects model under the null hypothesis:

\[
D_{nsti} = \mu_{i} + \epsilon_{nsti} = \mu_{i} + \alpha_{n} + \gamma_{nst} + \xi_{nsti}, \tag{4.4}
\]

where \( \mu_{i} \) represents the mean of \( D_{nsti} \) and \( \epsilon_{nsti} \) is partitioned into three independent
components of random effects: the subject-specific variation $\alpha_n$, the scan-specific variation $\gamma_{nst}$ and the per-measurement error $\xi_{nsti}$. More discussions can be found in Section 2.2. in (14).

4.2.2 Overview of anisotropic diffusion (AD) filter

Nonlinear anisotropic diffusion filtering is a general scale-space approach for edge-preserving piecewise smoothing of the original image. Note that diffusion in this content refers to a diffusion process that creates a nonlinear scale space, which is different from physical process of molecular diffusion measurement in DTI. In the scale-space, an image is seen as one sample from a continuous range of scaled images with the level zero representing the original image. The linear scale-space filtering technique, first introduced by Witkin (85), involves convolving the original image with Gaussian kernels with different bandwidths to generate gradually smoother versions of the original image. This technique demonstrates the natural way of quantitatively presenting image ambiguity at different scales. However, the isotropic filtering method with a Gaussian kernel has the disadvantage of blurring object boundaries and of the suppression of fine structures at a large scale (78). The nonlinear scale space generated by nonlinear diffusion filtering, which is proposed from an analogy to thermal equations describing the diffusion process, provides an alternative way to maintain the primary properties of a scale space. The equation for the process was first presented by Perona and Malik (78), and is referred to as the Perona-Malik (PM) equation as shown in Eq. (4.5).

$$\partial_t I(z, \tilde{t}) = \text{div}(g(z, \tilde{t}) \cdot \nabla I(z, \tilde{t})) .$$  \hspace{2em} (4.5)
where function $I(z, \tilde{t})$ is taken as the image intensity (e.g. FA or MD map in our study), $\tilde{t}$ is the processing ordering, referred to as iteration steps in discrete case. The conductance function $g(z, \tilde{t})$ controls the diffusion strength, which is usually taken as a parametric function of image local gradient ($|\nabla I(z, \tilde{t})|$) given by Eq. (4.8). The original image $I(z, 0)$ is regarded as the initial state of the diffusion process. The filtered versions are from its temporal evolution, evolving toward equilibrium. The entire diffusive process prefers intra-region smoothing over inter-region smoothing with proper spatial diffusion strengths $g(z, \tilde{t})$ selected. Adiabatic boundary condition is often used to guarantee the consistency of the average grey value of the image during the filtering process (86). However, the fixed boundary condition (also called the first boundary condition) was also used and proved to have superior deblurring effects and inner edge enhancement compared with adiabatic boundary condition (79), and therefore was used in our study.

For each iteration step, the image intensity change is defined as the flow contributions from 26 neighboring pixel intensities within a $3 \times 3 \times 3$ neighborhood, and the contribution of each neighboring voxel is inversely proportional to the distance between the centroid and the corresponding neighboring voxels. The stability of the iterated processing scheme is obtained by properly adjusting the integration constant $\Delta t$ according to the $3 \times 3 \times 3$ neighborhood structure and is given in (82) by Guido Gerig et al. as in Eq. (4.6):

$$\Delta t \leq \frac{1}{1 + \sum_{i=1}^{N} \frac{1}{(\Delta d_i)^2}}$$

(4.6)

where $N$ is the number of nearest neighbors ($N = 26$ in a 3D case), and $\Delta d_i$ is the
distance between the centroid and its neighboring voxels. The proper selection of the integration constant guarantees a stable evolution of the PM equation.

Conductance function $g$ has been well studied and can take many forms, but the following function was chosen in our study because it is considered to provide better balance between smoothing-efficiency and edge preserving and have more stable performance (87,88).

$$g = \frac{1}{1 + \left( \frac{|\nabla I(\bar{x}, t)|}{\kappa} \right)^2}, \quad (4.7)$$

where $|\nabla I|$ is taken as rough edge detector and $\kappa$ is the diffusion contrast parameter. The ratio $\frac{|\nabla I(\bar{x}, t)|}{\kappa}$ controls the flow strength. The maximum flow $\partial_t I(z, t)$, is obtained when gradient $|\nabla I| = \kappa$, which represents inhomogeneous regions, and reduces to 0 when $|\nabla I| \gg \kappa$ or $|\nabla I| \ll \kappa$, which represents either potential edges or homogeneous regions. For each iteration, the gradient map was first calculated to identify the conductance function (Eq. (4.7)) and those calculated parameters were used for image smoothing. The gradient threshold $\kappa$ was estimated according to the following $p$-norm method proposed by Francesco Voci et al (87):

$$\kappa(m\Delta t) = \frac{\lambda \sigma \|I(m\Delta t)\|_p}{N_{column} \times N_{row}}, \quad (4.8)$$

where the constant $\sigma$ is proportional to the image average intensity and $\lambda$ is a trade-off parameter and was set to be 0.5 in this study. $\|I(m\Delta t)\|_p$ represents the $p$-norm of image $I$ at time step $m\Delta t$, usually $p = 2$. The columns and rows of the image is $N_{column}$ and $N_{row}$ respectively. The $p$-norm method was chosen here because it is a close estimation with a simple calculation.
4.2.3 Monte Carlo simulations

Monte Carlo simulations were performed to conduct repeated measurements comparisons of the same subject with a predefined area of simulated pathology; the effectiveness and statistical power of iSPREAD and SPREAD were evaluated and compared. For Monte Carlo simulations, synthetic DW data were generated using a DTI dataset of the brain from a healthy volunteer. The original DTI dataset was used as the template for the first group (DTI\textsubscript{pre}), which represented the baseline scan when no disease effect was presented. The template for the second group (DTI\textsubscript{post}) was generated by defining a “diseased” region in the first group with different effect sizes (es) of the largest eigenvalue ($\lambda_1$) added in each voxel. Two shapes of disease areas were considered: a $5 \times 5 \times 3$ cubic region and a $3 \times 10 \times 3$ cuboid region located at the center of the splenium of the corpus callosum were simulated to mimic real brain abnormality. In this study, es took values from $10\%$ to $50\%$. Repeated measurements of the same subject were simulated by adding Gaussian-distributed noise to both DTI\textsubscript{pre} and DTI\textsubscript{post} templates to obtain SNR $\approx 50$ in non-diffusion weighted images. The effect of thermal noise was first generated using complex random numbers with their real and imaginary parts sampled independently from a Gaussian distribution function with a zero mean and a standard deviation determined by the desired SNR level (68,69); the real parts of the complex noise signals were then added to the noise-free baseline signal $S_0$ and DW signals $S_l$. The magnitude of the final complex data was then used to synthesize the noisy DTI datasets that were further used for calculations of the noisy tensors. The magnitude of DTI\textsubscript{pre} and DTI\textsubscript{post} templates were then calculated from the envelope of the complex
signals. For each group, \( n (n = 2 \sim 5) \) repeated measurements of the same subject were simulated. A total of 100 simulations were generated for each combination of es and n.

Repeated measures comparisons were conducted using both SPREAD and iSPREAD. For each subject, the scan and time label were randomly permuted for \( N = 1000 \) times to derive a distribution of \( p \)-values for each voxel under the null hypothesis. True Positive Runs (TP_Runs) and False Positive Runs (FP_Runs), which are defined as the total number of simulations within which at least one voxel was correctly detected in the disease area or incorrectly detected in the non-disease area in those 100 simulation data, respectively, were calculated. Sensitivity and Specificity values are defined as follows:

\[
\text{Sensitivity} = \frac{\text{TP}_\text{Runs}}{\text{Total}_\text{Simulation}_\text{Runs}}
\]
\[
\text{Specificity} = 1 - \frac{\text{FP}_\text{Runs}}{\text{Total}_\text{Simulation}_\text{Runs}}, \quad (4.9)
\]

ROC curves were drawn by selecting the per-voxel \( p \)-value from 0.01 to 0.3 with an increment of 0.01.

4.2.4 In vivo human brain Data

4.2.4.1 Subjects and image acquisition

Our method was first validated on the simulated data and then applied to repeated measures comparisons of \textit{in vivo} human brain data. Five healthy volunteers and seven multiple sclerosis (MS) patients were included in this study. All participants were given written informed consent and datasets were acquired using protocols approved by the local institutional review board.
Five healthy volunteers were chosen to compare the ability for false positive control of both SPREAD and iSPREAD, when no biological changes over time were assumed. For five healthy volunteers (2 male and 3 female, mean age 22 ± 8 years, right-handed), data were scanned twice within one week using a 3T Siemens TIM Trio scanner (Erlangen, Germany). The standard protocol included a 3D axially acquired high-resolution T1-weighted MP-RAGE sequence (1 × 1 × 1 mm³), and a DTI scan. DTI data were acquired using a single-shot echo-planar (SS-EPI) sequence with a voxel resolution of 2 × 2 × 2 mm³ voxel, TR/TE = 9100/89 ms, 60 non-coplanar diffusion-encoding directions at b = 1200 s/mm² as well as 10 non-diffusion weighted images with a minimum b value of 0 s/mm².

