

Assessment of Image Quality Requirements in Magnetic Resonance Imaging for Quantitative Brain Morphometry

THÈSE N° 7193 (2016)

PRÉSENTÉE LE 9 DÉCEMBRE 2016

À LA FACULTÉ DES SCIENCES ET TECHNIQUES DE L'INGÉNIEUR
LABORATOIRE DE TRAITEMENT DES SIGNAUX 5
PROGRAMME DOCTORAL EN GÉNIE ÉLECTRIQUE

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Pavel FALKOVSKIY

acceptée sur proposition du jury:

Dr G. Boero, président du jury
Prof. J.-Ph. Thiran, Dr G. Krüger, directeurs de thèse
Prof. T. Stoecker, rapporteur
Prof. E.-W. Radue, rapporteur
Dr D. Gallichan, rapporteur



ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Suisse
2016

To consult the statistician after an experiment is finished is often merely to ask him to conduct a post mortem examination. He can perhaps say what the experiment died of.

— Ronald Fisher

To my parents and Sveta...

Acknowledgements

Over the past 4 years I have learned a great deal and met a lot of great people. I would like to thank Gunnar Krueger and Jean-Philippe Thiran for giving me the opportunity to do my Ph.D. at EPFL, being supportive, and guiding me through many research problems. I truly value the experience that I had at EPFL. I would like to thank all my colleagues that I had a pleasure to work with. Alexis Roche, for always being there when I needed help with mathematics, image processing and everything else. Your rigorous and pragmatic approach to complicated problems gave me a lot of inspiration and it is something that I look up to. Tobias Kober, for helping me with IDEA programming, showing me how to operate scanner, providing me with guidance, and helping me not to lose my focus. It was good fun going to CHUV for the scanning for two sessions. I truly value it. Maryna Waszak, for being a good friend and a great colleague. It has been a lot of fun optimizing the imaging navigator to work towards motion correction problems. Mario Fartaria de Oliveira, for the discussions during the sleepless nights before deadlines. It has been a real pleasure knowing you as a friend and a colleague. Davide Piccini for teaching me the essential expressions in Italian language. Benedicte Marechal, for being a good neighbour and a great colleague. It was a lot of fun doing the reproducibility studies with you. Tom Hilbert for sharing the desk with me and always willing to hear my rants about my trials and tribulations. Guillaume Bonnier, Tom Hilbert, Mario Fartaria de Oliveira and Theo Galy-Fajou for epic Fridays. Emilie Mussard for helping me to learn the Swiss French. Daniel Brenner for data collection marathons and showing me around Bonn. It was a pleasure working with you. Michael Amann and Alain Thoni for brightening up the weekends collecting data in Basel. David Romascano, Daniel Schmitter, Michael Rieger, Tania Bacoyannis, Jean-Marc Jaeger, Ruslan Aydarkhanov, Robin Demesmaeker, Thomas King, Christina Granziera, and Kieran O'Brien for great atmosphere in the office. I would like to thank all of the colleagues at CIBM, and last but not least I would like to thank my wife Sveta and my parents for putting up with me writing my Ph.D. for the past 4 years.

Abstract

Structural T1-weighted magnetic resonance imaging (MRI) provides sufficient anatomical details to measure and track changes in volumes of brain structures. The volumes of brain structures and changes in them can be used to study the effects of disease, treatment monitoring, aging, learning and brain development. The present thesis investigates the requirements for performing reproducible quantitative brain volume measurements with automated brain tissue segmentation tools and gives an error bound on the measurements under various experimental conditions. A short introduction into the challenges of performing reproducible brain volume measurements and the main issues that impede the adoption of quantitative volumetric measurements in clinical practice is given, followed by an overview of the acquisition, reconstruction and automated image segmentation methods used to perform quantitative brain volume measurements. The first part of this study was carried out to investigate the reproducibility of volumetric measurements performed on different systems with a standardized ADNI protocol. Systematic biases in volume measurements were observed when there were changes in systems between the first scan and rescan. An important finding in the context of patient management was that neither repositioning nor a two-week gap between the measurements did significantly contribute to the uncertainty in volumetric measurements when compared to the uncertainty in a back-to-back scan-rescan scenario. In the second part of this study, the impact of new highly-accelerated acquisition protocols on automated brain tissue volume measurements was investigated. A single system was used to collect the data and acquisition time was varied at the expense of the SNR. An important outcome of this study was that for qualitative assessment accelerated protocols provided similar information. However, the automated volume measurements with the highly-accelerated protocols were found biased compared to the measurements with standardized ADNI protocol. In the final part of this study, scaling procedures were investigated as means for compensating for the observed differences in sequential automated brain volume measurements. A new image-property-based compensation strategy was proposed and compared to the current state-of-the-art protocol-based approaches. The main outcomes of this study were that there are limitations to the current state-of-the-art protocol-based approaches, namely that volume correction coefficients used in the protocol-based approaches can vary as a function of age, and there is an indication that the proposed image-property-based approach can be more robust to the age and contrast-dependent effects compared to protocol-based approaches.

Key words: magnetic resonance imaging, structural MRI, accelerated MPAGE protocols, brain morphometry analysis, reproducibility of brain volume measurements

Résumé

L'imagerie par résonance magnétique (IRM) structurale pondérée T1 fournit des détails anatomiques suffisants pour mesurer et suivre les changements volumétriques des structures cérébrales. Les volumes des structures cérébrales et leurs changements peuvent être utilisés pour étudier les effets d'une maladie, de l'âge, de l'apprentissage et du développement cérébral, ou pour contrôler un traitement. La présente thèse étudie les conditions nécessaires pour réaliser des mesures quantitatives reproductibles des volumes cérébraux avec des outils de segmentation automatique du tissu cérébral. Elle donne également une erreur liée aux mesures faites avec différentes conditions expérimentales. Une courte introduction présente les défis que pose la réalisation de mesures de volume cérébraux reproductibles ainsi que les principales difficultés qui entravent l'adoption de mesures volumétriques quantitatives dans la pratique clinique. Elle est suivie d'une présentation générale des méthodes d'acquisition, de reconstruction et de segmentation automatique de l'image qui sont utilisées pour réaliser des mesures de volumes cérébraux. La première étude a été conduite pour étudier la reproductibilité des mesures volumétriques réalisées avec différents systèmes en utilisant le protocole ADNI standard. Des biais systématiques ont été observés lorsqu'il y avait des changements de système entre la première acquisition d'une image IRM et la répétition de cette acquisition. Une importante découverte dans le contexte de la prise en charge des patients est que ni le repositionnement ni une pause de deux semaines entre deux mesures ne contribue de manière significative à l'incertitude des mesures volumétriques lorsque celles-ci sont comparées à l'incertitude d'un scénario d'acquisition avec réacquisition immédiate. Dans la seconde étude, l'impact des nouveaux protocoles d'acquisition hautement accélérés sur les mesures volumétriques automatiques du tissu cérébral a été examiné. Un système unique a été utilisé pour collecter les données et le temps d'acquisition a été modifié aux dépens du rapport du signal au bruit. Le principal résultat de cette étude est que, pour une évaluation qualitative, les protocoles accélérés fournissent une information similaire aux protocoles non-accelérés. Cependant, des biais dans les mesures automatiques des volumes sont présents dans les protocoles hautement accélérés en comparaison au protocole ADNI standard. Finalement, des procédures de graduation ont été étudiées pour compenser les différences observées dans les mesures automatiques des volumes cérébraux. Une nouvelle stratégie de compensation basée sur les propriétés de l'image a été proposée et comparée aux approches actuelles basées sur les protocoles. Le principal résultat de cette étude est qu'il y a des limitations aux approches actuelles basées sur les protocoles, à savoir que les coefficients de correction de volume utilisés dans ces approches peuvent changer en fonction de l'âge. Il

Il y a un indice qu'une approche basée sur les propriétés de l'image peut être plus fiable par rapport à l'âge et aux effets dépendants du contraste en comparaison aux approches basées sur les protocoles.

Mots clés : IRM structurel, analyse morphométrique cérébrale, les protocoles MPRAGE accélérés, reproductibilité des résultats d'analyse morphométrique cérébrale

Contents

Acknowledgements	i
Abstract (English/Français)	iii
List of figures	ix
List of tables	xiii
1 Introduction	1
1.1 Objectives and overview of this thesis	1
1.2 Main contributions	4
2 Background	5
2.1 Magnetic Resonance Imaging	6
2.1.1 Acquisition and Reconstruction	8
2.1.2 MPRAGE Acquisition	8
2.1.3 Conventional parallel imaging (GRAPPA)	9
2.1.4 Automated Brain Morphometry	11
3 Reproducibility of volumetric data acquired with a standardized T1-weighted brain-imaging protocol	13
3.1 Introduction	14
3.2 Materials and Methods	15
3.2.1 Acquisition	15
3.2.2 Data Analysis	16
3.3 Results	19
3.4 Discussion	23
3.5 Conclusion	25
4 Optimization of acquisition strategies and comparison of state-of-the-art T1-weighted whole-brain protocols	27
4.1 Introduction	28
4.2 Theory	29
4.3 Materials and Methods	30
4.3.1 MR Acquisition	30

Contents

4.3.2	Image Processing Pipeline and Analysis	31
4.4	Results	35
4.4.1	Observer Ratings	35
4.4.2	Volumetric Measurements	36
4.4.3	Signal-to-Noise Ratio (SNR)	39
4.4.4	Contrast-to-Noise Ratio (CNR)	40
4.4.5	Noise Effects	41
4.4.6	Numerical experiment with addition of synthetic noise	42
4.5	Discussion	43
4.6	Conclusion	47
5	Correction and optimization strategies towards reliable and robust estimation of data elements	49
5.1	Introduction	50
5.2	Materials and Methods	51
5.2.1	Acquisition	51
5.2.2	Data Processing	52
5.2.3	Statistical analysis	53
5.3	Results	55
5.3.1	Protocol-based compensation strategies	55
5.3.2	Image-property-based compensation strategy	61
5.3.3	Bootstrap	66
5.4	Discussion	67
5.5	Conclusion	71
6	Conclusion and outlook	73
6.1	Conclusion	73
6.2	Future Research	76
A	Appendix	77
A.1	Reproducibility of volumetric data acquired with a standardized T1-weighted brain-imaging protocol	77
A.2	Correction and optimization strategies towards reliable and robust estimation of data elements	80
A.2.1	Protocol-based compensation strategy	80
A.2.2	Image-property-based compensation strategy	87
	Bibliography	101
	Curriculum Vitae	103
	Publications	105

List of Figures

2.1	Illustration of the effect of the application of an external radio frequency (RF) pulse on spins placed in a constant magnetic field(B_0)	7
2.2	Sample MPRAGE images from a single subject showing saggital, axial, and coronal slices.	9
2.3	Top: sequence diagram of the MPRAGE pulse sequence. Phase encoding gradient (G_{PE}), readout gradient (G_{RO}),radio frequency pulse(RF). Bottom: evolution of longitudinal magnetization for WM, GM, and CSF structures as a function of time. 10	
2.4	Sample segmentation results from a single subject showing saggital, axial, and coronal slices. Blue: white matter; Green: grey matter; Red: CSF	11
3.1	Images within one session with the same subject during R0 session.	19
3.2	SNR for different scanner combinations using the segmentation masks from MorphoBox. Error bars show standard deviations across the subjects.	20
3.3	CNR for different scanner combinations (R0 session) using the segmentation masks from MorphoBox. Error bars show standard deviations across the subjects. The outliers are indicated as points.	20
3.4	Regression coefficients as fit by the linear mixed-effect model. Fixed effects are shown in red. Random effects are shown in blue. * $p < 0.05$; ** $p < 10^{-2}$; *** $p < 10^{-3}$	21
3.5	Volume differences for all possible scanner combinations. Each coloured bar represents a difference between scanner regression coefficients. The solid grey rectangle represents inter-subject variability (+/- 1 standard deviation). The dashed rectangle represents intra-subject back-to-back scan-rescan variability (+/- 1 standard deviations). * $p < 0.05$; ** $p < 10^{-2}$; *** $p < 10^{-3}$	21
3.6	Random components of variability expressed as percentage of between-subject variance, scan-rescan (R1)- red, additional variance due to repositioning (R2) - green, additional variance due to 2-week-gap between rescan (R3) - blue. . . .	23
4.1	K-space sampling patterns for the accelerated MPRAGE variants. Solid squares represent acquired lines. Squares with stripes represent lines acquired with a second echo. (a) 2D-GRAPPA acceleration with 2x2 undersampling; (b) CAIPIRINHA with 2x2 undersampling and shift 1; (c) Assignment of echoes in segmented MPRAGE	29

List of Figures

4.2	Sample images showing representative sagittal views of all protocol variants for 3 subjects (A, B, C). Please note the increased noise level of the accelerated protocols that is most visible in the brain stem region of the subjects.	35
4.3	Observer Rankings. Relative scale (1, best to 5, worst). * indicates significant difference from the ADNI-2 protocol.	35
4.4	Segmentation results from a single subject within one session. Blue: white matter; Green: grey matter; Red: CSF.	36
4.5	Intra-class correlation coefficients for different brain structures between all possible protocol combinations.	37
4.6	Volumetric percent difference with ADNI-2 protocol used as reference scan (V_r). * indicates difference from 0% median at the% significance level. * $p<10^{-2}$; ** $p<10^{-4}$;	38
4.7	Volumetric percent difference with all possible choices of reference protocol (V_r).	38
4.8	Voxel-wise SNR maps obtained using the pseudo multiple-replica method for a single subject within one session.	39
4.9	SNR measurement for different brain structures and protocols. Error bars show the standard deviations across the subjects. * indicates significant difference between two protocols.	40
4.10	CNR averaged across subjects. Error bars indicate the standard deviations across the subjects. * indicates significant difference between protocols.	41
4.11	Mean distance between the peaks of the intensity distributions and their variance for the different protocol variants. * indicates significant difference between protocols.	41
5.1	Regression coefficient for protocol-based model fit separately for each structure on ADNI, Basel, interpolated Basel, and combined datasets. Error bars indicate 95% confidence interval computed via boot strap with 500 simulations. * $p<0.05$; ** $p<10^{-2}$; *** $p<10^{-3}$;	55
5.2	Distributions of relative volume differences 1.5T/3T (Adjusted, Original) and 3T/3T (pure scan-rescan variability) of ADNI dataset with protocol-based compensation strategy computed in leave-one-out cross-validation settings	56
5.3	Distributions of relative volume differences 1.5T/3T (Adjusted, Original) and 3T/3T (pure scan-rescan variability) of interpolated Basel dataset with protocol-based compensation strategy computed in leave-one-out cross-validation settings	57
5.4	Distributions of relative volume differences ADNI-2/accelerated protocols (Adjusted, Original) of Bonn dataset with protocol-based compensation strategy computed in leave-one-out cross-validation settings	57
5.5	Distributions of relative volume differences 1.5T/3T (Adjusted, Original) and 3T/3T (pure scan-rescan variability) of interpolated Basel dataset with protocol-based compensation strategy computed when the model is trained on ADNI dataset and applied to scale interpolated Basel dataset	58

5.6 Distributions of relative volume differences 1.5T/3T (Adjusted, Original) and 3T/3T (pure scan-rescan variability) of ADNI dataset with protocol-based compensation strategy computed when the model is trained on interpolated Basel dataset and applied to scale ADNI dataset	59
5.7 Distributions of relative volume differences 1.5T/3T (Adjusted, Original) and 3T/3T (pure scan-rescan variability) of combined dataset with protocol-based compensation strategy computed in leave-one-out cross-validation settings. ADNI and Interpolated Basel datasets are shown separately.	60
5.8 Regression coefficient for CNR, and voxel volume effects fit separately for each structure on ADNI, Basel, interpolated Basel, and combined datasets. Error bars indicate 95% confidence interval computed via boot strap with 500 simulations. * p<0.05; ** p<10-2; *** p<10-3;	61
5.9 Distributions of relative volume differences(Adjusted, Original) 1.5T/3T and 3T/3T of ADNI dataset with image-property-based compensation strategy computed in leave-one-out cross-validation settings	62
5.10 Distributions of relative volume differences(Adjusted, Original) 1.5T/3T and 3T/3T of interpolated Basel dataset with image-property-based compensation strategy computed in leave-one-out cross-validation settings	62
5.11 Distributions of relative volume differences ADNI-2/accelerated protocols (Adjusted, Original) of Bonn dataset with image-property-based compensation strategy computed in leave-one-out cross-validation settings	63
5.12 Distributions of relative volume differences (Adjusted, Original) between 1.5T/3T and 3T/3T of interpolated Basel dataset with image-property-based compensation strategy trained on ADNI dataset	64
5.13 Distributions of relative volume differences (Adjusted, Original) between 1.5T/3T and 3T/3T of ADNI dataset with image-property-based compensation strategy trained on interpolated Basel dataset	64
5.14 Distributions of relative volume differences (Adjusted, Original) 1.5T/3T and 3T/3T of combined dataset with image-property-based compensation strategy computed in leave-one-out cross-validation settings. ADNI and Interpolated Basel datasets are shown separately.	65
5.15 Bootstrap validation of protocol-based compensation strategy. Testing database age is plotted on the y-axis and training database age is plotted on the x-axis. Colour represents the difference in mean RVDs of adjusted and unadjusted data.	66
5.16 Bootstrap validation of image-property-based compensation strategy. Testing database age is plotted on the y-axis and training database age is plotted on the x-axis. Colour represents the difference in mean RVDs of adjusted and unadjusted data.	67

List of Tables

4.1	Intra-class correlation coefficients (95% confidence interval shown in brackets).	37
4.2	Intra-class correlation coefficients (95% confidence interval shown in brackets) for the numerical experiment with addition of noise to the reference ADNI-2 scan.	42
A.1	Regression coefficients as fit by the linear mixed-effect model. * p<0.05; ** p<10 ⁻² ; *** p<10 ⁻³	78
A.2	Volume differences for all possible scanner combinations. * p<0.05; ** p<10 ⁻² ; *** p<10 ⁻³	79
A.3	Mean relative volume differences 1.5T/3T (Adjusted, Original) and 3T/3T (Target, pure scan-rescan variability) of ADNI dataset with protocol-based compensation strategy computed in leave-one-out cross-validation settings	80
A.4	Mean relative volume differences 1.5T/3T (Adjusted, Original) and 3T/3T (Target, pure scan-rescan variability) of interpolated Basel dataset with protocol-based compensation strategy computed in leave-one-out cross-validation settings	81
A.5	Mean relative volume differences ADNI-2/accelerated protocols (Original, Adjusted) of Bonn dataset with protocol-based compensation strategy computed in leave-one-out cross-validation settings	82
A.6	Mean relative volume differences 1.5T/3T (Original, Adjusted) and 3T/3T (Target pure scan-rescan variability) of interpolated Basel dataset with protocol-based compensation strategy computed when the model is trained on ADNI dataset and applied to scale interpolated Basel dataset	83
A.7	Mean relative volume differences 1.5T/3T (original/corrected) and 3T/3T (Target, pure scan-rescan variability) of ADNI dataset with protocol-based compensation strategy computed when the model is trained on interpolated Basel dataset and applied to scale ADNI dataset	84
A.8	Mean relative volume differences of ADNI dataset 1.5T/3T (Adjusted, Original) and 3T/3T (Target, pure scan-rescan variability). Combined dataset with protocol-based compensation strategy computed in leave-one-out cross-validation settings.	85
A.9	Mean relative volume differences of interpolated Basel dataset 1.5T/3T (Adjusted, Original) and 3T/3T (Target, pure scan-rescan variability). Combined dataset with protocol-based compensation strategy computed in leave-one-out cross-validation settings.	86