Seven Relapsing-Remitting MS (RRMS) patients (5 male and 2 female, mean age 40 ± 8) from an ongoing longitudinal MS study were selected retrospectively to test the statistical power of both the SPREAD and iSPREAD methods to monitor the disease progression. RRMS is characterized by relapses when clearly defined symptoms of attacks occur, followed by the remission of symptoms. For each patient, there were at least two scans with an enhanced lesion visible on T1 enhanced scans. MS patients were scanned 2–4 times within a 2-year period on a GE HDX 3T scanner (Milwaukee, WI, USA). The typical MRI protocol consisted of a T2 FLAIR scan, a T1 contrast enhanced scan, a high resolution T1 SPGR scan, and a DTI scan. The SS-EPI DTI images were acquired with TR/TE = 10500/82 ms, FOV = 240 × 240 mm², acquisition matrix=128 × 128 zero-filled to matrix size = 256 × 256; slice thickness 3 mm with no gap; 24 DWIs with b = 1000 s/mm² and 4 b₀ s with minimum b value. Both the axial
T2 FLAIR and the T1 contrast enhanced scans with resolution $2 \times 2 \times 3 mm^3$ were used as image guidance for manual drawing of lesion masks used as the gold standard for evaluating automatic lesion detection by SPREAD and iSPREAD.

4.2.4.2 Steps of analysis

Data analysis was performed using MATLAB (MATLAB 2013b, The Mathworks, Natick, MA, USA) and FSL (FSL5.0.4, FMRIB Analysis Group, Oxford University, Oxford, UK). The analysis of *in vivo* human data consisted of the following five steps:

*Step 1: Data preprocessing*

The eddy_correct Tool of FSL package was first applied to DTI data to correct eddy current and motion-induced artifacts. Non-brain tissues were removed using the Brain Extraction Tool in FSL. FA and MD maps were then generated using the DTIFIT tool of FSL.

*Step 2: Image registration between different time points*

For healthy volunteers, the registration was performed between two time points. In the first step, linear registrations were performed between FA/MD images at different time points using the FLIRT tool of FSL (degree of freedom = 12). The resultant transformation matrix was saved and split into one forward and one backward half transformation. Next, one of the FA/MD maps (either *pre* or *post*) was transformed to the “halfway” using the halfway transform matrix and used as the reference image for nonlinear registration. This guarantees the equivalent interpolation bias to both time points(89). In the final step, the nonlinear registration tool of FSL (FNIRT) was used to perform the nonlinear registration.
For each MS patient, the baseline FA/MD maps were used as the reference image, and FA/MD maps at later time points were co-registered to this reference image using the FNIRT tool of FSL. The co-registered FA/MD maps were used for iSPREAD/SPREAD analyses.

**Step 3: Permutation testing**

For each subject, the scan and time labels were randomly permuted at each and every voxel for N=1000 times to generate the permutation distribution under the null hypothesis.

**Step 4: Spatial regression**

For iSPREAD, nonlinear anisotropic filtering was applied to the co-registered FA/MD maps combined with the first boundary condition, where the original FA/MD maps were used as the initial condition. The tradeoff parameter $\lambda$ in Eq. (4.8) was chosen to be 0.5 and time step $\Delta t$ was calculated based on spatial configuration of the $3 \times 3 \times 3$ neighborhood of each FA/MD map according to Eq. (4.6). In order to ensure stable evolution while save computational cost, the time step $\Delta t$ was set to the largest value possible according to Eq. (4.6). Number of iterations was fixed to 4 in all simulations. A Gaussian kernel with fixed $FWHM = 2 \times 2 \times 2$ voxels was used for SPREAD.

**Step 5: Statistical analysis for healthy volunteers and MS patients.**

The total number of false positive voxels were counted in healthy volunteers and compared between SPREAD and iSPREAD in terms of the false positive control when no biological changes were presumed. For MS patient data, to quantitatively compare both methods in their power for detecting lesion progression, lesion masks were drawn
manually based on T1 enhanced images (transformed to baseline DTI space) by an experienced radiologist and served as the gold standard. We calculated True Positive Ratio in lesions (TPR_L) and False Positive Ratios in non-lesion white matter (FPR_{NLWM}) for each subject and compared them between SPREAD and iSPREAD. The Westfall-Young method (22) was used to control the FWER.

\[
TPR_L = \frac{\text{Significant Voxel Number in Lesion Mask}}{\text{Total Voxel Number in Lesion Mask}} \times 100\%
\]

\[
FPR_{NLWM} = \frac{\text{Significant Voxel Number in NonLesion WM}}{\text{Total Voxel Number in nonLesion WM}} \times 100\%.
\]

(4.10)

4.3 Results

4.3.1 Monte Carlo simulations
(a) Effect size = 0.2, Sample Size = 3
(i) Sensitivity
(ii) Sensitivity
(iii) Sensitivity
(iv) Sensitivity

(b) Effect size = 0.2, Sample Size = 5
(i) Sensitivity
(ii) Sensitivity
(iii) Sensitivity
(iv) Sensitivity

Effect size = 0.4, Sample Size = 3
(iii) Sensitivity
(iv) Sensitivity

Effect size = 0.4, Sample Size = 5
(i) Sensitivity
(ii) Sensitivity

Legend:
- ROC of ISpread
- ROC of SPREAD
Figure 4.2. ROC analyses for the two-group comparisons using iSPREAD (red crosses) and SPREAD (blue dots) from Monte Carlo simulated data under different combinations of effect size (es = 0.2, 0.4) and sample size (n = 3, 5) of each group. The disease area was simulated as either a 5 × 5 × 3 cubic region (Fig. 4.2 (a)) or a 3 × 10 × 3 cuboid region (Fig. 4.2 (b)). Sensitivity and Specificity were calculated using Eq. (4.9). Results for FA analysis are shown here as an example. The nonlinear anisotropic diffusion filtering was used for data smoothing in the iSPREAD method and a Gaussian kernel of fixed FWHM= 2 × 2 × 2 voxels were applied for estimation of spatial regression for SPREAD.

ROC analyses for the repeated measures comparisons showed that the iSPREAD method outperformed the original SPREAD method consistently; the differences became more obvious when effect size es and group size was small, while reduced with the increasing effect size or increasing sample size. The results were also affected by the different shapes of the disease area. The iSPREAD showed much better performance and the differences between the two methods became larger when the disease area was cuboid-shaped, which further proved that nonlinear anisotropic filtering has better performance in the presence of lesions of irregular shape. Similar results were also observed for the MD analysis.
4.3.2 Healthy volunteers

Figure 4.3 Location of false positive voxels detected (with zoom in images on the right) using iSPREAD analysis in longitudinal DTI data of one healthy volunteer. False positive of FA (red dots) and MD (blue dots) for iSPREAD are mostly occurred due to image mis-registration error or BET residuals, either at tissue boundaries or brain boundaries. Figures are displayed according to radiological convention.

On average, each DTI scan collected from five healthy volunteers consists of 75000 WM/GM voxels. The average false positive voxels for FA and MD analysis in five healthy volunteers were $11 \pm 4$ and $10 \pm 12$ voxels for iSPREAD, compared to $44 \pm 30$ and $74 \pm 74$ voxels for SPREAD. While both methods were shown to control the FP voxels at a reasonable rate, iSPREAD is better at controlling FPR with fewer FP voxels for both FA and MD. Careful examinations of the location of those false positive voxels indicated that most of them appeared at brain boundaries, mainly due to brain extraction residuals or mis-registration.
4.3.3 MS patients

Seven patients with RRMS were selected for the investigation of effectiveness of iSPREAD, as well as the comparison between iSPREAD and SPREAD. The patient characteristics, findings and clinical diagnoses are summarized in Table 4.1.

Table 4.1 Patient clinical characteristics and diagnoses

<table>
<thead>
<tr>
<th>Patient</th>
<th>Age</th>
<th>Gender</th>
<th>Number of scans</th>
<th>Location of T1 post contrast enhancement lesions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient 1</td>
<td>37</td>
<td>Female</td>
<td>4</td>
<td>Active lesion around the atrium of the left lateral ventricle at 6-month follow-up.</td>
</tr>
<tr>
<td>Patient 2</td>
<td>39</td>
<td>Female</td>
<td>2</td>
<td>Periventricular white matter lesion near the right temporal parietal area, 24 months after the baseline.</td>
</tr>
<tr>
<td>Patient 3</td>
<td>34</td>
<td>Female</td>
<td>2</td>
<td>Incomplete ring enhancing lesion in left frontal subcortical white matter area at baseline. The enhancing lesion resolved after 6 months.</td>
</tr>
<tr>
<td>Patient 4</td>
<td>41</td>
<td>Female</td>
<td>4</td>
<td>Lesion in the posterior limb of the right internal capsule at baseline and resolved in the follow-up scan 3 months later.</td>
</tr>
<tr>
<td>Patient 5</td>
<td>29</td>
<td>Male</td>
<td>3</td>
<td>Two enhancing lesions located in the right temporal subcortical white matter near the right occipital horn of the right lateral ventricle which resolved in the follow-up 3-month scan.</td>
</tr>
<tr>
<td>Patient 6</td>
<td>44</td>
<td>Female</td>
<td>2</td>
<td>Lesion in right corona radiata at baseline, which resolved</td>
</tr>
</tbody>
</table>
at one month follow-up scan.

| Patient 7 | 53 | Male | 2 | Incomplete ring enhancing lesion in the right corona radiate, 7.5 months after the baseline |

The results of the iSPREAD vs. SPREAD on seven MS patients are listed in Table 4.1. Detailed results on the first three patients are shown in figures 4-6. The pairwise comparisons were conducted between the baseline scan and its first follow-up scan. The representative FA results are reported in Table 4.2.

Table 4.2 FA results for iSPREAD vs. SPREAD

<table>
<thead>
<tr>
<th>FA results</th>
<th>iSPREAD</th>
<th>SPREAD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TPR_L</td>
<td>FPR_{NLWM}</td>
</tr>
<tr>
<td>Patient 1</td>
<td>94.68%</td>
<td>1.88%</td>
</tr>
<tr>
<td>Patient 2</td>
<td>81.50%</td>
<td>0.89%</td>
</tr>
<tr>
<td>Patient 3</td>
<td>73.38%</td>
<td>0.31%</td>
</tr>
<tr>
<td>Patient 4</td>
<td>85%</td>
<td>0.2%</td>
</tr>
<tr>
<td>Patient 5</td>
<td>62.11%</td>
<td>0.86%</td>
</tr>
<tr>
<td>Patient 6</td>
<td>58.97%</td>
<td>1.8%</td>
</tr>
<tr>
<td>Patient 7</td>
<td>62.2%</td>
<td>0.8%</td>
</tr>
</tbody>
</table>

TPR_L: True Positive Ratio in lesions.

FPR_{NLWM}: False Positive Ratios in non-lesion white matter.
4.3.3.1 MS patient 1

<table>
<thead>
<tr>
<th></th>
<th>Baseline</th>
<th>6 month</th>
</tr>
</thead>
<tbody>
<tr>
<td>A.</td>
<td>T1+C</td>
<td></td>
</tr>
<tr>
<td>B.</td>
<td>T2 FLAIR</td>
<td></td>
</tr>
<tr>
<td>C.</td>
<td>FA</td>
<td></td>
</tr>
<tr>
<td>D.</td>
<td>iSPREAD</td>
<td></td>
</tr>
<tr>
<td></td>
<td>FA Results</td>
<td></td>
</tr>
<tr>
<td>E.</td>
<td>SPREAD</td>
<td></td>
</tr>
<tr>
<td></td>
<td>FA Results</td>
<td></td>
</tr>
<tr>
<td>F.</td>
<td>iSPREAD</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MD Results</td>
<td></td>
</tr>
<tr>
<td>G.</td>
<td>SPREAD</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MD Results</td>
<td></td>
</tr>
</tbody>
</table>
Figure 4.4 Comparison of significant voxels detected for FA (marked by magenta dots) and MD (marked by cyan dots) using SPREAD (Row E and G) and iSPREAD (Row D and F) analysis of longitudinal data in monitoring disease progression for patient 1. This patient had an active lesion around the atrium of the left lateral ventricle visible on the post contrast T1 images at 6-month (Row A) after baseline, which shows as hyperintensity on FLAIR image (Row B). The images are magnified in the disease area showing the detected significant voxels on the right column.

This patient had an active lesion around the atrium of the left lateral ventricle, visible on T1 enhanced images 6 months after the baseline. Voxels with significant longitudinal FA/MD changes between the baseline and 6-month are shown in Fig. 4.4. Based on the gold standard lesion mask, TPR_L values were 94.68% for FA and 90.48% for MD with iSPREAD, compared to the TPR_L values of 85.99% (FA) and 69% (MD) for SPREAD. FPR_{NLWM} values were 1.88% (FA) and 0.96% (MD) with iSPREAD in the non-lesion WM area, compared to the FPR_{NLWM} values of 1.6% (FA) and 0.8% (MD) for SPREAD. While the location of the lesion was precisely detected by both iSPREAD and SPREAD, it is clear that iSPREAD outperformed SPREAD in terms of detection sensitivity. It is worth noting that the FPR_{NLWM} for iSPREAD is slightly higher than that of the SPREAD method. Further investigation showed that false positive voxels occurred mainly at the brain boundaries due to FSL Brain Extraction Tool (BET) residuals or tissue/lesion boundaries caused by partial volume effect or atrophic changes. Since iSPREAD is more sensitive to longitudinal changes in FA/MD images, it is also prone to more FP voxels caused by image co-registration errors. This is not obvious with the healthy volunteers.
because the two scans were taken within one week using exact same imaging protocols, the atrophic changes were negligible and the co-registration errors were minor in such case. With the zoomed images showing the detected significant voxels in the lesion area on the right column of Figure 4, it is clear that the disease area was better defined with nonlinear anisotropic filtering, with clearer boundaries and fewer FP voxels around the lesion.

4.3.3.2 MS patient 2

Figure 4.5 Comparison of significant voxels detected for FA (marked by magenta dots) and MD (marked by cyan dots) using iSPREAD (Row B & C. second column) and SPREAD (Row B & C. third volume) analysis of longitudinal data during disease progression for patient 2. This patient had a lesion in the periventricular white matter near
the temporal parietal area, which shows enhancement on the post contrast T1 image (Row A) 24 months after the baseline.

This patient developed a new lesion in the periventricular white matter near the temporal parietal area 24 months after the baseline. Permutation testing was conducted from the FA/MD image at baseline vs. 24 months. TPR_L values were 81.50% (FA) and 71.68% (MD) for iSPREAD, compared to the 49.71% (FA) and 27.17% (MD) for SPREAD. The FPR_NLWM values were 0.89% (FA) and 0.22% (MD) for iSPREAD, compared to 0.43% (FA) and 0.05% (MD) for SPREAD. iSPREAD outperformed SPREAD in terms of lesion detection sensitivity while controlling FPR_NLWM at a relatively low rate.

4.3.3.3 MS patient 3
Figure 4.6 Comparison of significant voxels detected for FA (marked by magenta dots) and MD (marked by cyan dots) using iSPREAD (Row B & C. first column) and SPREAD (Row B & C. second column) analysis of longitudinal data in monitoring disease progression for patient 3. This patient had a lesion shown as a ring-enhancing lesion on T1-weighted post contrast image at baseline (Row A).

This patient had a lesion shown as a ring-enhancing lesion on T1-weighted post contrast image at baseline, which resolved 12 months after the baseline. Based on the gold standard lesion mask, TPR_L values were 73.38% (FA) and 53.24% (MD) for iSPREAD, compared to the 16.55% and 0 for SPREAD. The FPR_{NLWM} values were 0.31% (FA) 0.28 % (MD) for iSPREAD, compared to 0.2% (FA) and 0.1% (MD). It is clear that iSPREAD was more sensitive to longitudinal changes and could identify subtle changes not detectable by SPREAD. Although the FPR_{NLWM} was slightly higher for iSPREAD compared to SPREAD, it was still reasonable compared to the large voxel numbers in a whole image volume.

4.4 Discussion

In this Chapter, we proposed a method dubbed iSPREAD, designed to address potential pitfalls caused by the Gaussian kernel used in the SPREAD method. It is a non-parametric permutation-based statistical framework that combines spatial regression and resampling methods, and provides effective detection of subject-specific, localized longitudinal changes of DTI parameters within the whole brain without \textit{a priori} hypotheses. The improvements in lesion detection sensitivity introduced by anisotropic
filtering were clearly demonstrated by both Monte Carlo simulations and applications of clinical brain DTI data. The results show that the nonlinear anisotropic filtering provides better spatial regression analysis when the spatial correlation is heterogeneous among neighboring pixels. The proposed method provides an effective tool for individual-level automatic lesion detections or for quantify progression of brain abnormalities through longitudinal DTI studies, which can assist physicians in their prognoses of various diseases.

The adaptive filtering method allows for the improvement of spatial resolution thus enhanced lesion detection power. From our study, two improvements are shown in iSPREAD compared to the original SPREAD: 1) the nonlinear anisotropic filtering process provides piecewise smoothing while preserving important structures such as edges. The tissue boundaries in the filtered image are better determined, leading to better results in differentiating tissue types as well as the lesion areas; 2) As an iterative process, the heterogeneous structure between neighboring voxels are preserved at each iteration, which yields much better image enhancement results than the traditional Gaussian filtering.

The results from both simulated data and human in vivo brain data showed the improvement in sensitivity introduced by the nonlinear anisotropic filtering. Differences between iSPREAD and SPREAD were most significant when the effect size and sample size were small. The improvement in statistical power for iSPREAD was further validated by the results from MS patients for monitoring lesion progression. The FA results for iSPREAD vs. SPREAD are shown in Table 4.2., iSPREAD was able to detect
longitudinal changes in FA maps with an average TPR_L of 73.98%, compared to 44.40% for SPREAD for the seven selected MS patients. Given the high sensitivity of iSPREAD in detecting brain abnormalities, it is possible for iSPREAD to detect brain abnormalities even at an early stage of the disease, which merits further investigation.

There are also some limitations for the iSPREAD method. First and foremost, the errors caused by image preprocessing steps, namely the BET residuals and registration errors, remain a major challenge, which is also common to all voxel-based analyses of DTI data(9). Registration errors occur mostly at boundaries between different tissue types or lesion/normal tissue boundaries. Although the nonlinear anisotropic filtering method and nonlinear registrations with high degrees of freedom (DOF) were applied during the image preprocessing stage to better preserve tissue boundaries and to minimize errors from registration, mis-registration was still inevitable because of effects such as atrophic changes and lesion evolution. This is more severe for longitudinal studies that last for a long period of time. Those effects can bring uncertainties to image alignments which are driven by different tissue contrasts. It is worth noting that the $FPR_{NLWM}$ of iSPREAD (0.96% for FA) was slightly higher than that of SPREAD (0.61 % for FA) in MS patients. Further investigation showed that most of the FPs stemmed from misregistration and BET residuals. Since iSPREAD is more sensitive to longitudinal changes, it is more sensitive to those registration errors as well. These FP voxels will also cause problems when detecting very small lesions. In this case, small lesions (true positive voxels) are easily ‘submerged’ into the FP voxels that makes them difficult to be identified. More in-depth investigation of the lesion evolution over time and devising the test statistic
accordingly will bring more insight into the statistical framework and will work towards removing registration errors. Second limitation related to this study is that the parameters chosen for the anisotropic filtering were based on tests from both the simulation and clinical data and, therefore, more empirical than theoretical. As an iterative process, the number of iterations is very crucial to the result of the filtering and should be chosen based on the needs of different applications. A large number of iterations mean a strong smoothing effect, and the level of smoothing in the case of image enhancement should be used with more constraints than in the case of image segmentation (87). 3~4 times of iteration have been used for image smoothing and enhancement (90), and a larger number of iterations have been used for image segmentation (91,92). Given the fact that iSPREAD is a permutation-based method, a large number of iterations for AD filtering will certainly increase the computational burden. Our experiments show that satisfactory results can be obtained with around 4 iterations and iSPREAD took about 2 ~ 3 times of computation time of the original SPREAD method. Since the anisotropic filtering was implemented using MATLAB in the current study, a much faster performance can be expected when using C/C++.