List of Tables

A.10 Mean relative volume differences (Adjusted, Original) 1.5T/3T and 3T/3T of ADNI dataset with image-property-based compensation strategy computed in leave-one-out cross-validation settings	87
A.11 Mean relative volume differences (Adjusted, Original) 1.5T/3T and 3T/3T of interpolated Basel dataset with image-property-based compensation strategy computed in leave-one-out cross-validation settings	88
A.12 Mean relative volume differences ADNI-2/accelerated protocols (Original, Adjusted) of Bonn dataset with image-property-based compensation strategy computed in leave-one-out cross-validation settings	89
A.13 Mean relative volume differences (Adjusted, Original) between 1.5T/3T and 3T/3T of interpolated Basel dataset with image-property-based compensation strategy trained on ADNI dataset	90
A.14 Mean relative volume differences (Adjusted, Original) between 1.5T/3T and 3T/3T of ADNI dataset with image-property-based compensation strategy trained on interpolated Basel dataset	91
A.15 Mean relative volume differences of interpolated Basel dataset (Adjusted, Original) between 1.5T/3T and 3T/3T. Combined dataset with image-property-based compensation strategy computed in leave-one-out cross-validation settings.	92
A.16 Mean relative volume differences of ADNI dataset (Adjusted, Original) between 1.5T/3T and 3T/3T. Combined dataset with image property based compensation strategy computed in leave-one-out cross-validation settings.	93

1 Introduction

1.1 Objectives and overview of this thesis

This thesis investigates the requirements for performing reproducible quantitative volume measurements of brain structures with automated brain tissue segmentation tools and investigates the applicability of compensation strategies to mitigate discrepancies between the measurements. The following is a short history overview with an introduction into the challenges of reproducible brain volume measurements and the main issues slowing down the adoption of quantitative volumetric measurements in clinical practice.

Magnetic resonance imaging (MRI) is an imaging modality that is often used in both clinical and research settings to image anatomy and physiological processes within a human body. MRI is particularly useful for imaging soft tissues since it provides a better soft-tissue contrast compared to computed tomography (CT), and is a non-invasive imaging technique that does not use ionizing radiation in contrast to CT.

The early development of MRI started in 1950 with the detection of the spin echoes by Erwin Hahn, and generation of a one-dimensional nuclear magnetic resonance (NMR) spectrum by Herman Carr in 1952. An important milestone was achieved in 1971 by Raymond Damadian who demonstrated that tumors and normal tissue can be distinguished in-vivo using nuclear magnetic resonance. In 1973, Paul Lauterbur produced the first nuclear magnetic resonance image. During the following years, MRI has been evolving rapidly to allow imaging of most parts of the human body. In the present thesis, we focus on brain-imaging.

The first commercially available MRI scanners appeared 30 years ago and the adoption of MRI modality has grown since then. In clinical practice, three dimensional brain-images produced by MRI are often used to exclude pathology, for disease detection and treatment monitoring. For instance, the anatomical details provided by structural T1-weighted MRI make it possible to assess gray matter concentrations within the brain, leading to an active field of research known as brain morphometry. Brain morphometry served as a valuable tool to diagnose and track changes associated with neurodegeneration, multiple sclerosis, inflammatory and

Chapter 1. Introduction

neurological diseases as well as normal aging, learning and evolution [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15]. In order to quantitatively assess these changes, manual, semi-automated, and automated brain volume segmentation tools are developed by the research community.

However, there is a considerable gap between state-of-the-art research techniques and everyday clinical practice. Despite promising findings, quantitative assessment of brain-tissue volumes is not yet widely used in clinical practice neither it is reflected in clinical guidelines or diagnostic criteria. The reasons of such situation and proposed solutions are summarized below.

There are several hindrances that slow down the adoption of quantitative volumetric measurements in clinical practice:

1. Manual tissue segmentation approaches are not practical in clinical settings since they are labor-intensive, require a high degree of expertise, and may introduce operator-dependent bias. However, these issues can be largely addressed with the use of automated brain segmentation tools. Automated brain segmentation tools have the potential to remove the operator-related bias and reduce the labour requirements.
2. Both automated and manual segmentation approaches can be sensitive to variations in imaging protocols of MRI data and, as a result, may potentially reduce precision of the volume measurements. In the context of Alzheimer's disease and dementia, there has been a substantial effort to standardize MRI acquisition protocols across vendors and field strength [16, 17]. The reproducibility of the standardized ADNI protocol has been extensively studied [18, 19, 20, 21, 22, 23, 24, 25, 26, 27]. However, the current thesis extends previous reproducibility studies by decoupling systematic biases on volume measurements due to hardware (scanner/protocol) and physiology from random offsets due to repositioning and inter-subject variability.

Rigorous application of the standardized protocols can be challenging. For instance, the typical acquisition time for a T1-weighted structural brain image is about 10 mins using the standardized ADNI-1 protocol [16] and 5 mins using ADNI-2 [17]. However, with uncooperative patients, the use of standardized protocols like ADNI can be difficult due to patient motion, and reduced acquisition time is desirable. More generally, in a radiology department, reduction of acquisition time can increase the patient throughput, which is beneficial both in terms of costs, and overall efficiency of utilization of resources available to the department. Recent advances in MRI hardware and acquisition methodology can potentially enable further reduction of acquisition time compared to conventional protocols in both clinical and research settings. This thesis aims to gain further insight into the exact impact of the use of the highly accelerated protocols on the outcomes of automated morphometry tools.

3. Normative ranges for volumes need to be available for assessing how brain structure volumes of a particular subject compare with a healthy population. In order to construct

normative ranges, a large number of subjects with various demographics needs to be imaged to capture the inter-subject variation across the population as well as systematic effects such as normal aging or gender effects. This can be a rather challenging task from the perspective of protocol standardization across hardware setups and vendors. For instance, despite the standardization efforts of ADNI, it has been shown that biases between scanners of different field strength are statistically significant [20]. Moreover, in a long-term perspective, accommodating new technological developments would either require incorporation of compensation strategies or construction of normative ranges specific to hardware and protocol setups. This second option would be very inefficient, time consuming, and costly. This work proposes and investigates applicability of compensation strategies as means to account for differences between protocols and hardware setups in the construction of normative ranges for brain volumes.

Overall, the present thesis aims to address all of the 3 main issues impeding the adoption of quantitative volumetric measurements in clinical practice. To achieve these goals, this work first examined the reproducibility of automated volume measurements performed on data acquired with standardized ADNI protocol[16, 17] on several systems under different experimental conditions(repositioning, 2-week gap between a rescan). The impact of highly-accelerated MPRAGE protocols on automated volume measurements and clinical readings was studied. Finally, the applicability of scaling procedures that can be used to put the sequential volume measurements acquired with different systems and protocols into the same frame of reference was investigated. The thesis structure is as follows.

- Chapter 1 is an overview of the thesis and a summary of its main contributions.
- Chapter 2 provides the necessary background on image acquisition and reconstruction. Describes the basis of automated segmentation algorithm used for the purpose of volumetric assessment.
- Chapter 3 summarizes the effects of using different hardware setups on the outcomes of automated morphometric assessment when standardized protocols are used.
- Chapter 4 describes the effects of using highly accelerated protocols on the outcomes of automated morphometric assessment with data acquired on a single system.
- Chapter 5 investigates the applicability of scaling procedures to compensate for discrepancies between hardware setups and protocols.
- Chapter 6 summarizes the main outcomes of the thesis and provides an outlook on the future work.

1.2 Main contributions

The main contributions of this thesis are:

- Investigated intra- and inter-scanner reproducibility of volumetric assessment of brain structures based on T1-weighted whole-brain imaging with standardized protocols
- Compared state-of-the-art whole-brain T1-weighted brain imaging protocols and developed a new pulse sequence (Segmented MPRAGE)
 - P. Falkovskiy, D. Brenner, T. Feiweier, S. Kannengiesser, B. Maréchal, T. Kober, A. Roche, K. Thostenson, R. Meuli, D. Reyes, T. Stoecker, M. A. Bernstein, J.-P. Thiran, and G. Krueger, “Comparison of accelerated T1-weighted whole-brain structural-imaging protocols,” *Neuroimage*, vol. 124, pp. 157–167, 2016.
 - P. Falkovskiy, D. Brenner, T. Feiweier, S. Kannengiesser, B. Maréchal, T. Kober, A. Roche, K. Thostenson, M. Seeger, T. Stoecker, M. Bernstein, and G. Krueger, “Comparison of accelerated T1-weighted whole-brain structural imaging protocols,” in *Proc. Intl. Soc. Mag. Reson. Med.* 22, 2014.
 - P. Falkovskiy, T. Kober, D. Reyes, K. Steinert, M. Seeger, M. Bernstein, and G. Krueger, “Segmented Multi-Echo MPRAGE Acquisition for Accelerated T1-weighted Brain Imaging,” in *Proc. Intl. Soc. Mag. Reson. Med.* 21, 2013.
- Adapted Segmented MPRAGE pulse sequence to serve as a navigator for motion correction framework
 - M. Waszak*, P. Falkovskiy*, T. Hilbert, G. Bonnier, B. Maréchal, R. Meuli, R. Gruetter, T. Kober, and G. Krueger, “Prospective head motion correction using FID-guided on-demand image navigators,” *Magn. Reson. Med.*, 2016.
 - M. Babayeva*, P. Falkovskiy*, T. Hilbert, G. Bonnier, B. Maréchal, R. Meuli, J. Thiran, R. Gruetter, G. Krueger, and T. Kober, “Prospective motion correction with FID-triggered image navigators,” in *Proc. Intl. Soc. Mag. Reson. Med.* 23, 2015.
- Investigated the means of correcting for discrepancies between protocols and scanners
- Investigated the applicability of the skull stripping structures for data acquired with MP2RAGE pulse sequence
 - P. Falkovskiy, B. Maréchal, S. Yan, Z. Jin, T. Qian, K. O’Brien, R. Meuli, J. Thiran, G. Krueger, T. Kober, and A. Roche, Quantitative comparison of MP2RAGE skull-stripping strategies, *Proc. Intl. Soc. Mag. Reson. Med.* 24, 2016.

2 Background

This chapter presents a summary of the concepts that are used in the subsequent chapters. First, it gives a short overview of the physical basis of the magnetic resonance imaging, describes acquisition and reconstruction of T1-weighted contrasts with MPRAGE pulse sequences. Finally, it gives a summary of the automated segmentation algorithm that was used for the purpose of volumetric assessment throughout the present thesis.

This chapter is based on Haacke, Brown, Thompson, & Venkatesan, 1999; McRobbie, 2003.

2.1 Magnetic Resonance Imaging

Magnetic resonance imaging is a medical imaging modality that can image both anatomy and function of the human body through the observation of the interaction of a nuclear spins with an external magnetic field. Spin is an intrinsic property of matter. Due to their prevalence, spins of protons are often used for human imaging. The spins of protons have two discrete energy states and when protons are placed in an external magnetic field, their spins will precess around the direction of the external magnetic field. The precessional frequency of the spins of protons is proportional to the magnetic field and is given by the Larmor equation:

$$\omega_0 = \gamma B_0, \quad (2.1)$$

where γ is a gyromagnetic ratio and B_0 is the magnetic field. At room temperature, the number of spins in a lower energy state (same direction as magnetic field) will be greater than in the higher energy state giving rise to the net magnetization (M_0) parallel to the magnetic field.

It is possible to interact with the spins through the application of a radio frequency (RF) pulse. When placed in a magnetic field, a particle with a non-zero net spin can absorb a photon and make a transition to a higher energy state. However, since the spin is quantized, the RF pulse needs to deposit the exact energy that corresponds to the difference between lower and higher energy states. In the case of a proton placed in an external magnetic field, the precessional frequency of its spin is the same as the frequency of radiation needed to transition between two states.

Because net magnetization is very small compared to the main magnetic field, it is virtually impossible to measure magnetization in the body at room temperature when it is at equilibrium and parallel with the main magnetic field B_0 . However, if the magnetization is tipped in the transverse plane through the application of a RF pulse, the precession of the net magnetization in the transverse plane would result in a detectable signal. An important effect of the application of a RF pulse is that it would bring the spins into phase coherence.

A simplistic MR experiment is illustrated in Figure 2.1. Initially, all of the spins are at equilibrium precessing around the main magnetic field. The main magnetic field (B_0) is parallel to the z-axis. If the system of spins is at room temperature, there would be a small excess of spins that are in the lower energy state. This would result in a non-zero net magnetization that is parallel to the main magnetic field. Following the application of RF pulse, the net magnetization would be tipped due to non-zero components of the net magnetization in transverse (phase coherence) and parallel (spins transitioning to higher energy state) to the main magnetic field planes. The spins would start to precess and the system would emit a signal referred to as a free-induction decay (FID) as the system returns to the equilibrium state. This signal can be measured by an antenna placed around the subject that is imaged.

The process of the system returning to the equilibrium is known as relaxation. This process

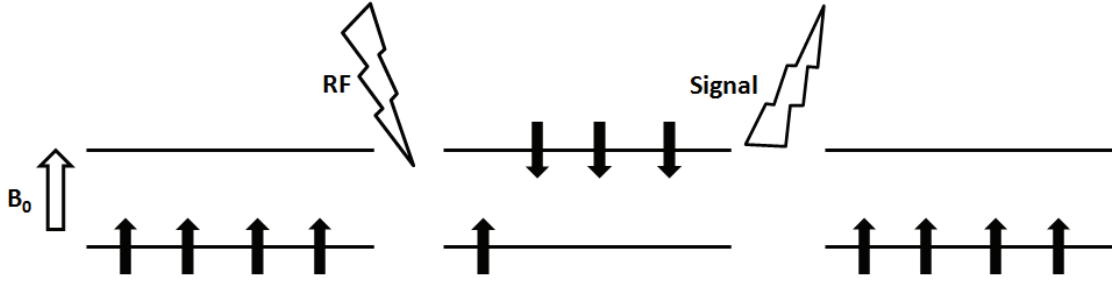


Figure 2.1: Illustration of the effect of the application of an external radio frequency (RF) pulse on spins placed in a constant magnetic field (B_0)

can be described phenomenologically by the Bloch equations.

$$\frac{d}{dt} \begin{pmatrix} M_x(t) \\ M_y(t) \\ M_z(t) \end{pmatrix} = \begin{pmatrix} -\frac{1}{T_2} & \gamma B_z(t) & -\gamma B_y(t) \\ -\gamma B_z(t) & -\frac{1}{T_2} & \gamma B_x(t) \\ \gamma B_y(t) & -\gamma B_x(t) & -\frac{1}{T_1} \end{pmatrix} \begin{pmatrix} M_x(t) \\ M_y(t) \\ M_z(t) \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ \frac{M_0}{T_1} \end{pmatrix} \quad (2.2)$$

where $\vec{M}(t)$ is net magnetization and $\vec{B}(t)$ is external magnetic field.

After the application of the RF pulse, the system will experience only the constant magnetic field B_0 that is parallel to the z-axis. By imposing this limiting condition, we can solve the Bloch equations to describe the relaxation to the equilibrium.

In the plane parallel to the main magnetic field, the system would return to the equilibrium magnetization M_0 as follows:

$$M_z(t) = M_z(0)e^{-\frac{t}{T_1}} + M_0(1 - e^{-\frac{t}{T_1}}) \quad (2.3)$$

where $M_z(t)$ is a component of magnetization parallel to the main magnetic field.

In the transverse plane, the system will return to the equilibrium as:

$$M_x(t) = (M_x(0)\cos(\omega_0 t) + M_y(0)\sin(\omega_0 t))e^{-\frac{t}{T_2}} \quad (2.4)$$

$$M_y(t) = (M_y(0)\cos(\omega_0 t) - M_x(0)\sin(\omega_0 t))e^{-\frac{t}{T_2}} \quad (2.5)$$

where $M_x(t)$, and $M_y(t)$ are the components of magnetization in the transverse plane.

There are two constants that describe this process of decaying to the equilibrium:

- Longitudinal relaxation time (T_1) that is due to "spin-lattice" interactions
- Transverse relaxation time (T_2) that is due to the "spin-spin" interactions

Chapter 2. Background

White matter (WM), gray matter (GM), and cerebro-spinal fluid (CSF) would exhibit different T_1 , and T_2 relaxation times. MRI exploits these properties of the tissues to generate the contrast in images.

2.1.1 Acquisition and Reconstruction

In order to produce 3D image-volumes, the spatial dependence would need to be encoded into FID signal. This can be accomplished by applying additional spatially varying magnetic fields. These spatially varying magnetic fields are referred as gradients.

Following the application of the RF excitation pulse, it is possible to encode the spatial information in the first 2 dimensions (\hat{y}, \hat{z}) using the phase of the precession of the spins. This can be accomplished by applying a spatially linearly varying magnetic field gradient in both dimensions (G_y, G_z) for a short period of time. Initially, right after the application of the gradient, the spins would precess with different frequencies depending on their spatial position, and then when the gradient is turned off, the spins would return to the original precessional frequency. Overall, this would give a spatial-dependent phase offset to the precession of the spins. For the one remaining dimension (\hat{x}), after the application of phase encoding gradients, a linearly spatially varying readout gradient (G_x) is applied in the direction of the remaining dimension and the signal is sampled while this gradient is applied.