The present work can be extended in at least two directions. First, only the scalar images (e.g. FA/MD map) were considered and the diffusivity was designed as a spatial varying scalar in the current study. However, when vector- or tensor-valued images are considered, it is more desirable to rotate the flux towards the orientation of interesting features, such as the tangential direction of structure boundaries (86). In such a case, a $3 \times 3$ adapted diffusion tensor can be used instead of the scalar diffusivity, in which the
smoothing effect along the edges is more preferred than perpendicular to it. Such a
filtering technique has been proposed and tested in (93), which proved to be more
suitable for the smoothing of bundle-like structures. With the more advanced filtering
design, similar analyses can be carried out on other quantities such as fiber tracts or
displacement from q-space imaging to help identify differences in generalized FA (GFA)
or peaks of diffusion orientation distribution function (ODF). Secondly, longitudinal
comparisons between scans at two different time points were considered and discussed
for the statistical inference test in this study. Although such repeated measures analyses
would give potentially the exact disease progression information at each time point, it is
very time consuming for longitudinal studies comprised of multiple time points. In such
cases, a voxel-wise statistical inferential test for a serial DTI (with more than two time
points) that provides general disease progression information would be preferred.
Moreover, it would be a natural extension of this work to carry other summary statistics
for investigating disease evolution and performing statistical inferential tests based on the
prior information about disease progression models.

4.5 Conclusion

A 3D nonlinear anisotropic filter was integrated into the iSPREAD method to
eliminate the potential shortcomings caused by the Gaussian kernel used in the SPREAD
method. The improvements in sensitivity and accuracy in lesion detection introduced by
anisotropic filtering were clearly demonstrated by both Monte Carlo simulation and in
vivo human brain data. The nonlinear anisotropic filtering allows for noise reduction and
reduces grey scale inhomogeneity as well as preserving important image detail, resulting in better lesion detection when the spatial correlation is heterogeneous among neighboring pixels. As a result, iSPREAD is an effective voxel-wise nonparametric method for detecting local changes in whole brain, subject-specific longitudinal DTI studies.
Chapter 5 Spatial regression analysis of serial DTI for subject-specific longitudinal changes of neurodegenerative disease

5.1 Introduction

Diffusion Tensor Imaging (DTI) (37,38), which measures the random motion of water molecules, provides a non-invasive way to investigate the structural integrity of the brain. It has been widely used in investigating white matter (WM) changes caused by brain development and aging (4), detecting abnormalities in normal-appearing WM due to disease (5), as well as identifying pathologic severity in patients with MS (94). In recent years, there has been increasing interest in the investigation of subject-specific changes within the brain without prior information regarding the spatial distribution of the pathology. Consequently, whole brain voxel-based methods (11-13) have gained much favor during recent years as an important alternative to region of interest (ROI) analysis in detecting localized changes within the brain and are most suitable when changes/effects are diffuse among individual subjects. Both parametric and nonparametric methods have been used to help identify regionally specific changes such as differences due to activation in fMRI (19), neuroanatomical differences in structure MRI data (20) and pathophysiology in longitudinal studies (21,22).
Due to the non-Gaussian nature of DTI data, nonparametric voxel-based methods that do not need any parametric assumptions such as bootstrap (26-28) and permutation-based methods (19), are more suitable. The nonparametric permutation-based method is able to devise a data-driven null distribution with only minimal assumptions, which gives the user more freedom in devising test statistics of interest. Any sensible test statistic that summarizes the local effect can be used in these hypothesis-testing procedures and the strong control of type I error is guaranteed under very mild assumptions of the null distribution. Such methods have been widely used in the area of fMRI to investigate the regionally specific effect in neuroimaging data (19). However, few of the aforementioned methods have been applied to subject-specific longitudinal studies. This is mainly because the number of available scans in a longitudinal study is often limited by practical factors such as the cost of patient recruitment, and the obtained data lacks sufficient information for a rigorous statistical inference test due to their low degrees of freedom.

The Spatial Regression Analysis of Diffusion tensor imaging (SPREAD) method previously presented (14) combines spatial regression and resampling methods, which provides a novel and efficient whole brain analysis method for detecting localized changes in subject-specific longitudinal study without an a priori hypothesis, for DTI-derived metrics such as fractional anisotropy (FA) and mean diffusivity (MD). SPREAD requires only one scan per time point for a valid statistical inferential test, which greatly reduces the granularity of permutation. The iSPREAD method (95) further improves the detection sensitivity and accuracy of SPREAD (21) substantially by incorporating a
three-dimensional (3D) nonlinear anisotropic diffusion filtering method. Both SPREAD and iSPREAD utilize a novel and effective permutation-based statistical method for whole brain analysis that relies on permuting time/scan labels and spatial kernel regression. They do not require adjustment of signal gains due to different DTI protocols at different time points and are effective for monitoring subject-specific lesion progression in longitudinal studies. However, aside from their many advantages, the following limitations exist for SPREAD / iSPREAD, which are also general to most permutation-based voxel-wise subject-specific methods applied in longitudinal studies:

1) The comparison is often taken pairwise between each time point vs. baseline, which is time consuming in the presence of serial DTI studies with multiple time points.

2) The potential differences caused by registration error or anatomical differences due to atrophic changes may manifest as false-positive voxels in the results. The consequences for such misalignment can either falsely identify positives or neglect true positives, both of which greatly reduce the statistical power and reliability of the results obtained.

3) The apparent and useful prior information of lesion progression models is largely neglected in these existing methods.

Therefore, a general statistical framework that accommodates a serial DTI study with multiple time points while taking into consideration the specific disease progression model is desired.
The main purpose of many longitudinal studies is to identify localized temporal changes within the brain. One crucial step towards detecting localized changes is to choose test statistics that are likely to be the most sensitive and informative in depicting possible departures from the null hypothesis, which assumes that there is no difference between data obtained at different time points. The statistical properties of any given hypothesis-testing procedure depend on both the null hypothesis, which specifies the distributional properties of the measurements without true signal, and the alternative hypothesis, which specifies the possible forms of true signal (temporal changes, in this case). The non-parametric permutation-based methods, such as SPREAD / iSPREAD, permit the use of a wide range of test statistics without the need to derive closed-form distributions of these statistics under the null hypothesis with specific parametric assumptions. This flexibility enables us to focus on choosing the optimum statistics based on different alternative hypotheses.

In this study, we proposed to extend the current SPREAD / iSPREAD method to a general statistical framework that accommodates a wide range of alternative hypotheses used in longitudinal studies. Five test statistics, which were divided into two major types, were implemented in the current statistical framework to help identify several different forms of temporal changes within individual subjects. One type is based on model-free test statistics; another is based on a general linear model that incorporates a certain disease evolution model. Our statistical method is similar in spirit to two well-established bodies of theories in statistical parametric maps (SPM) for assessing the regionally specific effects within the brain, namely the subtractive method (96,97) and the general
linear model (16), both of which have been widely used in the fMRI field for detection of brain activation. We use those theories within a nonparametric framework with carefully designed test statistics where the empirical null distribution is generated by permutation.

The aim of the present study was to describe and illustrate a statistical method that enables the investigation of longitudinal changes quantitatively. Simulation data with a predefined region of pathology and disease effects were used to evaluate the effectiveness of the proposed method. A series of DTI scans in three patients suffering from relapsing-remitting multiple sclerosis (RRMS) were used as human brain *in vivo* examples to demonstrate the implementation and utility of this method. Both simulations and *in vivo* results show that the proposed method is able to detect temporal changes in serial DTI with high sensitivity and accuracy. Extension of the proposed statistical framework to include other disease evolution/ drug effect models as well as different global statistics is discussed. The method is an extension of SPREAD/ iSPREAD, as well as an independent statistical framework that can be easily applied to a wide variety of longitudinal studies.
5.2 Material and methods

5.2.1 Overview of iSPREAD for serial DTI analysis

Based on the exchangeability of the time and scan labels under the null hypothesis (21), in the first step of iSPREAD analysis, the scan/time labels for FA/MD maps from each subject are randomly permuted at each voxel for $N = 1000$ times to generate a permutation distribution under the null hypothesis at each voxel. The permutated images are then smoothed using the nonlinear anisotropic filtering method for edge-preserving image enhancement, as well as for preserving spatial correlation between neighboring voxels. In the third step, various voxel test statistics are chosen to depict the temporal changes in a serial DTI analysis and will be discussed in detail in the next section. The Westfall-Young procedure (84) is used to control the FWER in the last step. Voxels are identified as significantly changing (i.e. lesion areas) if their $p$-value is less than a
predefined $p$-value (e.g. 0.05). The flowchart of the proposed framework is shown in Fig.5.1.

5.2.2 Statistical model for test statistics

Two types of voxel-wise test statistics (five test statistics in total) are used to detect temporal changes due to disease evolution in longitudinal studies. The first type (Test Statistics MF1 and MF2) uses intra-subject variation, which is a natural extension of SPREAD / iSPREAD from pairwise group comparison for two time points to a multiple group comparison for longitudinal studies that include multiple time points. The second type (Test Statistics LM, QM1, and QM2) is based on general linear model for an across-time regression analysis. Other test statistics can also be used in longitudinal studies based upon specific needs.

The following models were used to depict the lesion evolution over time: (i) Simple linear model, which assumes that FA/MD changes by an amount equal to the regression coefficient over time; (ii) Quadratic model (second-order polynomial model), which assumes a nonlinear relationship over time.

5.2.2.1 Model-free voxel-wise test statistic

This model-free test statistic detects intra-subject variation across time. In the following, FA is used as an example, but similar analysis also applies to MD.

\[ \Delta FA_i = \delta_{n,t} , \quad (5.1) \]
where $\delta_{n,i}$ can take one of the following two forms:

$$
\delta_{n,i} = \frac{1}{\sqrt{T-1}} \sqrt{\sum_{t=t_1}^{t_T} (\overline{FA}_{n.t_i} - \overline{FA}_{n.t_1})^2} \quad \text{(Test Statistic MF1)},
$$

$$
\delta_{n.i} = \frac{1}{\sqrt{T-1}} \sqrt{\sum_{t=t_1}^{t_T} (\overline{FA}_{n.t_i} - \overline{FA}_{n.t_{t_1}})^2} \quad \text{(Test Statistic MF2)},
$$

where $FA_{nst_i}$ is the $n$th subject’s $s$th scan at time $t$ measured for the $i$th voxel ($n = 1, 2, ..., N, s = 1, 2, ..., S, i = 1, 2, ..., I, t = t_1, t_2, ..., t_T), \overline{FA}_{n.t_i} = \frac{1}{S} \sum_{s=1}^{S} \overline{FA}_{nst_i}$ is the scan average and $\overline{FA}_{n.t_i}$ is the time average for each subject respectively, and $\overline{FA}_{n.t_1}$ is FA map at baseline ($t = t_1$). There are two ways to select the baseline image: (i) with Test Statistic MF1 use the mean image of all time points as the baseline image (Eq. (5.2)); (ii) with Test Statistic MF2 choose the image at the first time point as the baseline (Eq. (5.3)). Both test statistics represent localized temporal variation for FA/MD images at each voxel.

### 5.2.2.2 Voxel-wise statistics based on general linear model

All linear regression models for a response variable $Y_{ij}$ (e.g. FA/MD at certain time point) at voxel $j = 1, ... J$ can be expressed in general form (98), where a linear model with correlated errors is fitted for each individual voxel time-series $Y_{ij}$. The standard general linear model for an across-time regression analysis is shown as Eq. (5.4):

$$
Y_{ij} = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_K x_{iK} + \epsilon_{ij},
$$
where \( i = 1, \ldots, I \) indexes the observation (e.g. scans at different time points), \( \epsilon_{ij} \) is the error terms, and \( \beta_k \) are \( k^{th} \) unknown parameters for each voxel \( j \). The response variable \( Y \) at voxel \( j = 1, \ldots, J \) is expressed as a linear combination of explanatory variables \( x_{ik} \) representing the conditions under which observation \( i \) was made. In a longitudinal study, \( x_{ik} \) is typically time; however, the explanatory variables might include covariates (e.g. age) or dummy variables (e.g. gender, drug type or dosage).

In our study, two special cases of the general linear model (7) were considered: a simple linear model (Eq. (5.5)) and a quadratic model (Eq. (5.6)):

\[
Y_{ij} = \beta_0 + \beta_1 t_i + \epsilon_{ij},
\]

\[
Y_{ij} = \beta_0 + \beta_1 t_i + \beta_2 t_i^2 + \epsilon_{ij}.
\]

In both cases, the explanatory variable \( t_i \) is time (in months) and the dependent variable \( y_{ij} \) is the FA/MD value at \( j^{th} \) voxel with which the observation \( i \) (scan at the \( i^{th} \) time point) is made. Eq. (5.6) can be considered as a special case of (5.5), in which \( \beta_2 = 0 \), namely the quadratic term vanishes. Together Eq. (5.5) and Eq. (5.6) are useful in representing a large variety of realistic FA/MD changes over time in longitudinal studies.

The analysis of variance (ANOVA) \( F \)-test for the overall effect and \( t \)-test for individual covariates are the two main hypothesis-testing tools for multiple linear regressions. They can be used to generate the per-voxel \( t \)- or \( F \)-statistic map and check the significance of linear association under the assumption that the measurement error is normally distributed. Both \( t \)- and \( F \)-statistic can be viewed as signal-to-noise ratios in which the denominators serve as a clever way to eliminate the unknown nuisance
parameter $\sigma^2$, the variance of the measurement errors. In the case of permutation-based statistical methods, the distribution of measurement errors under the null hypothesis is simulated from a resampling procedure thus it is unnecessary to resort to division to eliminate any nuisance parameter of the error distribution. The following test statistics were chosen in our study.

1) Simple linear regression

The statement for the hypotheses:

$$H_0: \beta_1 = 0, \text{ versus } H_1: \beta_1 \neq 0. \quad (5.7)$$

The permutation test is carried out using the following statistic:

$$|\hat{\beta}_1| \text{ (Test Statistic LM)} \quad , \quad (5.8)$$

where $\hat{\beta}_1$ is the estimated slope value based on the standard least-squares principle. The larger this value is, the stronger the evidence to reject $H_0$.

2) Quadratic regression

If only the coefficient for the quadratic term (nonlinearity) is of interest, the statement for the hypotheses is:

$$H_0: \beta_2 = 0, \text{ versus } H_1: \beta_2 \neq 0. \quad (5.9)$$

The permutation test is carried out using the following:

$$|\hat{\beta}_2| \text{ (Test Statistic QM1)} \quad , \quad (5.10)$$

where $\hat{\beta}_2$ is the estimated regression coefficient for the quadratic term using standard least square estimates.
If the significance of the whole regression model is of interest, the statement for the hypotheses:

\[ H_0: \beta_j = 0, \text{ versus } H_1: \beta_j \neq 0, \text{ for } j = 1, 2. \]  

(5.11)

The permutation test is carried out using the following:

\[ \text{var}(\hat{y}) \quad \text{(Test Statistic QM2)} \quad , \]  

(5.12)

where \( \hat{y} = \hat{\beta}_0 + \hat{\beta}_1 t_i + \hat{\beta}_2 t_i^2 \) is the fitted value and \( \text{var}(\hat{y}) = \sum_{i=1}^{N} (\hat{y} - \bar{y})^2 / N - 1 \) is the variance of the fitted value \( \hat{y} \).

The five test statistics mentioned above were calculated at each voxel to form the statistic images. Permutation and spatial regression were used to construct the null distribution of these statistic images. Significant voxels were identified by comparing the original (un-permuted) statistic images to their permutation counterparts. Because all five test statistics are larger under \( H_1 \), the permutation \( p \)-value at a given voxel is defined as the proportion of permutation-generated test statistic that is \textit{larger} than the un-permuted version. We then applied a suitable multiple testing procedure such as Westfall-Young procedure (99) and the Benjamini Hochberg procedure (100) to these \( p \)-value maps to control for familywise error rate and false discovery rate, respectively. Technical details about the resampling procedure and multiple testing adjustment can be found in (21).
5.2.3 Monte Carlo simulation

Effectiveness and statistical power of the proposed method were first validated by a Monte Carlo (MC) simulation of group comparisons of repeated measurements of the same subject with a predefined simulated disease area in serial DTIs. Simulations were performed in a similar manner as previously described in (21). Instead of performing a two-group comparison, a multiple-group comparison was simulated to mimic the disease progression in a longitudinal study. Lesions were simulated as a $5 \times 5 \times 3$ cubic region at the center of the splenium corpus callosum with different effect sizes (es) of the largest DTI eigenvalue ($\lambda_1$) added in each voxel to imitate a real brain abnormality. Values of es ranged from 10~50%. Repeated measurements of the same subject were simulated by adding Gaussian-distributed noise to both DTI\textsubscript{pre} and DTI\textsubscript{post} templates to achieve signal-to-noise ratio (SNR) $\approx 50$ in non-diffusion weighted images. The effect of thermal noise was first generated using complex random numbers with their real and imaginary parts sampled independently from a Gaussian distribution function with a mean of zero and a standard deviation determined by the desired SNR level (68,69); the real parts of the complex noise signals were then added to the noise-free baseline signal $S_0$ and DW signals $S_i$. The magnitude of the final complex data was then used to synthesize the noisy DTI datasets that were further used for calculations of the noisy tensors. The magnitude of DTI\textsubscript{pre} and DTI\textsubscript{post} templates were then calculated from the envelope of the complex signals. Numbers of repeated measurements were chosen from $n = 2$~5 for each group in the simulation, and a total of 100 simulations were generated for each combination of es and n.
Both a linear trend and a nonlinear trend of disease progression models over time were considered, two models (Effect size on $\lambda_1$ vs. Timepoints) are illustrated in Fig. 5.2. The linear and nonlinear trends on the largest eigenvalue $\lambda_1$ also led to linear and nonlinear trends on tensor-derived parameters such as FA/MD.