To summarize, this process would result in a signal received by the antenna that would contain the precessional frequencies of the spins encoding the spatial locations in \hat{x} dimension and the phases of the spins encoding the spatial locations in the remaining dimensions \hat{y}, \hat{z} . The signal received from a sample can be summarized as:

$$S(k_x, k_y, k_z) = \iiint dx dy dz \rho(x, y, z) e^{-i2\pi(xk_x + yk_y + zk_z)} \quad (2.6)$$

where ρ is a density of spins emitting the signal and $k_x = \frac{\gamma}{2\pi} \int^t dt G_x(t)$, $k_y = \frac{\gamma}{2\pi} \int^t dt G_y(t)$, $k_z = \frac{\gamma}{2\pi} \int^t dt G_z(t)$ define the k-space trajectory through the application of gradients (G_x, G_y, G_z).

It is possible to reconstruct the original image $\rho(x, y, z)$ by applying an inverse Fourier transform. However, a big challenge is to manipulate the gradients in a way that would efficiently sample the k-space.

2.1.2 MPRAGE Acquisition

Magnetization Prepared Rapid Acquisition by Gradient Echo (MPRAGE) [28] pulse sequence is often used in 3D T1-weighted (T1w) brain-imaging due to an excellent GM-WM contrast. A typical use case in brain-imaging for this pulse sequence would be to provide anatomical information. This anatomical information can be later used to track the changes in tissues associated with disease, ageing, treatment, or to provide the supporting anatomical information to other imaging sequences or modalities. Typical images acquired using an MPRAGE pulse

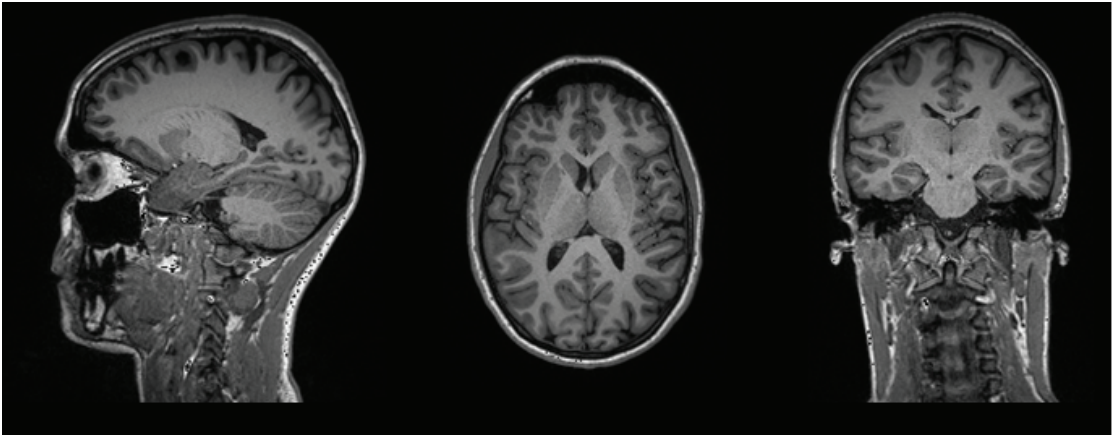


Figure 2.2: Sample MPRAGE images from a single subject showing sagittal, axial, and coronal slices.

sequence with ADNI-2 protocol parameters[17] are illustrated in Figure 2.2.

The sequence can be split into 3 parts: preparation, acquisition and recovery (Figure 2.3). Typical preparation module consists of a 180° inversion pulse followed by a delay. First, the inversion pulse would invert the longitudinal magnetization of all excited tissues. During the delay period, the longitudinal magnetization of GM, WM, CSF would recover at different rates due to differences in the tissue-specific T1 values. The differences in the recovery of longitudinal magnetization between the tissues would give rise to a contrast between the tissues (Figure 2.3).

The acquisition part consists of a gradient echo readout that would use low-flip-angle excitation pulses. Typically per preparation part, the acquisition module would apply phase encoding and frequency encoding gradients to sample data along the readout direction (\hat{x}) and partition direction (\hat{z}) while keeping the line direction (\hat{y}) fixed. Finally, following the acquisition module, the delay portion of the sequence allows the magnetization to recover.

The subsequent repetitions of these steps would fill the k-space through sampling the data along the line direction (\hat{y}) dimension.

2.1.3 Conventional parallel imaging (GRAPPA)

In the current clinical and research practice GRAPPA- [29] or SENSE-type accelerations [30] are often used reduce the acquisition time. The conventional parallel imaging techniques reduce the acquisition time through undersampling of the k-space data in one phase-encoding direction (1D).

Let us consider 1D GRAPPA reconstruction that is used in the ADNI-2 protocol. The data is undersampled in k_y direction. First, we define $S(k_y)$ as acquired signal, and $S(k_y + m\delta k_y)$ as

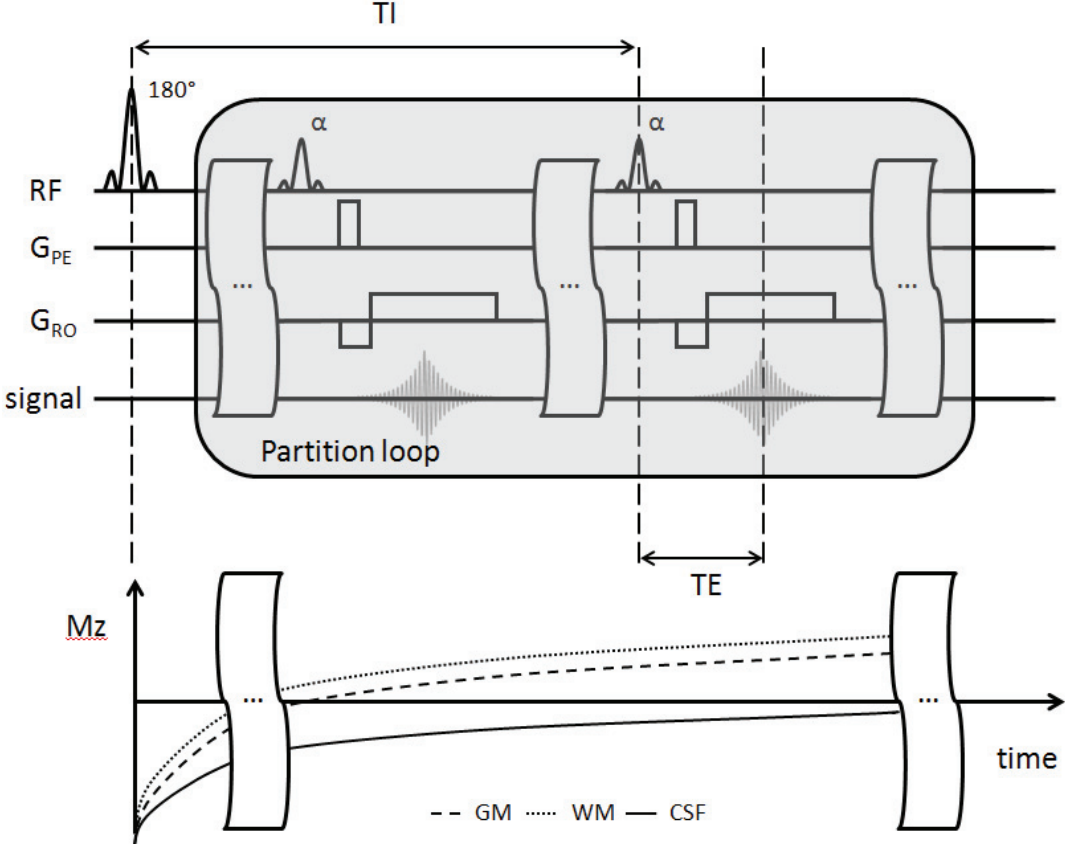


Figure 2.3: Top: sequence diagram of the MPRAGE pulse sequence. Phase encoding gradient (G_{PE}), readout gradient (G_{RO}), radio frequency pulse (RF). Bottom: evolution of longitudinal magnetization for WM, GM, and CSF structures as a function of time.

reconstructed signal. We can write 1D GRAPPA reconstruction as:

$$S(k_y + m\delta k_y) = \hat{G}_{(y, m)} S(k_y), m = 1..R - 1 \quad (2.7)$$

where $\hat{G}_{(y, m)}$ is a weight matrix that can be calculated from the reference lines.

When reference lines are acquired, we have the knowledge of both $S(k_y)$ and $S(k_y + m\delta k_y)$, but we do not know the weight matrix. The weight matrix $\hat{G}_{(y, m)}$ can be calculated by computing a pseudo-inverse as follows:

$$\hat{G}_{(y, m)} = S(k_y + m\delta k_y)(S(k_y)^H S(k_y))^{-1} S(k_y)^H \quad (2.8)$$

2.1.4 Automated Brain Morphometry

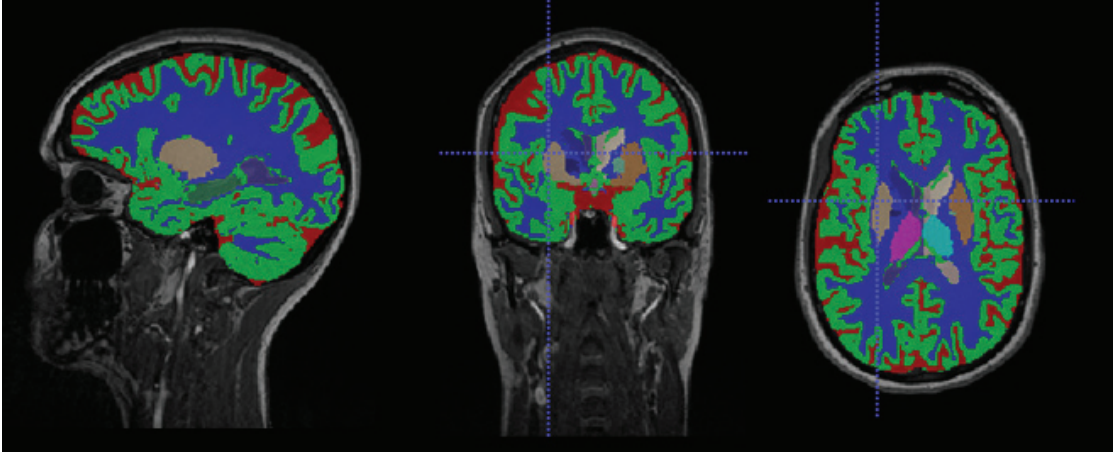


Figure 2.4: Sample segmentation results from a single subject showing sagittal, axial, and coronal slices. Blue: white matter; Green: grey matter; Red: CSF.

The automated brain tissue segmentation algorithm used in the present thesis is described in [12]. A short overview of the algorithm is presented in this section. The segmentation algorithm uses a template. The template consists of a T1-weighted MPRAGE scan acquired with ADNI-2 [17] protocol parameters and a mask that defines voxels that belong to various anatomical structures. The mask image was drawn and reviewed by several neurologists. The following anatomical structures are present in the mask image: the total intracranial volume (TIV) defined by the hemispheric and cerebellar gray matter (GM), white matter (WM) and the intracranial cerebrospinal fluid (CSF), lateral, third and fourth ventricles, cerebellum, thalamus, putamen, pallidum, caudate nucleus, and hippocampus. The flow of the algorithm can be summarized as follows.

First, an image template that outlines the tissue structures of interest is registered into the input image space. The registration of the template consists of estimating the 9 parameter(translation, rotation, and anisotropic scaling) affine transform [31] followed by the estimation of free-form diffeomorphic displacement field [32]. Following the template registration,

Chapter 2. Background

an additional bias field correction similar to [33] is applied to the input image volume.

Based on the estimated parameters of the transformation of the template image, masks defining the voxels that belong to each structure of interest are resampled into the image space. The voxels that belong to the TIV in the input image are defined based on the resampled masks of the template.

The following steps of the tissue segmentation algorithm are restricted to the TIV. The signal intensities are modeled as a five-class Gaussian mixture model that additionally incorporates a stationary Markov–Potts prior model[34, 35]. The classes used in this model roughly correspond to ventricular CSF, sulcal CSF, cortical GM, deep GM and WM. The model is fit using the variational expectation-maximization (VEM) algorithm and will output five posterior probability maps. The masks and volume estimates corresponding to CSF/GM/WM structures are then computed from the posterior probability maps. For the remaining structures, the tissue probability maps are combined with the masks of the resampled template to produce tissue volume estimates and corresponding masks.

A sample output of the label masks is illustrated in Figure 2.4.

3 Reproducibility of volumetric data acquired with a standardized T1-weighted brain-imaging protocol

With advances in medical imaging technology, clinical radiology is increasingly requested to provide sensitive markers to measure disease progression, to differentiate pathology from healthy aging, and to monitor therapy response. This chapter quantitatively characterizes MPRAGE scans acquired with standardized ADNI protocol parameters across different systems, field strength, shimming, and brain physiology based on basic image quality metrics (SNR, CNR) and examines the impact of scanner's model, field strength, protocol parameters, shimming, and brain physiology on the reliability and reproducibility of automated volumetric brain measurements in clinical settings.

Chapter 3. Reproducibility of volumetric data acquired with a standardized T1-weighted brain-imaging protocol

3.1 Introduction

Automated brain morphometry of magnetic resonance imaging (MRI) data is becoming an important tool in studies aiming to quantify disease progression, drug efficiency or normal aging. For instance, hippocampal atrophy is recognized as a biomarker for Alzheimer Dementia (AD) and mild cognitive impairment (MCI) [36, 37]. In the context of multiple sclerosis (MS), in addition to lesion count and volume, the atrophy rate of white matter (WM), grey matter (GM), thalamus, and caudate structures can be used as predictor for clinical status [7, 11, 13].

For the purpose of morphometric assessment, the magnetisation-prepared rapid acquisition gradient-echo (MPRAGE) pulse sequence [28] is often used to produce T1-weighted images because it exhibits excellent grey-white matter contrast. In practice, it is difficult to conduct a large cross-sectional or longitudinal study on the same system and as a result there has been a substantial effort to optimize reliability and reproducibility of the morphometric measurements across different systems and field strength through the use of standardized protocols and regular phantom calibration scans [8, 16, 17].

However, even when standardised protocols are used, automated volume measurements may be affected by physiological factors (e.g., dehydration), and system-specific factors such as shim settings, field strength, scanner model, and software version [20, 22]. Depending on the exact experimental setup, these factors may affect both cross-sectional and longitudinal exams. In cross-sectional studies, they may affect the ability to detect differences between groups if subjects are scanned on different types of systems. In longitudinal studies, important anatomical changes over time may remain undetected due to the large variance in image content of incompatible sequential measurements.

Some of these pitfalls can be addressed by careful adoptions of the patient scheduling, often at the cost of increased complexity, and reduced efficacy and productivity of the radiology department.

This work quantitatively characterizes MPRAGE scans acquired using different scanners, protocols, shim settings, and physiological conditions to investigate reliability and reproducibility of brain volumetry under clinical conditions. The reproducibility of the standardized 5-minute ADNI-2 protocol has been extensively studied [18, 19, 20, 21, 22, 23, 24, 25, 26, 27]. However, the current study aims to extend previous reproducibility studies by decoupling systematic offsets to volume measurements due to hardware (scanner/protocol) effects from random offsets due to repositioning, and physiology with linear mixed-effect model analysis. This decoupling of fixed effects from random effects makes it possible to examine the importance of each effect individually and to get a deeper insight into the merit of applying compensation strategies.

3.2 Materials and Methods

3.2.1 Acquisition

Experiments were performed on four clinical MRI scanners (Siemens Healthcare, Erlangen, Germany):

1. 3 T MAGNETOM Skyra (20-channel head/neck coil)
2. 3 T MAGNETOM Prisma (64-channel head/neck coil)
3. 1.5 T MAGNETOM Avanto (12-channel head coil)
4. 1.5 T MAGNETOM Espree (12-channel head coil)

22 healthy young subjects (20-39 years old) were imaged on each scanner after obtaining a written and informed consent. The physiological parameters hydration status and arterial blood pressure of each subject were controlled prior to each acquisition. The following MPRAGE protocols were used:

- (a) 3 T - scanners: 5:12-minute protocol as used in ADNI-2 ($TR/TI/BW/\alpha = 2300 \text{ ms}/900 \text{ ms}/240 \text{ Hz}/\text{px}/9 \text{ deg.}$) [17] but with isotropic 1mm resolution
- (b) 1.5 T - scanners: 4:42-minute protocol as used in ADNI ($TR/TI/BW/\alpha = 2400 \text{ ms}/1000 \text{ ms}/180 \text{ Hz}/\text{px}/8 \text{ deg.}, 1.25 \times 1.25 \times 1.20 \text{ mm}^3$) [16] but two-fold accelerated using GRAPPA ($R=2$)

For Avanto and Prisma scanners, the measurement program consisted of four MPRAGE volume acquisitions:

- (R0) Reference MPRAGE scan
- (R1) Repeat scan acquired in the same session and back-to-back with R0 ("best case" scan-rescan variability)
- (R2) Scan with repositioning of the subject during the same session as R0 (adds the effect of repositioning and shim changes to R1)
- (R3) Scan conducted within 2 weeks after R0 (adds physiological variances to R2)

At Skyra and Espree scanners, the measurement program consisted of only (R0) and (R3) acquisitions.

For all acquisitions, FOV placement was guided by a scanner-integrated AutoAlign feature. Intensity normalization was performed using a scanner software-integrated pre-scan procedure,

Chapter 3. Reproducibility of volumetric data acquired with a standardized T1-weighted brain-imaging protocol

and an adaptive coil combination mode [38] was used to reconstruct the image volumes for all acquisitions leading to 264 MPRAGE image volumes. In addition to the reconstructed imaging data, all raw data from 3T systems were saved for retrospective modified reconstructions.

3.2.2 Data Analysis

Quantitative volumetric measurements and label maps defining voxels that belong to each structure were computed using the prototype morphometry package MorphoBox [39, 12] for grey-matter (GM), white-matter (WM), normalized whole brain volume (NBV), cerebrospinal fluid (CSF), hippocampus, thalamus, putamen, caudate, and pallidum. Segmentation software was run in a fully automated manner and no manual editing was done at any stage of the processing. Segmentation results were visually inspected for gross errors.

Contrast-to-Noise Ratio (CNR)

Prior to CNR assessment, a separate B1 bias field correction [35] was applied to all acquired data in addition to intensity normalization performed using a scanner-integrated pre-scan procedure.

CNR was assessed using label maps provided by MorphoBox as:

$$CNR_{csf-gm} = \frac{(\mu_{gm} - \mu_{csf})^2}{(\sigma_{csf}^2 + \sigma_{gm}^2)} \quad (3.1)$$

$$CNR_{gm-wm} = \frac{(\mu_{wm} - \mu_{gm})^2}{(\sigma_{gm}^2 + \sigma_{wm}^2)} \quad (3.2)$$

where μ_{csf} , μ_{gm} , μ_{wm} were median intensities and σ_{csf}^2 , σ_{gm}^2 , σ_{wm}^2 were image intensity variances within CSF, GM and WM volumes. The intensity variances were estimated through computing median absolute deviation as opposed to computing the variances directly since this would render them less sensitive to outliers. The differences in CNR of different scanner/protocol variants were tested with the Wilcoxon rank-sum test [40].

Signal-to-Noise Ratio (SNR)

Using a retrospective reconstruction based on the raw data, voxel-wise SNR ($\mu_{signal}/\sigma_{noise}$) quantification was performed using a prototype implementation of the pseudo-multiple replica approach [41] with 8 pseudo-replicas. The variance of additional synthetic noise used to compute σ_{noise} was estimated based on a short prescan consisting of 128 readouts having 512 sampling points each, including oversampling, without an excitation pulse. To improve the estimation of σ_{noise} , a 2D 5x5 region of neighboring pixels was used to estimate σ_{noise} using a moving-averages fashion according to [42]. Average SNR in the brain stem, WM, and GM was computed for 5 subjects from the anatomical label maps produced by MorphoBox.

Brain stem, WM and GM were chosen as those tissue types that represent regions-of-interest (ROI) of known differences in the coil performance in the centre of the coil and the proximity of the coil elements [43].

Linear Mixed-Effect Model Analysis

Linear mixed-effect models are prime tools to analyze cross-sectional data with repeated measurements per subject [44, 45]. They are able to deal with missing data in contrast to the repeated measurement ANOVA [44, 45]. In the present study, we did not have R1 and R2 sessions acquired on the Espree and Skyra scanners. Therefore, due to the presence of the missing data, a linear mixed-effect model was used for the subsequent analysis.

Prior to further analysis, volumes of all structures were normalized by the total intracranial volume (TIV). A linear mixed-effect model [44] of the form:

$$y_i = X_i\beta + Z_i\gamma_i + \epsilon_i \quad (3.3)$$

was fit on the volumetric data from all subjects with the R software (version 3.1.1) using lme4 package [44]. In this model, y_i was a vector of serial volume measurements for subject i , X_i was a matrix that models the scanner/protocol type, gender and age of the subject (fixed effects). Random effects consisted of both subject-dependent and session-dependent offsets modelled by the matrix Z_i ; β and γ_i were vectors of unknown regression coefficients and the residual error ϵ_i represented the scan-rescan variability. Regression coefficients β were assumed to be the same across all of the subjects while γ_i and ϵ_i were treated as a random variables. Note that effects of scanner/protocol, subjects' gender and age, subject-dependent and session-dependent offsets were modelled as additive terms and are in percentage of TIV units.

Both variability due to scan-rescan and additional variability (repositioning, physiology etc.) as detected with (R2) and (R3) were assumed to be same across scanner models based on the findings previously demonstrated in [20].

The Satterthwate's approximation was used to estimate the number of degrees of freedom with lmerTest package [46], and generalized t-tests with Bonferroni correction[47] were used for post-hoc multiple comparisons of the normalized volume estimates between all combinations of scanner-dependent regression coefficients. The differences between the scanner-dependent regression coefficients were compared to the intra-subject standard deviation in order to assess the importance of the differences.

The linear mixed-effect model not only estimated the mean normalized volume for each condition defined through the fixed effects but also gave structure to the random effects and estimated of the variability due to scan-rescan, additional variability due to repositioning (R2), and additional variance due to a 2-week gap between a rescan (R3). An important consequence of formulating our problem in this form(Equation 3.3) was that we were able to distinguish

Chapter 3. Reproducibility of volumetric data acquired with a standardized T1-weighted brain-imaging protocol

subject-specific variance $var(y_i|\gamma_{i,0})$ and means $E(y_i|\gamma_{i,0})$ from marginal variance $var(y)$ and means $E(y)$.

The estimates of variance due to scan-rescan, additional variability due repositioning(R2), and additional variability due to a 2-week gap between a rescan (R3) were compared to the inter-subject variance to assess the importance of the random effects. It is important to note that the conditional subject-specific variance of uncertainties in measurements is given by the sum of the back-to-back scan-rescan variance and depending on the acquisition scenario either the variance due to repositioning or the variance due to the 2-week gap between a rescan.

3.3 Results

Figure 3.1 shows sample images from one subject scanned during R0 session from all scanner-*s*/protocols (4 in total) acquired in this study. For all of the shown images, the windowing was set to the same level. Upon visual inspection, it is difficult to see large differences between these four acquisitions. However, acquisitions from Prisma scanner have visibly higher SNR and CNR in the cortical region compared to the rest.

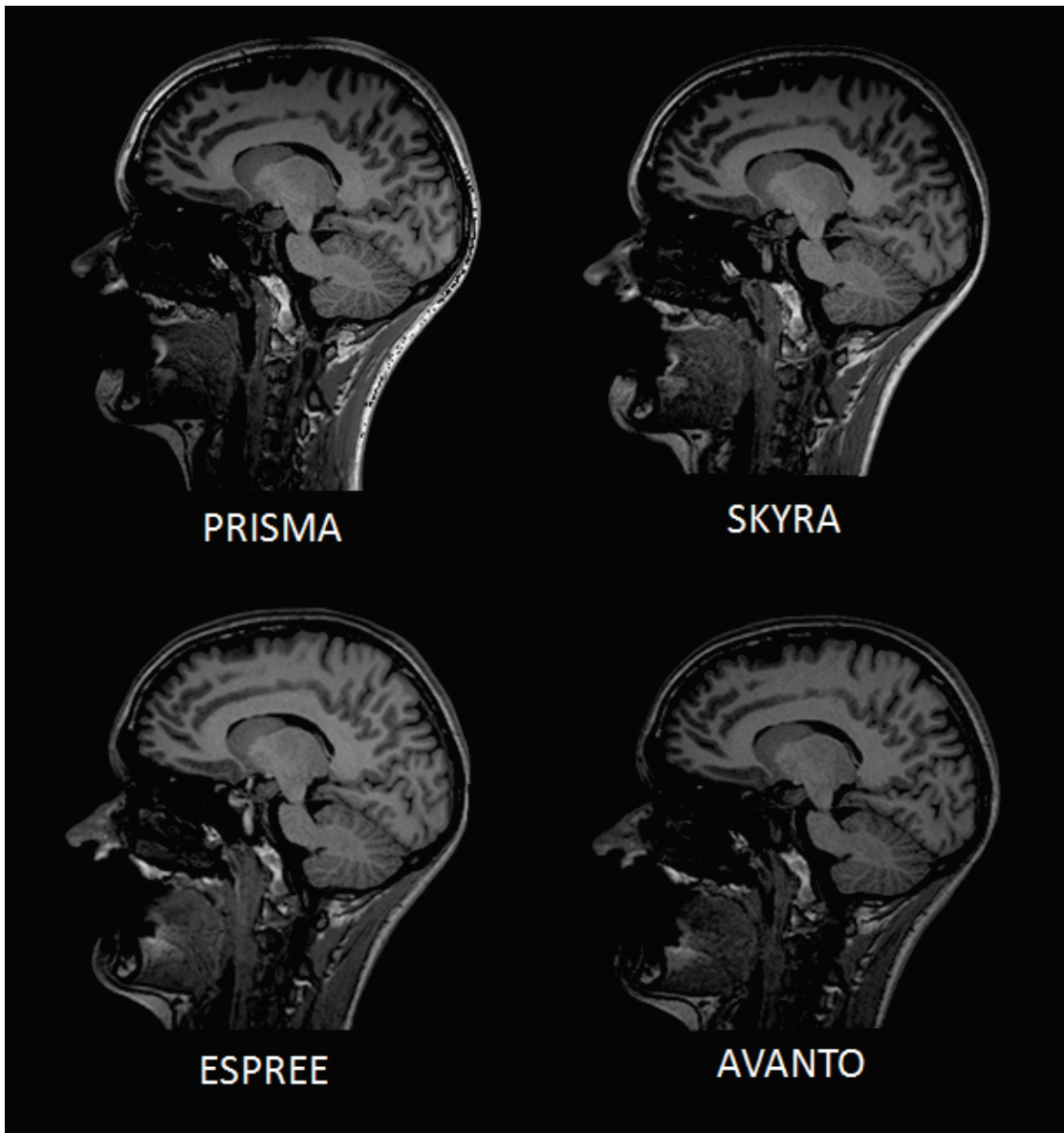


Figure 3.1: Images within one session with the same subject during R0 session.

Quantitative SNR assessment confirmed the visual inspection. SNR appeared to be well-matched on Avanto, Espree and Skyra scanners, but higher on the Prisma scanner for both GM

Chapter 3. Reproducibility of volumetric data acquired with a standardized T1-weighted brain-imaging protocol

and WM ($p < 0.05$, Figure 3.2). Mean SNR in WM: 38.64, 37.48, 49.84, 39.31; GM: 29.01, 28.95, 38.86, 28.19; Brain Stem: 23.30, 23.24, 21.71, 21.31 were observed on Avanto, Espree, Prisma, and Skyra scanners.

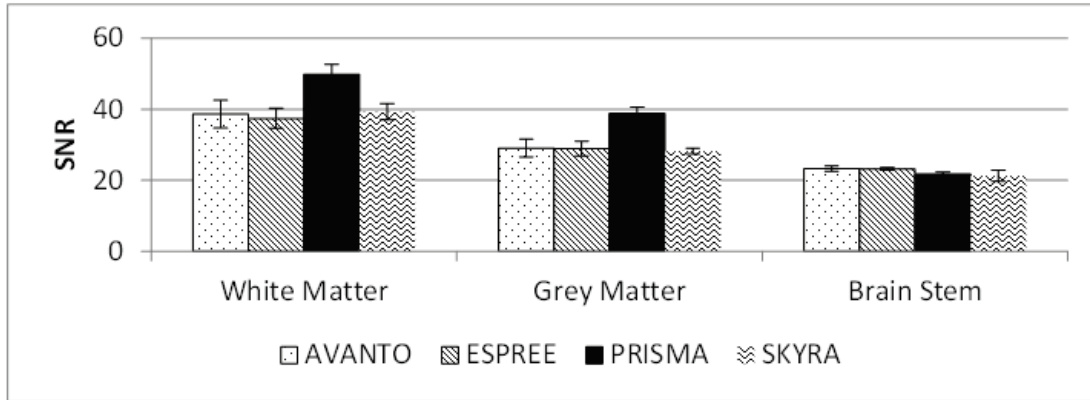


Figure 3.2: SNR for different scanner combinations using the segmentation masks from MorphoBox. Error bars show standard deviations across the subjects.

The results of the CNR assessment of all protocols/scanners used in this study are shown in Figure 3.3. Median GM-WM CNR 5.52, 5.23, 6.89, 6.63 and GM-CSF CNR 3.34, 3.25, 3.02, 3.01 were observed on Avanto, Espree, Prisma, and Skyra scanners. There were systematic differences in CNR (GM-WM and GM-CSF) between scanners of different field strength (Figure 3.3).

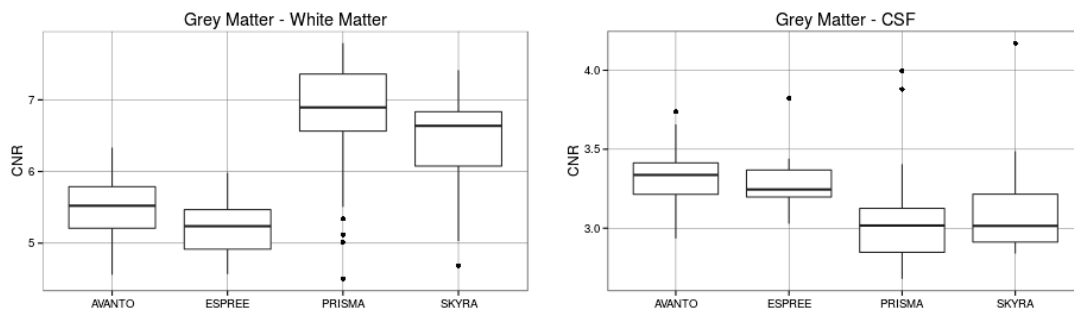


Figure 3.3: CNR for different scanner combinations (R0 session) using the segmentation masks from MorphoBox. Error bars show standard deviations across the subjects. The outliers are indicated as points.

The regression coefficients of fixed effects and estimates of the standard deviations of random effects fit by the linear mixed-effect model are shown in Figure 3.4 and Table A.1. Note that there is a significant effect of gender on the normalized volumes of grey matter, white matter and caudate. Males tend to have a significantly smaller percentage of grey-matter and caudate volumes whereas females tend to have a significantly smaller percentage of white-matter volume. The effect of age was found to be significant only for global grey matter volumes.

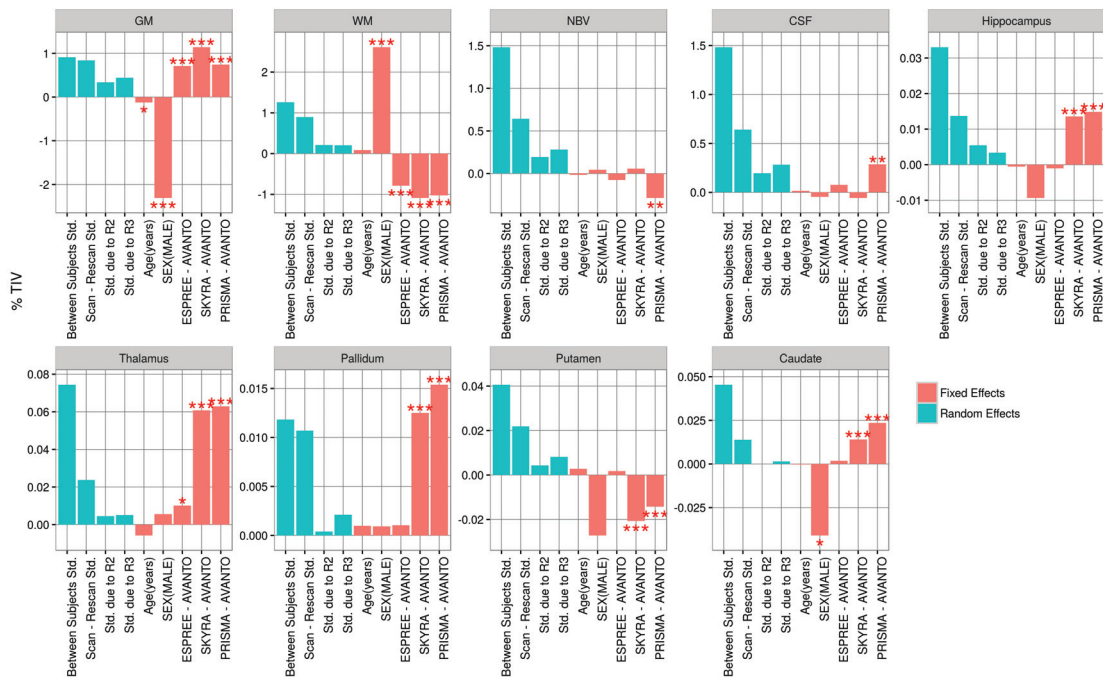


Figure 3.4: Regression coefficients as fit by the linear mixed-effect model. Fixed effects are shown in red. Random effects are shown in blue. * $p < 0.05$; ** $p < 10^{-2}$; *** $p < 10^{-3}$

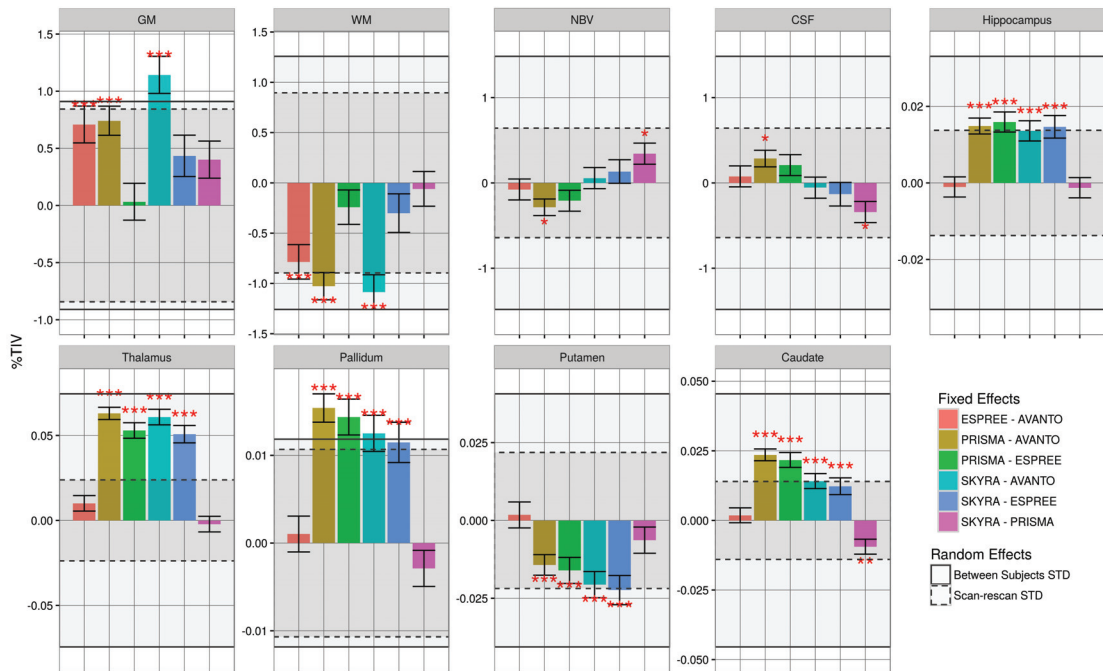


Figure 3.5: Volume differences for all possible scanner combinations. Each coloured bar represents a difference between scanner regression coefficients. The solid grey rectangle represents inter-subject variability (± 1 standard deviation). The dashed rectangle represents intra-subject back-to-back scan-rescan variability (± 1 standard deviations). * $p < 0.05$; ** $p < 10^{-2}$; *** $p < 10^{-3}$

Chapter 3. Reproducibility of volumetric data acquired with a standardized T1-weighted brain-imaging protocol

The post-hoc multiple comparison of the regression coefficients of each scanner is illustrated in Figure 3.5 and Table A.2. The differences in fixed effect regression coefficients represent a difference in mean estimates between two scanners and are in a percentage of the TIV units. Figure 3.5 also shows the inter-subject and scan-rescan standard deviations as estimated by the linear mixed-effect model and adjusted for the effects of scanner, age, sex, repositioning and 2-week gap between a rescan. The 1.5T - 3T differences constituted a considerable portion of inter-subject variability, i.e. 39% (caudate), 45% (hippocampus), 113%(pallidum), 45%(putamen), and 76%(thalamus) of the inter-subject standard deviation (solid grey area in Figure 3.5). Note that when compared to back-to-back scan-rescan variability, these differences were largest in the case of volume estimates of thalamus, and were of the order of scan-rescan variability for the rest of the small structures. However, they were mostly insignificant within the field-strength and constituted a small proportion of the inter-subject standard deviation. Results from GM, WM, NBV, and CSF tissues were heterogeneous in this respect.