(i) Simple linear model: implies that white matter changes by an amount equal to the regression coefficient over time. This type of model in a longitudinal study requires at least three time points. For MC simulation, the following effect sizes were chosen at each time point to mimic a linear relationship across time: $es_1 = 0, es_2 = 0.02, es_3 = 0.05, es_4 = 0.08, es_5 = 0.15$.

(ii) Quadratic model: implies a nonlinear relationship across time. This type of model in a longitudinal study requires at least four time points. For MC simulation, following effect sizes were chosen at each time point to mimic a nonlinear relationship across time: $es_1 = 0, es_2 = 0.5, es_3 = 0.2, es_4 = 0.1, es_5 = 0.05$, which represent different stages of the disease such as onset, peak and recovery.
Figure 5.2 Illustration of linear and quadratic simulated models (Effect size on $\lambda_1$ vs. Timepoints)

Group comparisons were conducted using iSPREAD with the five proposed voxel-wise test statistics. True Positive Runs (TP_Runs) are defined as the total number of simulations within which at least one voxel was correctly detected in the disease region. False Positive Runs (FP_Runs) are defined as the total number of simulations within which at least one voxel was incorrectly detected in the non-disease region. Sensitivity and Specificity values are defined based on TP_Runs and FP_Runs as follows:

$$\text{Sensitivity} = \frac{\text{TP}_\text{Runs}}{\text{Total}_\text{Simulation}_\text{Runs}},$$

$$\text{Specificity} = 1 - \frac{\text{FP}_\text{Runs}}{\text{Total}_\text{Simulation}_\text{Runs}}, \quad (5.13)$$

ROC curves were drawn by selecting the per-voxel p value from 0.01 to 0.3 with a gradual increment of 0.01. Results were compared between the five proposed test statistics.
5.2.4 Human in vivo brain data

5.2.4.1 Subjects and image acquisition

Multiple sclerosis (MS) is an inflammatory disease of the central nervous system that is thought to be a cell-mediated autoimmune disease leading to progressive neurologic dysfunction; it is classified into different categories that include relapsing remitting, progressive, and stable (5). Relapsing-Remitting MS (RRMS) is characterized by clearly defined occurrence of symptoms followed by the remission of symptoms.

In our study, three RRMS patients (2 male and 1 female, mean age 42 ± 2) from an ongoing longitudinal MS study were used as subjects. All subjects were given written informed consent and datasets were acquired using protocols approved by the local institutional review board.

Patients were scanned at baseline and every 3 to 6 months afterward for a duration of 2 years. Images were acquired using a GE HDX 3T scanner (Milwaukee, WI, USA). The Magnetic Resonance (MR) protocol consisted of: (1) a T2 FLAIR scan; (2) a T1 contrast enhanced scan; (3) a high resolution T1 SPGR scan, and (4) a DTI scan. Single-shot echo planar diffusion-weighted imaging (SS-EPI) was acquired with the following parameters: $TR/TE = 10500/82ms$ ; $FOV = 240 \times 240 \ mm^2$ ; acquisition matrix=128 × 128, zero-filled to matrix size = 256 × 256; slice thickness 3$mm$ with no gap; 24 $DWIs$ with $b = 1000s/mm^2$ and 4 $b_0$s with no diffusion weighting. Both the axial T2 FLAIR and T1 contrast enhanced scans with resolution $2 \times 2 \times 3mm^3$ were used for anatomic definition in the examination.
5.2.4.2 Analysis

The data were analyzed using MATLAB (MATLAB 2013b, The Mathworks, Natick, MA, USA) and FSL (FSL5.0.4, FMRIB Analysis Group, Oxford University, Oxford, UK). The FSL package’s eddy_correct tool was used to correct eddy current and motion-induced artifacts in the DTI data. The non-brain tissues were deleted using the BET tool in FSL; FA and MD maps were generated subsequently using DTIFIT tool in FSL. To facilitate the voxel-wise comparison at each time point, FA/MD maps at later time points were first co-registered to the FA/MD images at baseline using the FNIRT tool in FSL. The co-registered FA/MD maps were then averaged, and these averaged maps were used as the subject-specific templates to avoid asymmetry-induced bias in image processing(101). The FA/MD maps at different time points were then registered to the template using the FLIRT tool of FSL. The final co-registered FA/MD maps were used for the iSPREAD analysis.

For each subject, the scan and time labels were randomly permuted at each and every voxel for N=1000 times to generate the permutation distribution under null hypothesis. Five voxel test statistics proposed above were used to model changes in FA/MD over time. For the general linear model, both simple linear and quadratic fits were used to describe the lesion progression over time using FA/MD values as the dependent variable. The Westfall-Young method (84) was used to control the FWER. Both True Positive Ratio in lesions (TPRL) and False Positive Ratios in non-lesion white matter (FPRNLWM) were calculated for each test statistics to quantify the sensitivity and specificity in lesion detection, which were defined as in Eq. (5.14).
\[
TPR_L = \frac{\text{Significant Voxel Number in Lesion Mask}}{\text{Total Voxel Number in Lesion Mask}} \times 100\%
\]
\[
FPR_{NLWM} = \frac{\text{Significant Voxel Number in NonLesion WM}}{\text{Total Voxel Number in nonLesion WM}} \times 100\%, \quad (5.14)
\]

5.3 Results

5.3.1 Monte Carlo simulation

Figure 5.3 ROC analysis for group comparisons using iSPREAD with five proposed test statistics to identify localized changes from Monte Carlo simulated data. Disease
progression models were simulated by either (a, b) a simple linear model \((e_{s1} = 0, e_{s2} = 0.03, e_{s3} = 0.05, e_{s4} = 0.08, e_{s5} = 0.15)\) or (c, d) a quadratic model \((e_{s1} = 0, e_{s2} = 0.5, e_{s3} = 0.2, e_{s4} = 0.1, e_{s5} = 0.05)\). Sample sizes of \(n = 3\) and \(n = 5\) were considered. The disease area was simulated as a \(5 \times 5 \times 3\) cubic region at the center of splenium corpus callosum. Sensitivity and Specificity were calculated using Eq.(5.13).

Results for FA analysis are shown here as an example.

From the MC simulations, it is clear that a high sensitivity can be obtained by all test statistics for a simple linear trend (Fig. 5.3 (a) and (b)) with the exception of Test Statistic QM1, which tested the significance of a quadratic trend. Test Statistic LM, which was the estimated slope in a simple linear regression model, yielded the highest statistical power. As a side note, while the quadratic model (Eq. (5.6)) includes the linear model (Eq. (5.5)) as a special case, Test Statistic QM2 also summarizes the variance explained by the linear component in Eq. (5.6). However, Test Statistic QM2 is not as “sharp” as Test Statistic LM because Eq. (5.6) has one more unknown parameter (the quadratic term) to estimate, therefore it is not as efficient as the simple linear model.

As for the quadratic model (Fig. 5.3 (c) and (d)), a high sensitivity can be obtained by all test statistics except for Test Statistic LM, which was the estimated slope for simple linear regression. Both Test Statistic QM1 and QM2, which were designed specifically for the quadratic model, yielded slightly superior performance than the other test statistics. The differences in results between using different test statistics were reduced with the increasing sample size and effect size of lesions. For instance, a disease progressed with a fast rate (e.g. with a relatively large effect size change at each time point) will yield a
higher detection sensitivity than a disease progressed with a slow rate (e.g. with a relatively small effect size change at each time point).

5.3.2 Human in vivo data

5.3.2.1 MS patient 1

This patient had an active lesion in the posterior limb of the right internal capsule visible on the post contrast T1 images at baseline and fading over time in the follow-up scans.

Figure 5.4 This patient had an active lesion (red arrows) in the posterior limb of the right internal capsule visible on the post contrast T1 images at baseline and resolved in the follow-up scans. The degree of hyperintensity within the lesion lessened after 3 month and 4 month compared to the baseline scan.
Figure 5.5 Comparison of significant voxels detected for FA (magenta dots) using iSPREAD with five proposed test statistics (Fig. 5.5 (a)). Mean FA values calculated from gold standard ROIs at different time points were drawn in Fig. 5.5 (b), together with a fitted simple linear model and a quadratic regression model for mean FAs over time.

Table 5.1 $\text{TPR}_L$ and $\text{FPR}_{\text{NLWM}}$ calculated using each test statistic for MS patient 1

<table>
<thead>
<tr>
<th>Test Statistics</th>
<th>MF1</th>
<th>MF2</th>
<th>LM</th>
<th>QM1</th>
<th>QM2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{TPR}_L$</td>
<td>88.01%</td>
<td>97.62%</td>
<td>83.33%</td>
<td>83.33%</td>
<td>88.10%</td>
</tr>
<tr>
<td>$\text{FPR}_{\text{NLWM}}$</td>
<td>0.017%</td>
<td>0.30%</td>
<td>0.003%</td>
<td>0.004%</td>
<td>0.02%</td>
</tr>
</tbody>
</table>
All five proposed test statistics were able to achieve an average $\text{TPR}_L$ of 88.10% and $\text{FPR}_{\text{NLWM}}$ of 0.07%, with a slightly decreased sensitivity achieved by Test Statistic QM1. While the model-free test statistics (Test Statistic MF1 and Test Statistic MF2) were able to control the FWER at a reasonable rate, with an average $\text{FPR}_{\text{NLWM}}$ of 0.16%, it is clear that test statistics based on a general linear model (Test Statistics LM, QM1 and QM2) were able to control the false positive even better, with an average $\text{FPR}_{\text{NLWM}}$ of 0.009%. FP voxels due to mis-registration and atrophy were largely suppressed. This was not obvious with simulated data because there were no such effect as mis-registration and atrophy that will potentially cause the FPs. Due to limitation of the proposed linear model, which could only accounts for majority of the variations within the disease area across time, a slight decrease in $\text{TPR}_L$ (84.92% vs. 92.86%) were observed when using test statistics based on a general linear model compared to the model-free test statistics.