Figure 3.6 illustrates the contribution of individual components to random effects (back-to-back scan-rescan, repositioning, and physiology) as fit by the model. The scan-rescan differences in back-to-back scanning scenario significantly contributed to the total variance and represented a significant proportion of between-subject variance for all of the investigated structures. Note that the scan-rescan variance of volume estimates constituted a large proportion of the inter-subject variance for pallidum, white matter and grey matter while scan-rescan variances of thalamus and caudate volumes were small compared to inter-subject variance.

In the case of repositioning between the measurements or a 2-week-gap between a rescan, the total variance of an uncertainty in the measurements for a given subject is the sum of the back-to-back scan-rescan variance(R_0) and the additional variance either due to repositioning(R_2) or 2-week-gap between a rescan (R_3). Both repositioning (R_2) and 2-week-gap between a rescan (R_3) did not significantly contribute to the total variability compared to back-to-back scans and between-subject variability. Note that compared to the other investigated structures, volumes of grey matter had a relatively large back-to-back scan-rescan variability as well as strong contribution from both repositioning and a 2-week gap between scanning sessions.

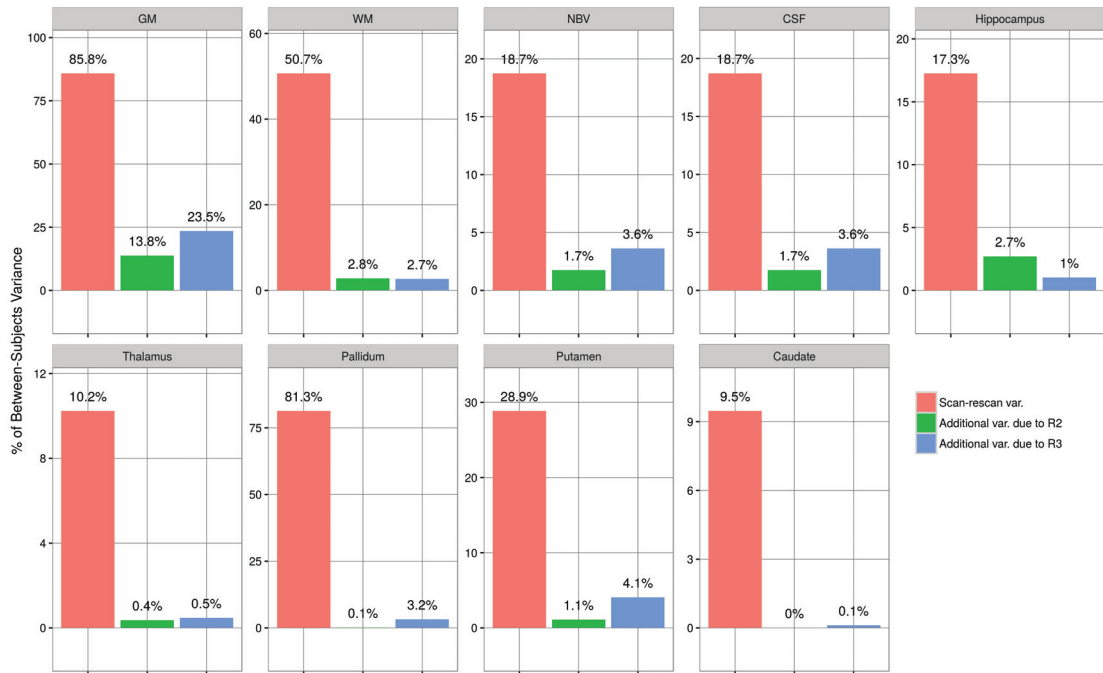


Figure 3.6: Random components of variability expressed as percentage of between-subject variance, scan-rescan (R1) - red, additional variance due to repositioning (R2) - green, additional variance due to 2-week-gap between rescan (R3) - blue.

3.4 Discussion

This work quantitatively characterised MPRAGE scans acquired with protocol parameters similar to ADNI and tested the impact of scanner's model, field strength, protocol parameters, shimming, and brain physiology on the reliability and reproducibility of volumetric brain assessment in clinical settings.

Consistent with theoretical considerations taken in ADNI protocol design, SNR appeared to be well-matched on Avanto, Espree and Skyra scanners. However, it appeared to be higher on the Prisma scanner for both GM and WM ($p < 0.05$, Figure 3.2), which can be presumably explained by the use of the 64-channel head coil (Prisma) vs. 12- and 20-channel coils otherwise. Note that GM-WM CNR appeared to be higher on 3 T systems compared to 1.5 T. This increase of GM-WM CNR on 3 T systems can be attributed to the small differences in the respective protocol parameters including the inversion times.

In this study, we decoupled systematic offsets to volume measurements due to hardware, and protocol effects from random offsets due to repositioning, and physiology using a linear mixed-effect model analysis.

There was a significant effect of gender on the normalized volumes of grey matter, white matter and caudate that was consistent with previously reported studies [48, 49]. However, the effect

Chapter 3. Reproducibility of volumetric data acquired with a standardized T1-weighted brain-imaging protocol

of age was only significant for grey matter. The absence of significant age effect for other structures can be attributed to the narrow age range of the subjects used in this study.

Systematic biases were observed in variability of volume measurements of the caudate nucleus, hippocampus, pallidum, putamen, and thalamus structures between 1.5T and 3T systems. However, they were mostly insignificant within field-strength. These systematic differences between 1.5T and 3T systems can be explained by the changes in contrast and differences in protocol parameters that, for example, differences in voxel size can give rise to different partial volume fractions. Total volumes of GM, WM and CSF were also found different across scanners of the same field strength. One of the factors responsible for this heterogeneity can be the system-specific residual bias-field left after application of B1 bias field correction [35]. Within small structures we do not expect to observe large variations of residual bias field and expect to have pure CNR/protocol effects.

In the context of patient management, an important finding is that overall the effects from repositioning (R2), and 2-week rescan (R3) scenarios did not significantly contribute to the total variability when compared to the back-to-back scan-rescan (R1) scenario. Note that scan-rescan standard deviation of volumes of thalamus structure was small compared to the 1.5T - 3T protocol differences and between-subject standard deviation. For the rest of small structures examined in this study, the scanner/protocol related effects were of the same order as scan-rescan variability. Comparatively smaller scan-rescan variability of the thalamus can be presumably explained by its more stable shape across subjects compared to other central nuclei that may facilitate robust automated segmentation.

3.5 Conclusion

In summary, without the use of sophisticated correction strategies, compatibility of serial and cross-platform scans can be best reached by minimizing differences in field, protocol, hardware, software, and physiology. Intra-subject scan-rescan variabilities were on average a factor of two smaller than the inter-subject variability and might be impacted by algorithmic instabilities, e.g. during registration (i.e. for small brain structures) and noise figures (e.g. for GM and WM) [50].

Two encouraging findings of this study were that (a) repositioning, re-shim and physiology conditions appeared as minor contributors to total variability in volumetric assessment and (b) protocols at equal field strength introduced an insignificant bias for the NBV, caudate nucleus, hippocampus, pallidum, putamen, and thalamus structures.

Significant biases found in this study appeared when field strengths, voxel size, contrasts differed and noise levels were modulated. The employed 1.5T and 3T protocols slightly differ in contrast and resolution – factors which cannot be disentangled using the data collected in the current study. However, since these effects are deterministic, correction strategies can be applied to enable comprehensive pooling of imaging data in cross-sectional studies or improved atrophy assessment in longitudinal studies.

4 Optimization of acquisition strategies and comparison of state-of-the-art T1-weighted whole-brain protocols

Recent advances in MRI hardware and acquisition methodology promise improved leverage of the MR signal and more benign artefact properties in particular when employing increased acceleration factors in clinical routine and research. In this chapter, four variants of a four-fold-accelerated MPRAGE protocol (2D-GRAPPA, CAIPIRINHA, CAIPIRINHA elliptical, and segmented MPRAGE) are quantitatively investigated based on clinical readings, basic image quality metrics (SNR, CNR), and results of automated brain tissue volume segmentations. The experiments were performed on a single system. The results are benchmarked against a widely-used two-fold-accelerated 3T ADNI MPRAGE protocol that was investigated in Chapter 3.

This chapter is adapted from already published work (Falkovskiy et al., 2016).

4.1 Introduction

The 3D magnetisation-prepared rapid acquisition gradient-echo (MPRAGE) pulse sequence [28] with Alzheimer’s Disease Neuroimaging Initiative protocol (ADNI-2) [17] parameters is a well-established standard for multi-site and longitudinal MRI studies that involve T1-weighted imaging of the human brain. Both in routine clinical and research settings, reduced MRI scan times are desirable for increased patient throughput, improved patient comfort, and better management of patient motion. The 3T imaging protocol of the ADNI employs conventional two-fold parallel imaging [29] to reduce the acquisition time of a whole-brain MPRAGE scan from 9 min in the original ADNI-1 protocol [16] to 5 min in ADNI-2. Recently, several strategies have been proposed to further reduce the acquisition time in 3D structural-brain-imaging protocols beyond conventional parallel imaging, e.g. 2D-GRAPPA [51], CAIPIRINHA [52, 53], and segmented MPRAGE [54].

Because of the excellent tissue contrast, the 3D T1-weighted MPRAGE images are appreciated for radiological reading. Moreover, quantitative assessment of brain tissues and the volume of individual brain structures has become an important tool in more research-oriented applications of the MPRAGE [5, 55, 56, 57, 12]. For instance, studies aiming at quantification of disease progression, drug efficiency or normal aging apply serial imaging to assess structural changes over time. Since the expected effect sizes in normal aging and disease (e.g. increased atrophy rates) are often subtle over time, it is vital to understand the reliability and reproducibility of the imaging-based quantitative measurements and to understand any inconsistencies that may appear when changing the pulse sequence between longitudinal repeat scans.

The reproducibility of the reference protocol (5-minute ADNI-2) has been studied extensively [18, 19, 20, 21, 22, 23, 24, 25, 26, 27] However, to our knowledge, the influence of further accelerations obtained by applying 2D-GRAPPA, CAIPIRINHA or segmented MPRAGE protocols on volumetric brain measurements and clinical readings has not been reported so far.

In this work, we aim to assess the reliability of the data obtained on the same platform across these accelerated protocols both qualitatively and quantitatively. The qualitative analysis is carried out by an experienced observer. The quantitative analysis is performed through assessing the reproducibility of volume measurements with automated brain segmentation algorithm [39, 12] contrast-to-noise ratio (CNR) and signal-to-noise ratio (SNR) measurements [41].

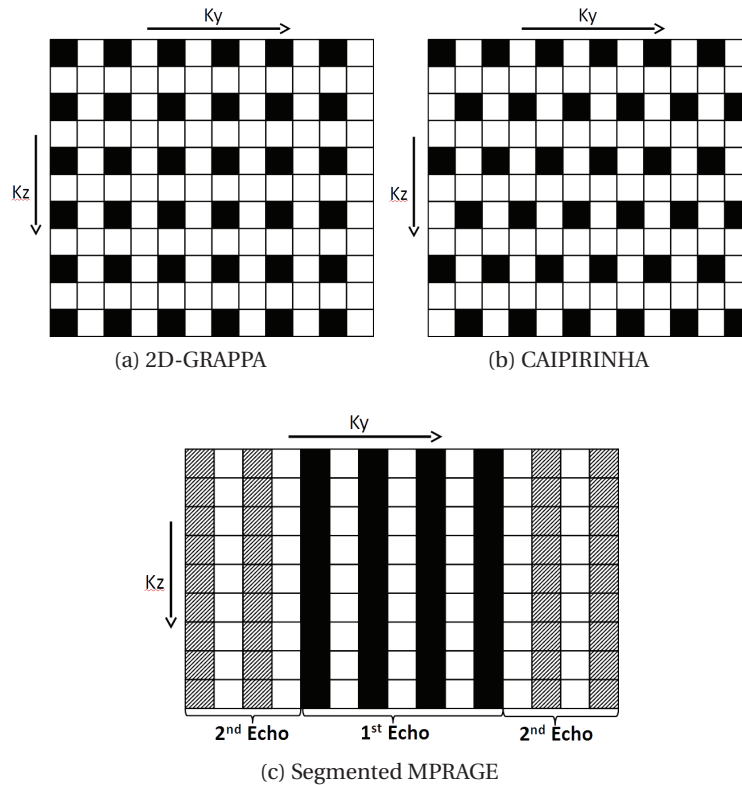


Figure 4.1: K-space sampling patterns for the accelerated MPRAGE variants. Solid squares represent acquired lines. Squares with stripes represent lines acquired with a second echo. (a) 2D-GRAPPA acceleration with 2x2 undersampling; (b) CAIPIRINHA with 2x2 undersampling and shift 1; (c) Assignment of echoes in segmented MPRAGE

4.2 Theory

Conventional GRAPPA- [29] or SENSE-type accelerations [30] are often used in clinical settings. Both methods obtain acceleration by undersampling of the k-space data in one phase-encoding direction (1D). In most cases, only moderate accelerations are employed due to increased aliasing and noise amplification (g-factor penalty) with higher acceleration factors. In order to overcome those limitations partly, and assuming that there is sufficient SNR, it is possible to further generalize the conventional GRAPPA approach to 2D for higher accelerations [51].

A modified undersampling pattern used in 2D-GRAPPA is illustrated in Figure 4.1a. 2D-GRAPPA allows for a more effective exploitation of the coil sensitivities of contemporary high-channel-count receive arrays which vary in both phase-encoding directions, resulting in a reduced g-factor compared to 1D-GRAPPA with the same total acceleration. This advantage can be further improved by using the CAIPIRINHA [52, 53] approach, which shifts the aliasing pattern by introduction of an additional CAIPIRINHA shift parameter (Figure 4.1b).

Chapter 4. Optimization of acquisition strategies and comparison of state-of-the-art T1-weighted whole-brain protocols

In addition to parallel imaging techniques, a common approach to reduce scan time is to use elliptical scanning [58]. Elliptical scanning does not sample the corners of k-space and hence reduces the acquisition time.

A segmented MPAGE approach accelerates the acquisition using a different method which does not rely on coil sensitivity variations in contrast to parallel imaging techniques [54]. The standard acquisition scheme is modified to allow acquisition of multiple echoes with bipolar gradient readouts. Between the readouts, phase-encoding blips are inserted so that multiple portions of k-space are sampled per excitation. The echoes are grouped so that the first echoes are acquired in the centre of k-space to ensure the intended contrast (Figure 4.1c). A phase correction algorithm similar to the one used for EPI acquisitions is applied to remove phase inconsistencies between k-space lines acquired in the different echoes with differing polarity. This correction is computed based on the lines acquired in the centre of k-space. Notably, this segmented approach can be introduced in addition to parallel-imaging-based acceleration.

4.3 Materials and Methods

4.3.1 MR Acquisition

All experiments were performed on a standard clinical 3T MRI (MAGNETOM Skyra, Siemens Healthcare, Erlangen, Germany) equipped with a 32-channel head coil array. The measurement protocol consisted of five 3D MPAGE volume acquisitions with protocol parameters similar to the ADNI-2 MPAGE protocol settings (TR/TI= 2300/900 ms, $\alpha = 9$ deg., BW=240 Hz/pixel, readout in superior-inferior direction) [17] but with 1 mm isotropic resolution at a FoV of 256x240x176 mm^3 : (a) T1w ADNI-2 protocol with 2-fold GRAPPA acceleration (TA=5:12 min) [17] which is used as reference and that we will refer to as reference ADNI-2; (b) 2D-GRAPPA with 4-fold (2x2) undersampling (TA=2:59 min) [51]; (c) CAIPIRINHA with 4-fold undersampling (2x2 shift 1 in z direction) (TA=2:59 min) [52]; (d) CAIPIRINHA same as c but with additional elliptical scanning option, further reducing scan time (TA=2:40 min) [52]; (e) segmented MPAGE with 4-fold acceleration based on combining conventional parallel imaging (2-fold) and a two-echo segmented acquisition (BW=480 Hz/pixel, TA=3:15 min) [54]. All 1-D undersampled scans utilized 32 reference lines for reconstruction of the GRAPPA weights, while the 2-D undersampled scans utilized a 32x32 lines reference region.

For the segmented MPAGE acquisition, the bandwidth was increased to maintain the echo spacing equal to the one of the reference ADNI-2 protocol. With the exception of the reference T1w ADNI-2 measurement, all scans were acquired using in-house prototype sequences. The order of the scans was randomized between sessions.

22 subjects (12 male and 10 female, age 20-44 years, mean 30 ± 6.0 years) were imaged. The subjects were screened prior to enrolment in this study based on a health-assessment questionnaire. Only the subjects in good health and without a history of neurological diseases were considered. The exclusion criteria included: any known active medical conditions;

hypertension; excessive smoking; excessive alcohol consumption; drug addiction; diabetes mellitus; history of significant vascular events (i.e., myocardial infarction, stroke or peripheral vascular disease); history of malignant neoplasia of any form; history of cardiac, lung, liver, or kidney failure; active or inadequately treated thyroid disease; active neurological or psychiatric conditions; history of head trauma with loss of consciousness. After obtaining written consent, each subject was scanned with the five MPRAGE variants within one session under an IRB-approved protocol. FOV placement was guided by the scanner-integrated AutoAlign feature. Intensity normalization was performed using a scanner software-integrated pre-scan procedure and an adaptive coil combination mode [38] was used to reconstruct the image volumes for all of the protocols used in this study.

4.3.2 Image Processing Pipeline and Analysis

Qualitative

The data were graded by an experienced image analyst from the ADNI MR Core. The analyst was blinded with respect to protocols. First, each volume was rated with respect to the presence of motion artefacts (none, mild, severe) and overall image quality on the scale of 1-3 based on the presence of blurring/ringing/ghosting or other artefacts (1: good, 2: fair, 3: very low quality). No separation on the nature of artefacts that contributed to overall image quality score was made.

Second, following this ranking, if two or more image volumes were assigned the same image quality grade, image volumes were presented side by side and the observer made a subjective decision which image had higher image quality relative to each-other. In this fashion, relative rankings within a session were constructed with a relative scale (1: best, 5: worst). The grade of 1 is assigned to the volume with the best image quality in the session and 5 to the volume with the worst image quality relative to the other scans in the session. Note that it should be considered as a limitation of these qualitative rankings the fact that the observer did not have to justify his choice, and only a single observer was used. These scores were averaged across the sessions, and statistical difference between the protocol variants was tested with the Wilcoxon rank-sum test [40] since we did not expect scores to be normally distributed.

Quantitative

The data was passed through the in-house-developed automatic segmentation framework MorphoBox [39, 12] to compute the volumes of a number of brain tissues and structures: total intracranial volume (TIV), grey matter (GM), cortical grey matter (cGM), white matter (WM), hippocampus, thalamus, caudate, putamen, pallidum and brain stem. Label masks that define voxels belonging to a particular tissue type or structure were also computed. In order to assess the consistency of the volumetric observations, intra-class correlation coefficients (ICC) [59] between all possible combinations of protocol variants were computed. The model used in

Chapter 4. Optimization of acquisition strategies and comparison of state-of-the-art T1-weighted whole-brain protocols

this study for ICC computation treated subjects as randomly sampled from a larger population and the acquisition protocols employed in this study as fixed parameters.

In order to evaluate a potential systematic bias in the volumetric results, normalized volumetric differences (D) between the reference protocol (V_r) and each variant (V_v) were computed for each structure as:

$$D(V_r, V_v) = \frac{(V_v - V_r)}{(V_r + V_v)}, \quad (4.1)$$

where $D(V_r, V_v)$ is in the range $[-1,1]$. Normalized volumetric differences were averaged across the subjects. All possible choices of the reference protocol (V_r) were examined. The statistical significance of the difference from a zero median in normalized volumetric differences was tested using Wilcoxon signed-rank test [40] because the differences were not expected to be normally distributed.

Based on the ICCs and normalized volume differences, the most consistent accelerated protocol was selected when compared to the reference ADNI-2 protocol. For this protocol, the normalized volume differences were recomputed using the methodology described in [21] and then compared to the scan-rescan reproducibility study of reference ADNI-2 protocol [21].

The gold-standard method for SNR computation is the multiple-replica approach that consists of imaging the same object several times. Voxel-wise standard deviations of noise (μ_{noise}) and mean intensities (μ_{signal}) are estimated from those measurements and voxel-wise SNR maps ($\mu_{signal}/\sigma_{noise}$) are constructed. However, the multiple-replica approach is not practical for calculating SNR maps for in-vivo imaging due to patient motion and excessive measurement time. For this reason, pseudo multiple-replica approach [41] was used instead, as it just requires knowledge of the noise correlation between coils. It allows the calculation of SNR maps from a single acquisition by mimicking the multiple replica method. First, synthetically generated random noise is scaled and correlated across coils based on the receive coil covariance matrix. The noise covariance matrix was measured with the subject and coil setup unchanged, by acquiring 128 readouts having 512 sampling points each, including oversampling, without an excitation pulse. This correlated and scaled noise is injected into the data at the beginning of the image reconstruction pipeline to produce a stack of images with different noise. The “true” acquisition noise from the subject is still present in this synthetic data and is assumed to be a part of the signal. From this stack of images, the voxel-wise means (μ_{signal}) and standard deviations (σ_{noise}) are computed, and SNR maps ($\mu_{signal}/\sigma_{noise}$) are produced.

The SNR computations were performed for all acquired 3D MPRAGE volumes with 8 pseudo-replicas. To improve the estimation of σ_{noise} , a 2D 5x5 region of neighboring pixels was used to estimate σ_{noise} in a moving-average fashion according to [42].

Individual SNR maps were spatially normalized [60] to the reference ADNI-2 image volume within the same session. Using the label maps from the reference ADNI-2 datasets, mean SNR

values were computed for the following brain tissue types and structures: brain stem (chosen due to its proximity to the centre of the head coil; representing the smallest expected SNR), hippocampus, white matter (representing an average of expected SNR), and cortical grey matter (chosen due to its proximity to the coil elements; representing the biggest expected SNR). The differences in SNR of different protocol variants were tested with the Wilcoxon rank-sum test [40] with Bonferroni correction for multiple comparisons [47].

SNR efficiency can be defined as: SNR/\sqrt{TA} , where TA is total acquisition time. SNR efficiency was computed for all protocol variants and differences were tested with the Wilcoxon rank-sum test [40] with Bonferroni correction for multiple comparisons [47].

In order to assess the contrast-to-noise ratio (CNR), data were spatially normalized [60] to the reference ADNI-2 image volume within session. In addition to intensity normalization performed using a scanner software-integrated pre-scan procedure, a separate B1 bias field correction was applied to all of the datasets [35]. This additional bias field correction was shown to be very robust in 3T settings. CSF, WM and GM were masked using the label masks from the reference ADNI-2 image volumes. Two CNR values were computed as follows:

$$CNR_{csf-gm} = \frac{(\mu_{gm} - \mu_{csf})^2}{(\sigma_{csf}^2 + \sigma_{gm}^2)} \quad (4.2)$$

$$CNR_{gm-wm} = \frac{(\mu_{wm} - \mu_{gm})^2}{(\sigma_{gm}^2 + \sigma_{wm}^2)} \quad (4.3)$$

where μ_{csf} , μ_{gm} , μ_{wm} are mean intensities and σ_{csf}^2 , σ_{gm}^2 , σ_{wm}^2 are image intensity variances within CSF, GM and WM volumes. The differences in CNR of different protocol variants were tested with the Wilcoxon rank-sum test [40] with Bonferroni correction for multiple comparisons [47].

CNR efficiency can be defined as: CNR/\sqrt{TA} , where TA is total acquisition time. CNR efficiency was computed for all of the protocols used in this study and differences were tested with the Wilcoxon rank-sum test [40] with Bonferroni correction for multiple comparisons [47].

It is expected that the SNR will be decreased in accelerated acquisitions when compared to the reference ADNI-2 protocols, to a first approximation, proportionally to the square-root of the acceleration factor. The decrease in SNR can be attributed to the lower total acquisition time. A further factor is the reconstruction-induced noise amplification described by the g-factor. Furthermore, the different protocol variants can induce a different signal modulation in k-space.

In terms of GM-WM and CSF-GM contrast, the inversion recovery curve with the segmented MPRAGE acquisition is mapped along the partition-encoding axis, and undersampling is performed along the phase-encoding axis. Therefore, the resulting contrast is not expected to change compared to the standard MPRAGE. In 2D-GRAPPA and CAIPIRINHA acquisitions,

Chapter 4. Optimization of acquisition strategies and comparison of state-of-the-art T1-weighted whole-brain protocols

the inversion recovery curve is mapped mostly along the partition-encoding, and the changes in contrast are consequently expected to be minimal [52, 61]. We expect that the changes in CNR using accelerated acquisitions are driven by changes in noise (due to different sampling durations between the reference and the four-fold accelerated scans) rather than by changes in contrast. Therefore, the last part of our image processing pipeline only addresses the noise-related effects on the results obtained with MorphoBox.

To study the noise dependency of the segmentation results, the following numerical experiment was designed: scaled noise was added to the raw data of the reference ADNI-2 protocols to mimic the SNR performance of the accelerated protocols. Based on the SNR measurements within-session, the average SNR value was calculated between 4-fold accelerated protocols. The level of noise added to ADNI-2 data was set to match the average SNR performance of 4-fold accelerated protocols on a per session basis. These synthetic data were reconstructed and then passed on to the automated segmentation framework. This procedure was repeated 32 times for each subject to produce 32 synthetic image volumes and segmentation results per subject.

The subsequently obtained volumetric results from synthetic data of the different brain structures, and tissue types were pooled from all the subjects. ICCs between synthetic volumetric results were used as a measure of consistency of data. The model used to assess the consistency of the synthetic noise experiment considered subjects to be randomly sampled from a larger population of subjects and 32 repetitions with additional noise to be judges that are sampled from a bigger population of judges [59]. Normalized volumetric differences between synthetic data and the reference ADNI-2 protocol were calculated to assess whether there is a systematic bias in the measurements. The statistical significance of the difference in normalized volumetric differences was tested using the Wilcoxon signed-rank test [40].

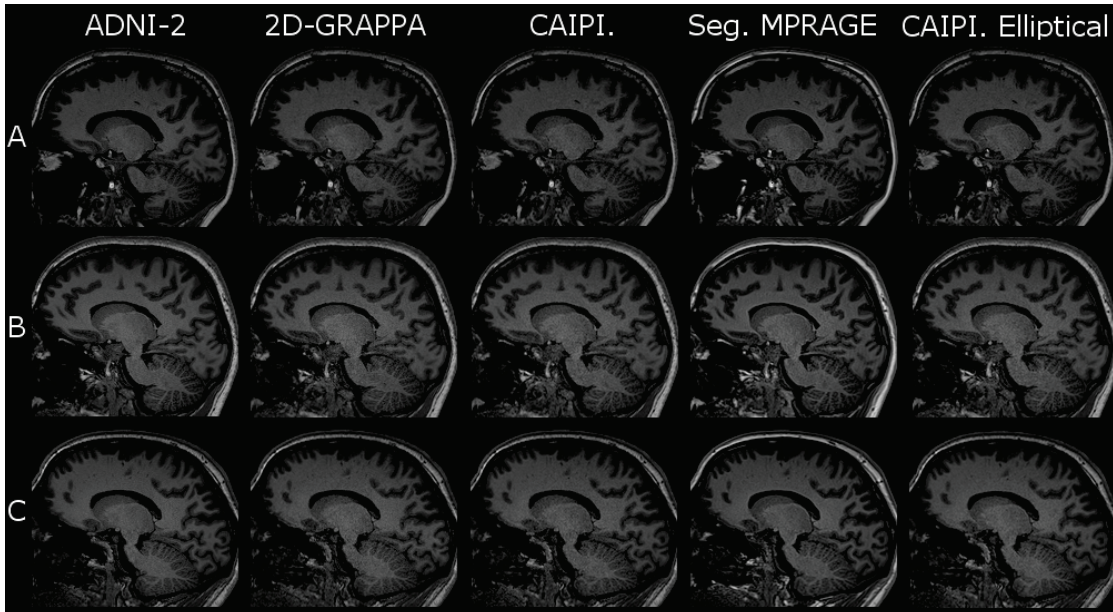


Figure 4.2: Sample images showing representative sagittal views of all protocol variants for 3 subjects (A, B, C). Please note the increased noise level of the accelerated protocols that is most visible in the brain stem region of the subjects.

4.4 Results

Each row in Figure 4.2 demonstrates images from the same subject (three shown in total) obtained within one session using the reference ADNI-2 protocol and the 3-minute accelerated protocol variants. The windowing was set to the same level for all shown images. In the brain stem region, there is a visible increase of noise in accelerated protocols when compared to the reference ADNI-2 protocol.

4.4.1 Observer Ratings

All raw images before noise injection were rated by the observer as diagnostically useful with high image quality (76 image volumes) or mild image quality issues (34 image volumes). Segmented MPRAGE acquisitions exhibited, in some cases, mild ringing artefacts largely arising from hyper-intense signals due to abundant fat in the neck. No systematic pattern of image quality issues was reported by the observer. Overall, qualitative observations yield that all protocols provide clinically useful image quality.

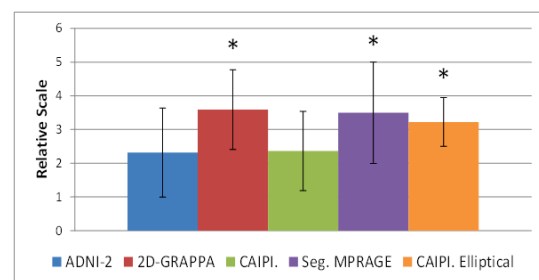


Figure 4.3: Observer Rankings. Relative scale (1, best to 5, worst). * indicates significant difference from the ADNI-2 protocol.

Chapter 4. Optimization of acquisition strategies and comparison of state-of-the-art T1-weighted whole-brain protocols

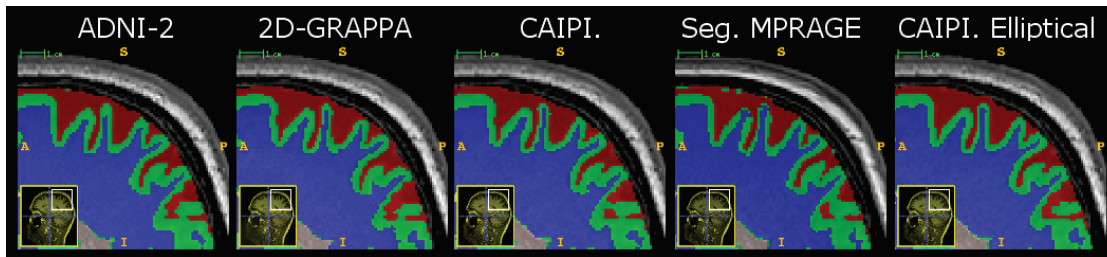


Figure 4.4: Segmentation results from a single subject within one session. Blue: white matter; Green: grey matter; Red: CSF.

The mean observer rankings (1, best to 5, worst) are shown in Figure 4.3. Note that the 3-minute CAIPIRINHA MPRAGE scans are perceived on average of identical value for radiological reading as the reference 5-minute ADNI-2 scans despite a subtle, but visible noise degradation. However, the other protocols appear statistically different ($p < 0.05$), as assessed by the Wilcoxon rank-sum test.

4.4.2 Volumetric Measurements

Visually, the segmentation results from all protocols used in this study display a high degree of similarity (Figure 4.4). The segmentation errors are at the single voxel level, and therefore hard to visually notice. Visual inspection suggests that most of these differences occur in regions affected by partial-volume effects. All segmentation results were visually inspected for large segmentation errors and no gross segmentation errors have been observed in this study. It is important to note that MorphoBox was run in a fully automated fashion and no manual editing was applied at any stage of the segmentation process.

Intra-class correlation analysis shows highly consistent volumetric measurements of accelerated protocols when compared to the reference ADNI-2 protocol (Table 4.1).

Figure 4.5 illustrates the intra-class correlation coefficients between all protocols used in this study. Each column represents an ICC between two protocol variants and each row represents a structure. The CSF volumes in segmented MPRAGE acquisitions exhibit smaller correlation coefficients relative to the other protocol combinations. Pallidum intra-class correlation coefficients are systematically smaller than the correlation coefficients of other structures.

4.4. Results

	ADNI-2 vs. 2D-GRAPPA	ADNI-2 vs. CAIPI	ADNI-2 vs. Seg. MPRAGE	ADNI-2 vs. CAIPI. Elliptical
TIV	1.00	1.00	0.99 [0.98-1.00]	1.00
Cortical Grey Matter	0.96 [0.90-0.98]	0.97 [0.93-0.99]	0.93 [0.84-0.97]	0.97 [0.92-0.99]
White Matter	0.99 [0.98-1.00]	0.99 [0.97-1.00]	0.99 [0.97-1.00]	1.00 [0.99-1.00]
Hippocampus	0.94 [0.85-0.97]	0.95 [0.88-0.98]	0.94 [0.87-0.98]	0.94 [0.86-0.97]
Thalamus	0.98 [0.94-0.99]	0.98 [0.96-0.99]	0.97 [0.93-0.99]	0.97 [0.94-0.99]
Putamen	0.96 [0.90-0.98]	0.97 [0.94-0.99]	0.95 [0.89-0.98]	0.95 [0.88-0.98]
Caudate	0.98 [0.95-0.99]	0.99 [0.97-0.99]	0.98 [0.94-0.99]	0.98 [0.94-0.99]
Pallidum	0.89 [0.76-0.95]	0.94 [0.86-0.97]	0.88 [0.74-0.95]	0.85 [0.68-0.94]

Table 4.1: Intra-class correlation coefficients (95% confidence interval shown in brackets).

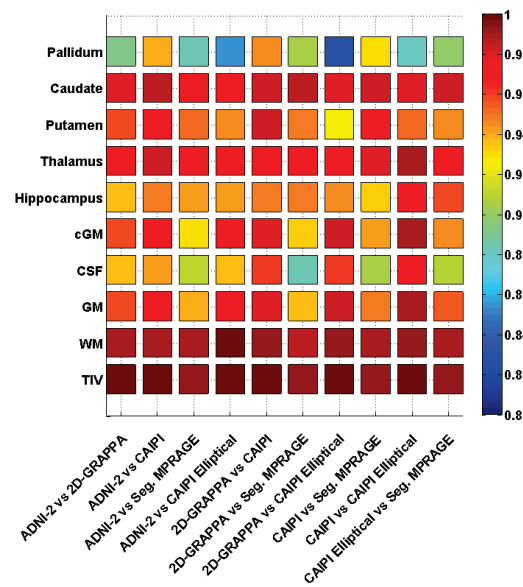


Figure 4.5: Intra-class correlation coefficients for different brain structures between all possible protocol combinations.

Chapter 4. Optimization of acquisition strategies and comparison of state-of-the-art T1-weighted whole-brain protocols

Figure 4.6 summarizes the outcome of the analysis of normalized volumetric differences when taking computed volumes from the reference ADNI-2 protocol as a reference volume (V_r). The label “ADNI-2+NOISE” corresponds to a synthetic dataset generated through addition of noise to match the SNR of accelerated protocols. There is a strong indication that the increase in noise levels introduces a systematic bias to the computed volumes of some structures when compared to the reference volume of the ADNI-2 protocol. Most apparent, a statistically significant increase in the white matter volumes in the accelerated acquisitions and a trend towards decreased cortical grey matter volumes were observed in this investigation.