5.3.2.2 MS patient 2

This patient had an active lesion around the atrium of the left lateral ventricle visible on the post contrast T1 images at 6 month from baseline and resolved later.
Figure 5.6 This patient had an active lesion around the atrium of the left lateral ventricle visible on the post contrast T1 images at 6 month (red arrows) from baseline and resolved late.

Table 5.2 TPR_L and FPR_NLWM calculated using each test statistic for MS patient 2

<table>
<thead>
<tr>
<th>Test Statistics</th>
<th>MF1</th>
<th>MF2</th>
<th>LM</th>
<th>QM1</th>
<th>QM2</th>
</tr>
</thead>
<tbody>
<tr>
<td>TPR_L</td>
<td>85.06%</td>
<td>96.10%</td>
<td>75.65%</td>
<td>69.16%</td>
<td>77.92%</td>
</tr>
<tr>
<td>FPR_NLWM</td>
<td>0.14%</td>
<td>4.50%</td>
<td>0.07%</td>
<td>0.06%</td>
<td>0.09%</td>
</tr>
</tbody>
</table>
It can be seen from the results that while all test statistics could achieve an average TPR_L of 80.78%, the test statistics based on a linear regression model yielded much lower FPs (0.07% vs. 2.32%), and had a higher ability to control the FWER. Specifically, the average TPR_L obtained by the model-free test statistics was 90.58% and the average FPR_{NLWM} was 2.32% compared to an average TPR_L of 74.24% and an average FPR_{NLWM} of 0.07% obtained by the test statistics based on a linear regression model.

5.3.2.3 MS patient 3

This patient has three lesions located in two different slices; results from iSPREAD are shown in Figs. 5.8-5.9 for one slice and Figs. 5.10-5.11 for the other slice.

Figure 5.8 This patient had two new gadolinium enhancing lesions visible in this slice on the post contrast T1 images at baseline and resolved in the follow-up 3-month scan and 6-month scan. One subtle peripheral enhancing lesion (blue arrows) was within the medial aspect of the occipital horn of left lateral ventricle in the left posterior temporal area. Another one lesion (red arrows) was located in the right subcortical temporal area (near
the occipital horn of right lateral ventricle). Both lesions demonstrate hyperintensity on FLAIR image at the baseline, with both the size of the lesions and degree of FLAIR hyperintensity gradually lessened in the follow-up scans, which suggested a stable state.

![Figure 5.9 Significant voxels detected for FA (marked by magenta dots) using iSPREAD with three proposed test statistics. TPR_L and FPR_NLWM calculated using each test statistic by Eq. (5.14) are listed in Table 3.](image)

Table 5.3 TPR_L and FPR_NLWM calculated using each test statistic on the first slice

<table>
<thead>
<tr>
<th>Test Statistics</th>
<th>MF1</th>
<th>MF2</th>
<th>LM</th>
</tr>
</thead>
<tbody>
<tr>
<td>TPR_L</td>
<td>90.71%</td>
<td>90.71%</td>
<td>82.86%</td>
</tr>
<tr>
<td>FPR_NLWM</td>
<td>2.31%</td>
<td>2.35%</td>
<td>1.81%</td>
</tr>
</tbody>
</table>
Figure 5.10 This patient had one new gadolinium enhancing lesions visible in this slice on the post contrast T1 images at baseline (first column, marked by red arrow) and resolved in the follow-up scans. The enhancing lesion was located in the right splenium of the corpus callosum extending to the right major forceps area and shown as hyperintensity on the FLAIR image at baseline, and gradually lessened both in size and intensity in the follow-up scans.

![Image of enhancing lesions](image)

Figure 5.11 Significant voxels detected for FA (marked by magenta dots) using iSPREAD with three proposed test statistics. TPR\textsubscript{L} and FPR\textsubscript{NLWM} calculated using each test statistic by Eq. (5.14) are listed in Table 5.4.

<table>
<thead>
<tr>
<th>Test Statistics</th>
<th>MF1</th>
<th>MF2</th>
<th>LM</th>
</tr>
</thead>
<tbody>
<tr>
<td>TPR\textsubscript{L}</td>
<td>94.09%</td>
<td>94.09%</td>
<td>91.40%</td>
</tr>
<tr>
<td>FPR\textsubscript{NLWM}</td>
<td>2.31%</td>
<td>2.35%</td>
<td>1.81%</td>
</tr>
</tbody>
</table>

iSPREAD was able to detect all lesions with a very high sensitivity. Compared to the gold standard lesion masks, lesions were detected with an average TPR\textsubscript{L} of 88.10% and FPR\textsubscript{NLWM} of 2.15% for the two lesions on one slice in Fig. 5.9; an average TPR\textsubscript{L} of
93.19\% and FPR_{NLWM} of 2.15\% for the lesion on the other slice in Fig. 5.11. Since only three scans available in this longitudinal study, only the model-free Test Statistics MF1, MF2 and Test Statistic LM (based on a simple linear model) were applied to test the localized temporal changes.

5.4 Discussion

In this study, a nonparametric statistical method comprised of two types of voxel-based statistics that helps in detecting subject-specific longitudinal changes in a serial DTI data is presented. Specifically, both the model-free voxel-based test statistics and the test statistics based on a general linear model were applied in the proposed statistical method to test against the null hypothesis of zero differences occurring between groups across time. While a high sensitivity and accuracy was obtained by the five proposed specific voxel-wise statistics belonging to the two types respectively, a significant improvement in specificity while including the prior information of the disease evolution. This indicates the possibility of differentiating the relative contributions of anatomical differences due to image mis-registration, atrophy (False positives) and differences in tissue compositions within a presumably homogeneous structure (True positives, e.g. lesion, tumor, normal appearing white matter). This statistical method is an extension of previously presented SPREAD/iSPREAD method, as well as an independent statistical framework that can be applied to a variety of longitudinal studies using carefully designed test statistics to help detect subject-specific local changes within the brain.
The nonparametric permutation-based methods are very suitable for analysis of data with low degrees of freedom to avoid noisy statistic images (19). However, aside from their numerous applications in the field of fMRI, their applications in longitudinal studies are sparse. Two major disadvantages of permutation-based methods prohibit their wide use; one is the computation burden they impose, the other is the need for sufficient scans in an experiment to give an enough number of possible labelings for generating the empirical null distribution (102). While the first one is partially solved by the fast developing power computing technology, the second one can be addressed by the previously proposed SPREAD/iSPREAD method.

In most cases, the motivation that drives a longitudinal study is the alternative hypothesis that there exist temporal changes between scans at different time points; the problem then comes to which test statistic is most informative in terms of identifying localized changes. Two types of test statistics were herein proposed and tested in this study. The first type, which measures the intra-subject variability across time, is a natural extension of the test statistic used in iSPREAD for pairwise comparison. As a model-free test statistic, it requires no prior information about the disease evolution. However, such test statistics are very sensitive to registration errors or anatomical differences between the scans, therefore were not able to distinguish between significant voxels caused by pathological changes (True Positives) and anatomical differences due to atrophy or image registration error (False positives). The general linear model used in the second type of test statistics is similar in spirit to those used in parametric methods, but without the need for a parametric assumption of measurement errors. For the general linear model, it is
very crucial to devise the design matrix according to the specific longitudinal study to reflect the disease evolution model. Both the simple linear model and quadratic model were chosen in this study to account for the temporal effects of the lesion progression. The models chosen in this study were simple, intuitive and easy to implement, which can depict a large number of white matter integrity changes in longitudinal studies.

The effectiveness of the proposed statistical method was validated by both simulation data and human in vivo brain data. For simulation data, while model-free test statistics (Test Statistic MF1 and MF2) were able to detect changes in lesion progressions with a high sensitivity and accuracy, Test Statistic LM, QM1 and QM2 were most useful when the lesion progression followed a specific model. Moreover, it is clear that when the disease progression model followed a specific model, the test statistics designed accordingly would have a higher statistical power than model-free test statistics. Differences between different test statistics were most significant when the effect size and sample size were small, and reduced with an increasing sample size and effect size.

For the MS patients chosen in this study, take the first two MS patient as examples, while all five test statistics could detect longitudinal changes in lesion evolution with an average TPR_L of 87.31%, test statistics based on a linear model had the ability to control the FWER at a relatively low rate (with a FPR_{NLWM} of ~ 0.04%) compared to an average FPR_{NLWM} of ~ 1.24% obtained from model-free test statistics. FP voxels due to mis-registration, atrophy, partial volume and random scan variations were largely suppressed when the lesion progression model could be specified explicitly. Compared to the model-free test statistics, the advantage of integrating the specific lesion progression model was
clear. Unlike typical independent and identically distributed (i.i.d) Gaussian noise, the effect of underlying anatomic differences caused by mis-registration, atrophy, etc., are spatially smooth thus will not be filtered out by spatial regression. These artifacts increase the total variation and drive up model-free test statistics, which in turn inflate type I error. However, these artifacts are in general specific to each scan and do not follow any specific disease progression model. By imposing a specific model of a true signal, these artifacts will be largely cancelled out in the temporal direction, which reduces the type I error as a result.

It is worth pointing that for the two model-free test statistics, Test Statistic MF2 uses the image at the first time point as the baseline image while the Test Statistic MF1 uses the time average image for this purpose. Averaging images in the temporal direction reduces variability, therefore the within-group variation measured by Test statistic MF2 is larger than Test statistic MF1, which can inflate the type I error (false positives) obtained from Test statistic MF2 as compared to that from Test statistic MF1. This is the reason why there were usually more FPs in the results calculated from Test Statistic MF2 than that of Test Statistic MF1 for MS patients.

Despite its advantage of better FP control with a high accuracy with the general linear model, there are two major limitations with its use. First and foremost, in order to devise the design matrix in the general linear model for a specific study, prior information about the disease progression is desirable. This is similar to the general linear model used in fMRI; the design matrix should be devised according to the experiment design. Moreover, in order to accurately model the disease progression, an accurate model should be
considered. Both the simple linear model and quadratic model are rough approximation of the temporal changes in FA/MD. The consequence of a lack of fit is usually a decrease in detection sensitivity. As showed in the results, due to limitations of the proposed linear model, which could only account for the majority of the temporal variations within the disease area across time but not all, a slightly decreased sensitivity were obtained by test statistics using a general linear model compared to the model-free test statistics, with a TPR_L of 92.17% for the model-free test statistics versus a TPR_L of 83.35% for the test statistics based on a general linear model. A more precise and robust time series regression model is therefore needed to accurately examine the localized temporal changes in a longitudinal study and the test statistics could be designed accordingly. However, this is not easy for a voxel-based method as the temporal change trends may vary within disease area. Fortunately, there are many previous fMRI studies we can refer to (16,103). Moreover, when a specific drug effect/disease evolution is of interest; such effect can be easily detected with corresponding design matrix, avoiding the disruption of other “unwanted” signals. The second limitation lies in situations when dealing with data of low degrees of freedom. Although the iSPREAD method greatly reduces the granularity of permutation and makes the permutation-based method feasible with data of low degrees of freedom, the multiple regression still needs the number of data points $N > k + 1$, where $k$ is the number of variables. Since the degrees of freedom for a multiple regression is $N - k - 1$, as the number of data points decreases, the ability to test the model erodes accordingly. Therefore, the simple linear model requires at least three scans, one at each time point and even more scans are needed for a quadratic model. However,
neither of these limitations mentioned above are specific to this approach, but rather are caused by the inherent characteristic of the methodology.

The current statistical framework can be improved in at least two aspects. In iSPREAD, a nonlinear diffusion filtering method was used to preserve the correlation of the heterogeneous spatial structure and greatly improve the statistical power. For a longitudinal study, the data can be seen as 4D, with the fourth dimension being the time series. In that sense, a 4D filtering method can be applied to preserve the correlation between neighboring voxels in the temporal direction, which is expected to further promote the statistical power of the method. However, in the case of longitudinal study with low degrees of freedom (with few time points available), this needs more careful consideration.

The proposed statistical method provides a novel and simple way of depicting subject-specific localized temporal changes, and is especially suitable for datasets with low degrees of freedom. However, the application for this study is not restricted to investigating lesion evolution, but also for detecting longitudinal changes due to aging, drug effects or other neurodegenerative diseases. The proposed statistical method can be easily generalized to include a broad spectrum of longitudinal studies as well as readily applied to other imaging applications such as fMRI when the degrees of freedom of data are small. It is also a natural extension of this study to include more test statistics according to the different lesion evolution models. Although the voxel-wise test statistics tend to be more sensitive and appropriate to identify focal differences (e.g. lesion, tumor) between groups, global summary statistics are usually used to identify diffuse changes.
between groups (19). Therefore, different choices of global test statistics such as different norms can be used in the presence of diffuse changes within the brain (e.g. mild traumatic brain injury) to give the researchers information regarding the severity of the brain injury, which merits further investigation.

5.5 Conclusion

We have presented a permutation-based voxel-wise whole brain analysis method as an extension of SPREAD/ iSPREAD that facilitates detecting longitudinal changes in serial DTI studies. This method can obtain a high statistical power even with limited scans available, and thus is very suitable for longitudinal studies with low degrees of freedom. The method described is shown to be accurate and able to achieve a high sensitivity while controlling FWER at a relatively low rate when the test statistic was designed according to the disease progression model being investigated. The proposed framework can be easily generalized to accommodate a variety of longitudinal studies using carefully designed test statistics, which will greatly increase the scope of analysis methods available for longitudinal studies at an individual level.
Chapter 6 Summary and future work

6.1 Summary

The work presented in this dissertation is intended to address two major challenges in DTI: (1) image quality control and (2) its application in subject-specific longitudinal DTI study. The first part of this dissertation focuses on DTI quality control. Although many Quality Control (QC) software tools are being developed and are widely used and each has its different tradeoffs, there is still no general agreement on an image quality control routine for DTIs and the practical impact of these tradeoffs is not well studied. An objective comparison that identifies the pros and cons of each of the QC tools is helpful for the users to make the best choice among tools for specific DTI applications. In Chapter 3, we presented a quantification protocol to evaluate three popular QC tools (TORTOISE, DTIstudio, DTIprep). Both simulation data and human in vivo data were used to compare quantitatively how major DTI artifacts, as well as the ratio and distribution of different artifacts, affected the output accuracy and precision of tensor calculation for three popular QC tools. The results of our study show that each tool had its own pros and cons. While the performances of all three QC tools were comparable - able to identify and compensate the outlier well when the data corrupted ratio was relatively low, some of them became less stable with increasing data corrupted ratio. Featured algorithms from each QC tools and the ways in which particular algorithm affects the output of each of the tools were also analyzed. The compatibility and integration of them with other popular image processing tools were discussed to help DTI
users to choose suitable QC tools for specific study as well as to build a general DTI processing pipeline.

Quantitative measurement of localized longitudinal changes in brain abnormalities at individual level may offer critical information for disease diagnosis and treatment. The second part of this dissertation focuses on nonparametric algorithm design for lesion detection in longitudinal DTI studies. In Chapter 4, we proposed a voxel-wise permutation-based method iSPREAD for whole brain longitudinal studies of local DTI changes or lesion evolution. iSPREAD improved the original SPREAD method by incorporating 3D nonlinear anisotropic filtering, which provided edge preserving noise reduction and led to better lesion detection. The improvement in lesion detection sensitivity and accuracy was well demonstrated by both ROC analysis for simulation data and clinical DTI data from MS patients. iSPREAD is able to achieve sufficient statistical power even with limited number of scans, and is an effective voxel-wise nonparametric method for detecting and monitoring local changes in whole brain longitudinal DTI studies at individual level.

In Chapter 5, a statistical method based on iSPREAD was presented to facilitate detection of longitudinal changes in serial DTI studies. Two types of voxel-wise test statistics (a model-free test statistic, which measures intra-subject variability, across time, and test statistics based on general linear model that incorporate specific lesion evolution models) were proposed. The implementation and utility of the proposed statistical framework were demonstrated by both Monte Carlo simulations and applications on clinical DTI data from human brain in vivo. By a design of test statistics based on disease
progression model, it was possible to apportion the true significant voxels that are attributed to the disease progression and those caused by underlying anatomical differences that cannot be explained by the model, which led to improvement in false positive (FP) control in the results. The proposed method could achieve sufficient lesion detection sensitivity and accuracy while controlling FWER at a relatively low rate when the test statistic was designed based on the disease progression model. The proposed framework can be applied to a variety of longitudinal studies by using carefully designed test statistics, which will greatly increase the scope of analysis methods available for subject-specific longitudinal studies.

6.2 Future work

In this dissertation, we propose iSPREAD to help in detecting of localized longitudinal diffusion changes within the whole brain at individual-level with improved accuracy and sensitivity compared to the original SPREAD. As an extension of iSPREAD, we also propose a statistical method that facilitates analysis of spatial regression analysis of serial DTI for subject-specific longitudinal changes of neurodegenerative disease. This statistical method is especially suitable for longitudinal datasets with low degrees of freedom, and can be easily applied to a wide variety of longitudinal studies to help detect localized changes at individual level with carefully designed test statistics.

Although the statistical method is complete in itself, as discussed before in Chapter 4 and Chapter 5, there still remains room for improvement of the present method.

- Vector- or tensor-valued images can be considered in addition to FA/MD maps.
In such case, a $3 \times 3$ adapted diffusion tensor can be used instead of the scalar diffusivity in the design of anisotropic filtering, in which the smoothing effect along the edges is more preferred than perpendicular to it. With the more advanced filtering design, similar analyses can be carried out on other quantities such as fiber tracts or displacement from q-space imaging to help identify differences in generalized FA (GFA) or peaks of diffusion orientation diffusion function (ODF).

- Data collected in a longitudinal study or an fMRI study can be seen as 4D, with the fourth dimension being the time series. A 4D filtering method can therefore be applied to preserve the correlation between neighboring voxels in time series, which is expected to further promote the statistical power of the method.

- The application for this study is not restricted to investigating lesion evolution, but is also for detecting longitudinal changes due to aging, drug effects or other neurodegenerative diseases. The proposed statistical framework can be easily generalized to include a broad spectrum of longitudinal studies as well as readily applied to fMRI when the degrees of freedom of data are small.

- It is also a natural extension of iSPREAD to include more test statistics according to the different lesion evolution models.

- Different choices of global test statistics such as different norms can be used in the presence of diffuse changes within the brain (e.g. mild traumatic brain injury) to give the researchers information regarding the severity of the brain
injury, which merits further investigation.

Currently, we are exploring the use of iSPREAD to help identify the diffuse changes within the brain such as mild traumatic brain injury using global test statistics.
References


55. Sotak CH. The role of diffusion tensor imaging in the evaluation of ischemic brain injury - a review. NMR in Biomedicine 2002;15(7-8):561-569.