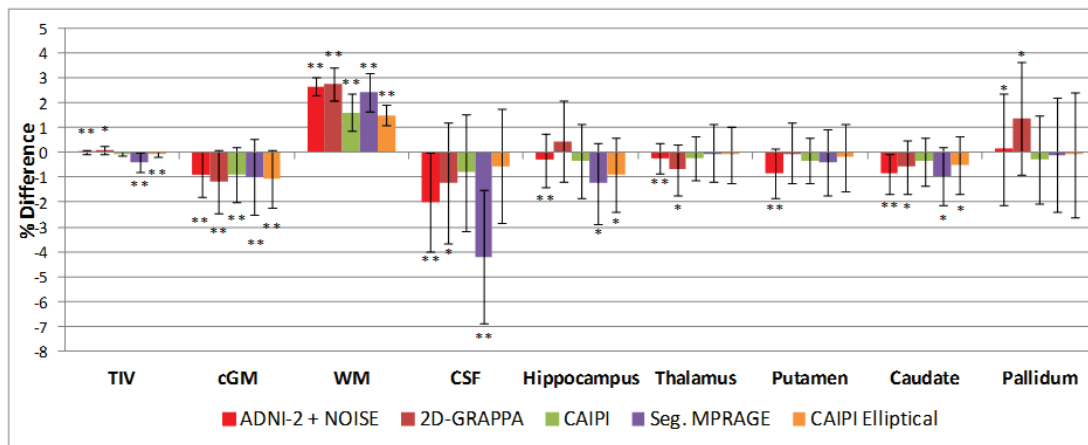


Figure 4.6: Volumetric percent difference with ADNI-2 protocol used as reference scan (V_r). * indicates difference from 0% median at the% significance level. * $p < 10^{-2}$; ** $p < 10^{-4}$;

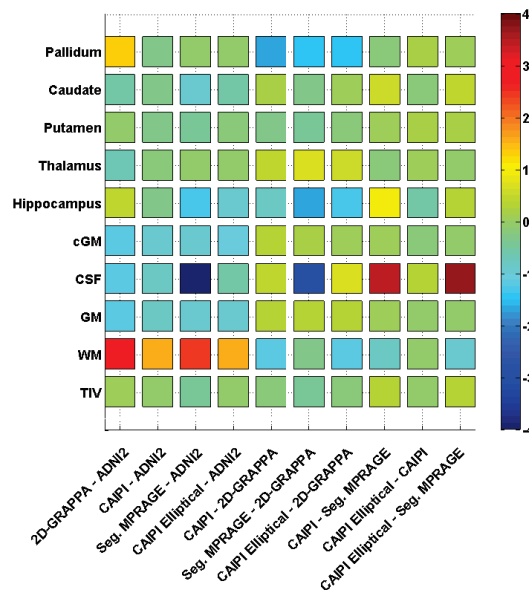


Figure 4.7: Volumetric percent difference with all possible choices of reference protocol (V_r).

Figure 4.7 summarizes the outcome of the normalized volumetric differences analysis for all possible choices of the reference volume (V_r). Each column represents the outcome of computing the normalized volume difference between two protocol variants used in this study and averaging it across the subjects. The results indicate that the segmented MPRAGE acquisitions introduce a change to estimated CSF volumes when compared to other accelerated protocol variants.

4.4.3 Signal-to-Noise Ratio (SNR)

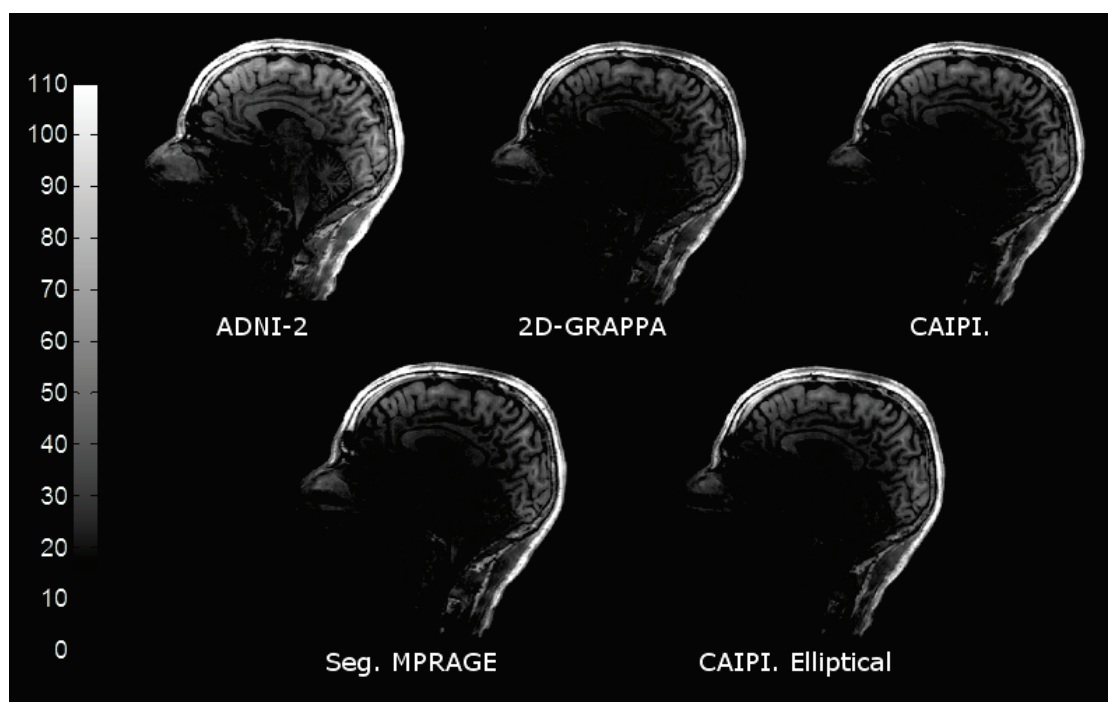


Figure 4.8: Voxel-wise SNR maps obtained using the pseudo multiple-replica method for a single subject within one session.

Figure 4.8 demonstrates sagittal SNR maps computed with the pseudo multiple-replica method [24] for a single subject within one session. Theoretically, a decrease in SNR proportional to $1/\sqrt{2}$ (reducing the acquisition time from 5 min from 3 min) is expected in the accelerated acquisitions when compared to the reference ADNI-2 protocol. Obtained SNR maps demonstrate the expected decrease of SNR in the accelerated protocols.

The results of the SNR analysis performed within different brain sub-structures are summarized in Figure 4.9. Mean SNR values of 50 (white matter), 29 (cortical grey matter), 22 (brain stem) and 20 (hippocampus structures) were observed, using the conventional ADNI-2 protocol. There is a statistically significant drop in SNR between ADNI-2 and four-fold-accelerated protocols. SNR dropped consistently by around 34%, 36%, 31%, and 34% for the accelerated variants: 2D GRAPPA, CAIPIRINHA, segmented MPRAGE and CAIPIRINHA with elliptical

Chapter 4. Optimization of acquisition strategies and comparison of state-of-the-art T1-weighted whole-brain protocols

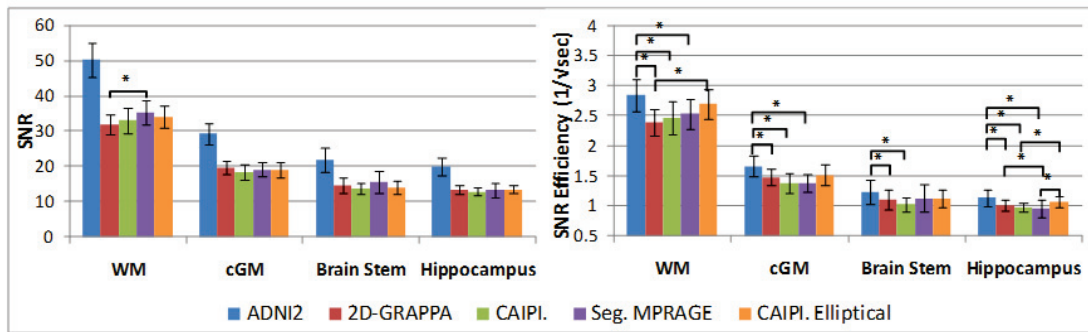


Figure 4.9: SNR measurement for different subjects. Error bars show the standard deviations across the subjects. * indicates significant difference between two protocols.

scanning. White matter SNR is significantly different between 2D-GRAPPA and segmented MPRAGE acquisitions. The rest of the four-fold-accelerated protocols are not significantly different from each other.

In terms of SNR efficiency, Figure 4.9 illustrates performance of the protocols used in this study and significant differences. With the conventional ADNI-2 protocol, average SNR efficiency values of 28 (white matter), 17 (cortical grey matter), 12 (brain stem) and 11 (hippocampus structures) were observed. Noteworthy is a decrease in SNR efficiency of 2D-GRAPPA, CAIPIRINHA, and segmented MPRAGE acquisitions in white matter, cortical grey matter and hippocampus regions when compared to the ADNI-2 protocol. CAIPIRINHA with elliptical scanning was not statistically different from the ADNI-2 protocol in terms of SNR efficiency. In the brain stem region, only 2D-GRAPPA and CAIPIRINHA appeared to be statistically different. On average, SNR efficiency decreased by 12%, 15%, 13% percent for 2D-GRAPPA, CAIPIRINHA, and segmented MPRAGE acquisitions when compared to ADNI-2 protocol.

4.4.4 Contrast-to-Noise Ratio (CNR)

The results of the CNR assessment of all protocols used in this study are shown in Figure 4.10. The greater the value of CNR, the better is the separation between the intensity distributions of the tissues, which simplifies the segmentation problem. Analogously to the SNR measurements, contrast to noise is decreased in accelerated acquisitions. There is a statistically significant decrease in CNR between ADNI-2 and four-fold-accelerated protocols. On average, CNR in the conventional ADNI-2 protocol was found to be 4.47 and 7.16 for CSF-GM and GM-WM, respectively, which significantly decreased on average by 29%, 26%, 38% and 28% in the 2D-GRAPPA, CAIPIRINHA, segmented MPRAGE, and elliptical CAIPIRINHA approaches. Both CSF-GM and GM-WM CNR appears to be significantly different between segmented MPRAGE and other four-fold-accelerated variants. The rest of the four-fold-accelerated protocols are not significantly different from each other.

Figure 4.10 illustrates CNR efficiency of the protocols used in this study. CNR efficiency in the conventional ADNI-2 protocol was found to be 0.25 and 0.41 for CSF-GM and GM-WM. Segmented MPRAGE appeared to be statistically different from the rest of the protocols in terms of CSF-GM and GM-WM CNR efficiency. 2D-GRAPPA is statistically different from the ADNI-2 protocol in terms of GM-WM CNR efficiency.

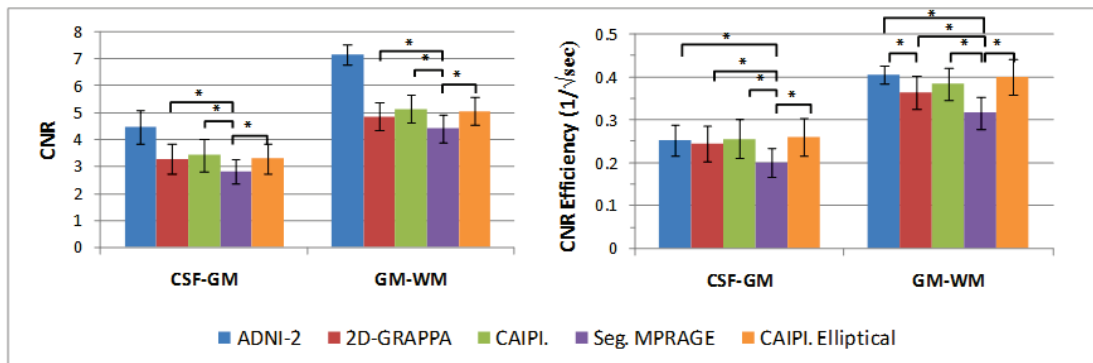


Figure 4.10: CNR averaged across subjects. Error bars indicate the standard deviations across the subjects. * indicates significant difference between protocols.

4.4.5 Noise Effects

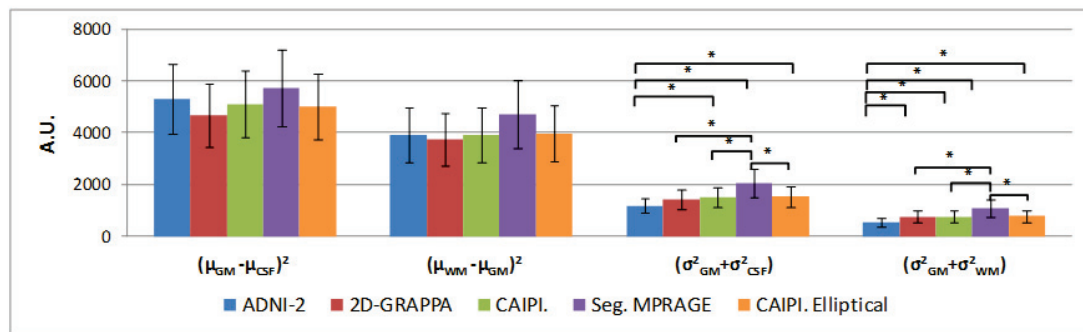


Figure 4.11: Mean distance between the peaks of the intensity distributions and their variance for the different protocol variants. * indicates significant difference between protocols.

In order to understand whether changes in the noise or signal properties cause the observed changes in SNR and CNR, Figure 4.11 illustrates the mean changes in the distance between the peaks of intensity distributions of the tissues and the sum of variances of those tissues as extracted by the MorphoBox. A statistically significant increase in sums of variances of $\sigma^2_{csf} + \sigma^2_{gm}$ and $\sigma^2_{wm} + \sigma^2_{gm}$ between ADNI-2 and CAIPIRINHA, segmented MPRAGE, and elliptical CAIPIRINHA acquisitions is observed. There is a statistically significant increase in $\sigma^2_{wm} + \sigma^2_{gm}$ sums of variances between ADNI-2 and 2D-GRAPPA acquisitions. Both sums of variances, $\sigma^2_{csf} + \sigma^2_{gm}$ and $\sigma^2_{wm} + \sigma^2_{gm}$ appear to be significantly different between segmented

Chapter 4. Optimization of acquisition strategies and comparison of state-of-the-art T1-weighted whole-brain protocols

TIV	1.00	Thalamus	0.99 [0.99-1.00]
Cortical Grey Matter	0.98 [0.97-0.99]	Putamen	0.97 [0.95-0.99]
White Matter	1.00	Caudate	0.99 [0.99-1.00]
Hippocampus	0.99 [0.98-0.99]	Pallidum	0.93 [0.88-0.97]

Table 4.2: Intra-class correlation coefficients (95% confidence interval shown in brackets) for the numerical experiment with addition of noise to the reference ADNI-2 scan.

MPRAGE and other four-fold-accelerated variants. Changes in the distance between the peaks of the intensity distributions were not significantly different.

4.4.6 Numerical experiment with addition of synthetic noise

For each subject, 32 volumes of synthetic data were computed by adding synthetic noise to the reference ADNI-2 raw data to reduce the SNR by approximately 35%, i.e. to match the SNR of the four-fold accelerated scans.

Table 4.2 illustrates ICCs between the segmentation results of synthetic image volumes. For all examined structures, ICC values greater than 0.93 were observed. The estimated ICC of pallidum volumes exhibited the smallest correlation coefficient of 0.93 when compared to other structures. Introduction of artificial 35% noise amplification to the reference ADNI-2 data introduced a bias in the volumetric data similar to the bias level observed with the accelerated acquisitions (Figure 6). A statistically significant increase in the measurements of white matter volumes and a trend towards decreased cortical grey matter volumes are evident when compared to the reference ADNI-2 volumes.

4.5 Discussion

In longitudinal studies that use morphometric assessments of brain tissues and structures, the choice of the imaging protocol can potentially influence qualitative readings and degrade reproducibility of serial automated brain segmentations. The main objective of this study was to investigate the impact that accelerated protocols (2D-GRAPPA, CAIPIRINHA, segmented MPAGE, and CAIPIRINHA with elliptical scanning) have on the observer's qualitative readings and on an automated brain segmentation procedure.

Qualitative analysis demonstrated that all of the data obtained using the accelerated 3-minute protocols have clinical value, even if artefacts are present and despite the changes in noise distribution. This observation can be explained by the fact that radiologists are used to "reading through" artefacts and indicates that, in these settings (3T with a 32-channel head coil), the "gold-standard" reference ADNI-2 protocol provides an SNR value that is sufficiently high. It is important to note that the radiological readings of the CAIPIRINHA scans were perceived to provide on average the same diagnostic image quality as the reference ADNI-2 scans despite a 36% reduction in SNR. This may further indicate that the images obtained with the conventional ADNI-2 protocol using a 32-channel coil operate in an SNR regime and/or have the artefact-to-noise ratio that provides an image quality beyond the minimum routine reading requirements. Also, the shorter scan time can reduce motion sensitivity and may lead on average in the longer scans to some subtle degradation of the overall image quality.

However, when volumetric data acquired with different protocols are pooled in longitudinal studies, one has to be cautious of potentially introducing a bias to any quantitative analysis. In our analysis, we found such a bias of the order of 2% in white matter and of the order of 1 % in cortical grey matter volumes when using the accelerated protocols.

If we look at the specific structures, for instance, we found that the hippocampus volume assessment differs most when comparing the reference ADNI-2 standard with the segmented MPAGE and CAIPIRINHA with elliptical-scanning acquisition protocols. Pallidum volumes were most affected by 2D-GRAPPA and caudate volumes were most affected by segmented MPAGE when compared to the reference ADNI-2 protocol.

Overall, the most consistent accelerated acquisition scheme for morphological analysis when compared to the reference ADNI-2 protocol is CAIPIRINHA, since the systematic changes in volumes that were observed are the smallest and a strong correlation between volumes is present. However, this MPAGE variant can only be applied in the situations when multi-channel coils are used and if there is a sufficient variation in coil sensitivity profiles.

For small structures, the observed changes in volumes segmented by MorphoBox were consistent with the results presented in several reproducibility studies [21, 23, 25, 26] comparing repeat scans with an identical protocol. In particular, if we recomputed our normalized volumetric differences between the CAIPIRINHA and the reference ADNI-2 protocol using the methodology described in [21] we observed: hippocampus 2.65 +/- 1.45 %; caudate 1.57 +/-

Chapter 4. Optimization of acquisition strategies and comparison of state-of-the-art T1-weighted whole-brain protocols

1.25 %; pallidum 1.45 +/- 1.35 %; putamen 1.63 +/- 1.03 %; thalamus 1.35 +/- 1.16 %. Those values are in agreement with values reported in a reproducibility study [21] processed with the FreeSurfer [62, 63] software package (version 5.1.0): hippocampus 3.26 +/- 0.93 %; caudate 2.57 +/- 0.36 %; pallidum 7.44 +/- 1.95 %; putamen 4.61 +/- 0.88 %; thalamus 4.97 +/- 1.29 %. Slightly smaller absolute normalized volumetric differences in the current study can be attributed to the use of a single scanner platform, not repositioning the subject within the session, time span between scan and rescan, hydration level changes, different degree of acceleration, different types of accelerations, different age group, smaller subject number, and may also be attributed to the differences in the segmentation software.

ICCs between the reference ADNI-2 protocol and the respective accelerated variants were also similar to previously reported values in reproducibility studies [21, 25, 27], comparing repeat scans with an identical protocol. In particular, the correlations of volumes of small structures from CAIPIRINHA and reference ADNI-2 protocols are comparable to the values reported in [27] processed with FreeSurfer [62, 63] (version 4.0.1): hippocampus 0.95 [0.88-0.98] (this work) vs 0.989 [0.976-0.997] [27], caudate 0.98 [0.95-0.99] vs 0.994 [0.988-0.998], pallidum 0.88 [0.74-0.95] vs 0.706 [0.445-0.897], putamen 0.97 [0.94-0.99] vs 0.971 [0.939-0.991], thalamus 0.98 [0.96-0.99] vs 0.984 [0.965-0.995]; numbers in brackets represent the 95% confidence intervals. The differences in ICCs are small and can be attributed to the bigger subject number and slightly different protocol parameters (TR/TI/TE/flip angle) in the current study as well as different scanner platform, not repositioning the subject within the session, time span between scan and rescan, hydration level changes, different degree of acceleration and different segmentation software.

Considering other possible protocol combinations, the most interchangeable protocol combinations are CAIPIRINHA / 2D-GRAPPA and CAIPIRINHA / CAIPIRINHA with elliptical scanning. This conclusion is based on the observed high ICC values and small changes in volumes relative to the other protocol combinations (Figure 4.5, Figure 4.7). There is a strong indication that segmented MPRAGE acquisitions introduce a change to CSF volumes. These differences can be attributed to the reduction in susceptibility-related distortions due to the increased bandwidth. Qualitatively, it can be observed that there is a better delineation of CSF in segmented MPRAGE acquisitions. We assume that this MPRAGE variant exhibits similar properties as the multi-echo MPRAGE variant MEMPRAGE [64].

The CNR and SNR analysis performed in this study depends on segmentation quality of MorphoBox and the accuracy of the registration algorithm. An alternative option would be to perform brain tissue segmentations manually. However, manual segmentations would be vulnerable to human error and are very time consuming. Another alternative option would be to not fully segment brain tissues but rather select small ROIs within them. However, using this method for SNR and CNR analysis will not necessarily capture the noise amplification (g-factor).

In order to assess the segmentation errors made by MorphoBox, it would be ideal to have the

ground truth segmentations. However, no brain segmentation method, be it automatic or manual, is error-free. Experimentally, MorphoBox and FreeSurfer [62, 63] segmentation tools have been compared on the ADNI [16] database in terms of disease detection accuracy and showed that MorphoBox and FreeSurfer achieve similar performance levels [12]. In the current study, no gross segmentation errors were observed in all of the acquired image volumes and the differences were at the single voxel level. Visually, most of these differences occurred in regions affected by partial volume effects. Potential segmentation errors can be a consequence of partial volume effects since they make the segmentation problem intrinsically ambiguous and CNR/SNR changes can amplify this ambiguity. An indication of this effect is given by the order of magnitude of the observed differences in volumes between the protocol variants, suggesting that only a small number of voxels are affected.

The SNR analysis yielded approximately a $1/\sqrt{2}$ decrease in SNR when moving from the accelerated reference ADNI-2 protocol to further accelerated variants. This result is expected from theoretical considerations when assuming that noise is proportional to $\sqrt{\text{sampling time}}$ [65]. In 2D-GRAPPA, CAIPIRINHA and CAIPIRINHA with elliptical-scanning acquisitions, the reduced sampling time, and thus the decrease in SNR, is attributed to undersampling of k-space. The small difference to the theoretically expected value indicates a negligible additional g-factor penalty. In segmented MPRAGE, the decrease in SNR is due to the increased bandwidth.

The SNR efficiency of 2D-GRAPPA, CAIPIRINHA, and segmented MPRAGE acquisitions decreased compared to the ADNI-2 protocol. This can be attributed to prolonged fraction of scan time as a result of the acquisition of reference lines in these protocol variants. CAIPIRINHA with elliptical scanning was not statistically different from the ADNI-2 protocol due to reduced scan time and increased SNR because of the use of elliptical scanning.

Systematic changes in the CNR were observed in the accelerated acquisitions. The analysis of the means and variances of intensities of the accelerated protocols revealed that there is a systematic change in the noise figure compared to the reference protocol. This change in variance is largely explained by the observed changes in SNR of the accelerated acquisitions.

It is important to note that the changes in SNR and CNR can also potentially influence the scan-rescan reproducibility of the four-fold-accelerated protocols in terms of segmentation results. Even though, a recent study [22] demonstrated negligible impact on the test-retest reproducibility within identical protocols, with our current experimental design, we are unable to fully address scan-rescan reproducibility of the four-fold-accelerated protocols and further research is needed. Therefore caution must be exercised when using the four-fold-accelerated protocols in research or clinical settings for volumetric analysis.

In the numerical experiment that added synthetic noise to the reference ADNI-2 data, there is a high degree of consistency between the segmentation results of synthetic data within a subject. This conclusion is based on the high ICC values for all considered structures. In terms of volumetric results, the synthetically noise-matched ADNI-2 scans and the accelerated

Chapter 4. Optimization of acquisition strategies and comparison of state-of-the-art T1-weighted whole-brain protocols

protocols exhibited similar bias when compared to the reference ADNI-2 protocol. Although the results suggest that noise properties affect volume bias, it should be noted that the exact relationship between noise and volume biases and the impact of other sources of biases cannot be disentangled by the current study design.

Overall, our analysis of the influence of noise on the segmentation results suggests that, in contrast to other reproducibility studies, inconsistencies between scans can be at least partly explained by the noise, CNR, and SNR – values which are measurable in dedicated settings. In other words, we hypothesize that it may be possible to correct for those systematic changes in the volumetric assessments of brain structures based on the knowledge of the noise level in the image. This statement is based on our experimental observations of volumetric measurements with high ICC values between protocol variants and the presence of a systematic bias.

However, further investigation is needed to generalize these findings, i.e. to determine if it is possible to extrapolate the findings to different coil setup, scanner field-strength, acquisition time, resolution, scan orientation and other parameters that may result in different levels of partial volume contamination and different SNR/CNR levels.

4.6 Conclusion

In summary, there are three main practical results of this study. First, accelerations of 3D structural brain scans beyond the routinely used acceleration factor of two have a measurable impact on some image analysis metrics (CNR, SNR, and noise). However, obtained images provide at least very similar information for qualitative readings in this 3T setting. Second, our analysis suggests that using or combining data from different variants of MPRAGE protocols should be done with caution. This statement is based on the findings of a number of quantitative image analysis metrics including SNR, CNR and volumetric assessments that all showed differences between the variants. This holds especially true for small brain structures that are subject to higher partial-volume effects than larger structures. Third, our results suggest that volumetric biases (at least under the given study conditions and subject to limitations mentioned in discussion) are largely affected by the noise properties of the images.

Overall, one has to carefully consider the exact use case of the accelerated protocols. In some situations, the benefits of using such accelerated protocols may potentially outweigh the drawbacks. For example, the management of patient motion is expected to improve through the use of accelerated protocols. In our study, it was demonstrated that accelerated protocols may be used in routine clinical readings. An indicator of this statement is given by the results of the qualitative readings that find the reference ADNI-2 and the CAIPIRINHA scans to have equal image quality based on qualitative assessment. We provide an error estimate on the volumetric results, when accelerated protocols are used and compared to a “standardized” ADNI-2 protocol. Therefore, if quantitative volumetric assessment is of interest, a careful consideration must be given to the effect size of changes in volumes for the structure of interest when using data from accelerated protocols. The future direction of this study will be to investigate scaling procedures as a feasible way to correct for inconsistencies in the accelerated protocols.

5 Correction and optimization strategies towards reliable and robust estimation of data elements

The primary objective of this chapter is to investigate the applicability of scaling procedures as a means to compensate for discrepancies between hardware setups, and protocols observed in Chapters 3 and 4. Data collected as part of the studies described in Chapters 3, 4 and an additional set of data collected as a part of ADNI initiative are considered in the present chapter. Current state-of-the-art protocol-based approaches are investigated and compared to the proposed image-property-based approach.

5.1 Introduction

In the analysis of different acquisition scenarios used for volumetric assessment in Chapter 4, we observed discrepancies between morphometric data acquired with different protocols on the same system with the same resolution (1mm x 1mm x 1mm). We varied the acquisition time at the expense of both contrast-to-noise ratio (CNR) and signal-to-noise ratio (SNR) which resulted in systematic biases in volumes of most notably grey matter (GM) and white matter (WM) structures.

In Chapter 3, we investigated the scenario under which the SNR matched protocols were used on different MR scanner models. Several scanning scenarios were analysed: a) back-to-back scan-rescan; b) repositioning; c) 2 week gap between scan and rescan. Both b) and c) acquisition scenarios were shown to be largely negligible compared to scan-rescan variability. Acquisition protocols of systems with different field strength were adjusted according to ADNI guidelines [17, 16] in order to provide similar SNR performance across the field strengths. The most notable difference between the 1.5T and 3T protocol parameters was the voxel volume (1mm x 1mm x 1mm vs. 1.25mm x 1.25mm x 1.2mm). Despite the similar SNR between 1.5T and 3T protocols, differences in contrast-to-noise ratio were observed. Also, differences in volumes of all of the investigated structures were observed between either protocols or systems.

Without the use of any sophisticated correction strategies as discussed in the further scope of this work, our findings indicate that the best reproducibility is achieved when the same system and protocol are used. This has important implications in the context of both research study designs and clinical workflows. For instance, a recent meta-analysis by Frisoni et al. includes 9 investigations addressing the rate of hippocampal atrophy in a total of 645 Alzheimer's disease (AD) patients and 348 controls [8]. The average rate of atrophy per year was 2.9-5.6%, mean 4.25% in AD and 0.3-2.2%, mean 1.25% in controls, which results in an average difference of 3% atrophy per year between AD versus controls. This means that the methodological variability should be less than 3% if baseline and follow-up investigations are done in an interval of 1 year. This highlights that both changes associated to normal aging and AD or mildly cognitively impaired (MCI) are small and further motivates the need to remove the effects that might confound the analysis.

However, in practice it is not always possible to perform the measurements on the same system both in the single centre settings and especially in the multi-center settings. Moreover, with uncooperative patients, the acquisition time can be a limiting factor making it difficult to keep the protocol parameters consistent. Allowing the use of different systems and the flexibility in protocol parameters can drastically accelerate the collection of data and result in an increase of statistical power by virtue of having a greater number of subjects. This increase in statistical power can allow detection of morphometric changes with greater precision in both cross-sectional and longitudinal studies.

The purpose of this chapter is to investigate correction strategies for the systematic biases

present in brain volumetric measurements due to heterogeneous acquisition systems and protocols. The current state-of-the-art approaches, propose to either remove the unwanted technological variability through preprocessing of the input data[66, 67] or to scale the volumes obtained with the automated segmentation software to a reference value based on the knowledge of the protocol[45, 68].

The use of protocol-based correction strategies has several drawbacks. First, a separate calibration study needs to be carried out whenever a new protocol, system, or site are introduced. Second, caution must be exercised when applying protocol-based compensation strategies to subjects that undergo changes in contrast due to disease, aging or other not system or protocol related factors. For instance, it has been reported that there is a decrease in T1-weighted contrast that is attributed to aging [69]. These changes in contrast can potentially confound the segmentation results due to partial volume effects and bias the estimated effect of aging, disease, or treatment.

To address these problems a novel image-property-based approach of scaling the data to the same reference level is developed and proposed in this thesis. It is compared to the state-of-the-art protocol-based approaches.

5.2 Materials and Methods

5.2.1 Acquisition

We had access to three datasets acquired under an IRB-approved protocol that contained 3D T1-weighted MPRAGE acquisitions.

Bonn dataset

This dataset consists of 22 healthy young subjects (20-44 years old). The experimental design and acquisition parameters are described in Chapter 4. The voxel size was maintained constant but the scan time was varied at the expense of SNR and CNR.

Basel dataset

This dataset consisted of 22 healthy young subjects (20-39 years old). The experimental design and acquisition parameters are described in Chapter 3. Note that unlike the Bonn dataset, the voxel size was adjusted to maintain the same SNR levels across the systems. In Chapter 3, we observed changes in CNR between data acquired at different field strengths. However, those changes are correlated to the changes in voxel size making it difficult to extrapolate the changes observed in the Bonn dataset to the current one.

Chapter 5. Correction and optimization strategies towards reliable and robust estimation of data elements

ADNI dataset

This multi-centre dataset consists of screening (1.5T)/baseline (3T) T1-weighted MPRAGE acquisitions. Clinical status (healthy, mildly cognitively impaired(MCI) or Alzheimer(AD)), age, gender parameters were available as a part of this database for each subject. To mitigate hardware effects, our analysis was restricted to the 335 subjects (54-91 years old) who were scanned on systems of one vendor (Siemens Healthcare, Germany) only. Within these subjects, 95 were healthy, 166 were MCI, and 166 were AD. Each subject had two screening scans acquired at 1.5T without repositioning with a voxel size of $1.25 \times 1.25 \times 1.2$ mm³. A fraction of 45 subjects also had two consecutive baseline scans at 3T with a voxel size $1 \times 1 \times 1.2$ mm³ (roughly SNR-matched to the 1.5T data). Note that similarly to the Basel dataset, the field strength is strongly correlated to the voxel size.

Interpolated Basel dataset

The slice thickness used in 3T ADNI protocol [17] differs from the 3T protocol used in the acquisition of the Basel dataset(1mm vs 1.2mm).

To investigate if we can apply the correction strategies learned from the Basel data set to the ADNI dataset, the raw k-space data(which was stored during the study) from all 3T Basel datasets was cropped to match the resolution of the 3T ADNI protocols(1.2mm slice thickness) and retrospectively reconstructed on the scanner.

Combined dataset

This dataset consisted of both the interpolated Basel and the ADNI datasets.

5.2.2 Data Processing

Automated segmentation

All scans were processed using the MorphoBox prototype [12] to estimate the volumes of total gray matter (GM), total white matter (WM), total cerebrospinal fluid (CSF), total normalized brain volume(NBV), hippocampus, thalamus, putamen, caudate, and pallidum structures.

CNR assessment

The CNR of all image volumes was quantitatively assessed using the label maps provided by MorphoBox. The methodology used to assess the CNR is described in Chapter 3.

5.2.3 Statistical analysis

In Chapter 3 a linear mixed effect model was used to detect the differences between the acquisition systems. However once the model is fit, it is possible to use it to make predictions about previously unseen data. In this section two models that were used to make predictions to scale volumetric data to the same reference are outlined.

Protocol-based compensation strategy

First, we examined a simple model which considered that the main parameter that defines the differences between the outcomes of two consecutive automatic brain volume segmentations with-in a subject is the protocol type.

For each structure of interest, a linear mixed-effect model[44]:

$$y = X\beta + Z\gamma + \epsilon \quad (5.1)$$

was fit on the estimated volumes normalized by total intracranial volume and logarithmically transformed. The purpose of the logarithmic transform is to warrant that volumes predicted by the linear mixed-effect model are always positive. In this model, protocol type, age, sex, and clinical status(healthy, MCI or AD) were modelled as fixed effects (X). Random effects consisted only of subject-dependent offsets (Z). β and γ were regression coefficients that represent the fixed and random effects. Mathematically, the main difference between regression coefficients is that β was modelled as a fixed parameter and γ as a random variable. The residual error ϵ represented the scan-rescan variability.

Image-property-based compensation strategy

There are various factors that can potentially affect the outcomes of the morphometric assessment (SNR, CNR, voxel-size, blurring, ghosting, ringing, etc). In this part of the experiment we focused on the CNR and voxel-size parameters. Blurring, ghosting and ringing were assumed being negligible as all data were found of high image quality based on visual inspection and automated quality control [70]. With the protocol-based correction strategies, we implicitly assumed that we can capture the changes with just one parameter, namely the system/protocol identifier. However, there are changes in contrast that are attributed to normal aging [69]. In Chapter 4, we observed that both CNR and SNR can potentially affect the outcome of an automatic brain segmentation. Therefore, we propose a slightly more complex model of the similar form as Equation 5.1 that aims to explain the discrepancies in volumetric observations with-in a subject purely based on the voxel volume, and the contrast-to-noise ratio.

For each structure of interest, a linear mixed-effect model was fit to the sequence of volumes normalized by TIV after logarithmic transformation with clinical status (healthy, MCI or AD), subjects age, voxel volume, and CNR normalized by voxel volume treated as fixed effects (X),

Chapter 5. Correction and optimization strategies towards reliable and robust estimation of data elements

and the subject identifier treated as a random effect (Z) in order to account for the repetition of measurements within subjects. The normalization of CNR by voxel volume was performed in order to mitigate the correlation of CNR and voxel volume.

Evaluation of compensation strategies

All models were fit with the R software (version 3.1.1) using the lme4 package [44] on each dataset separately. First, we looked at the regression coefficients produced by each model. The Satterthwaite's approximation was employed to estimate the number of degrees of freedom using the lmerTest package [46], and generalized t-tests were used to test the significance of the regression coefficients. The regression coefficients were plotted for each dataset separately.

All datasets were first examined separately using a leave-one-out cross validation. The fitted model parameters were used to correct volumes for fixed effects so as to match the arbitrary reference of a healthy subject scanned at 3T with ADNI protocol parameters.

However, an important question that needs to be answered is how generalizable are our findings. To answer this question, we would need both a training and a test dataset. We can consider ADNI and Basel databases to be as training and testing datasets and vice-versa. Since there are differences in resolution between the ADNI and Basel imaging protocols (1.2mm vs. 1mm slice thickness) we used the interpolated data from the Basel dataset for this part of the experiment. It is important to point out that the subjects in the ADNI database were old and there were diseased subjects in contrast to Basel dataset where there were only young and healthy subjects. Note that there is a strong correlation between voxel size and CNR. However, if our testing dataset is sampled in the same way as the training dataset, this co-linearity should not affect the predictive power of the model.

To study if the observed differences were due to insufficient sample size, data were combined into a combined dataset that consisted of the interpolated Basel and the ADNI datasets, and a leave-one-out cross validation on combined dataset was performed.

In order to assess the performance of each correction strategy within one subject, relative volume differences (RVD) were computed between all possible combinations of acquired data as:

$$RVD = \frac{|V_1 - V_2|}{V_1 + V_2} \quad (5.2)$$

where V_1 and V_2 were brain volume estimates.

Normalized histograms were constructed for both corrected and uncorrected data.

Finally, to obtain further insight into age dependent effects, we performed the following bootstrap procedure. We randomly drew a sample of 20 subjects from the combined dataset to train the model and 20 subjects to test the model. This procedure was repeated varying the

percentage of subjects from ADNI datasets from 0-100% with a step of 10% for the training dataset and 20 subjects would be selected randomly from the remaining subjects at each step. The mean difference between adjusted and unadjusted RVDs was plotted as a function of the training database age and testing database age.

5.3 Results

5.3.1 Protocol-based compensation strategies

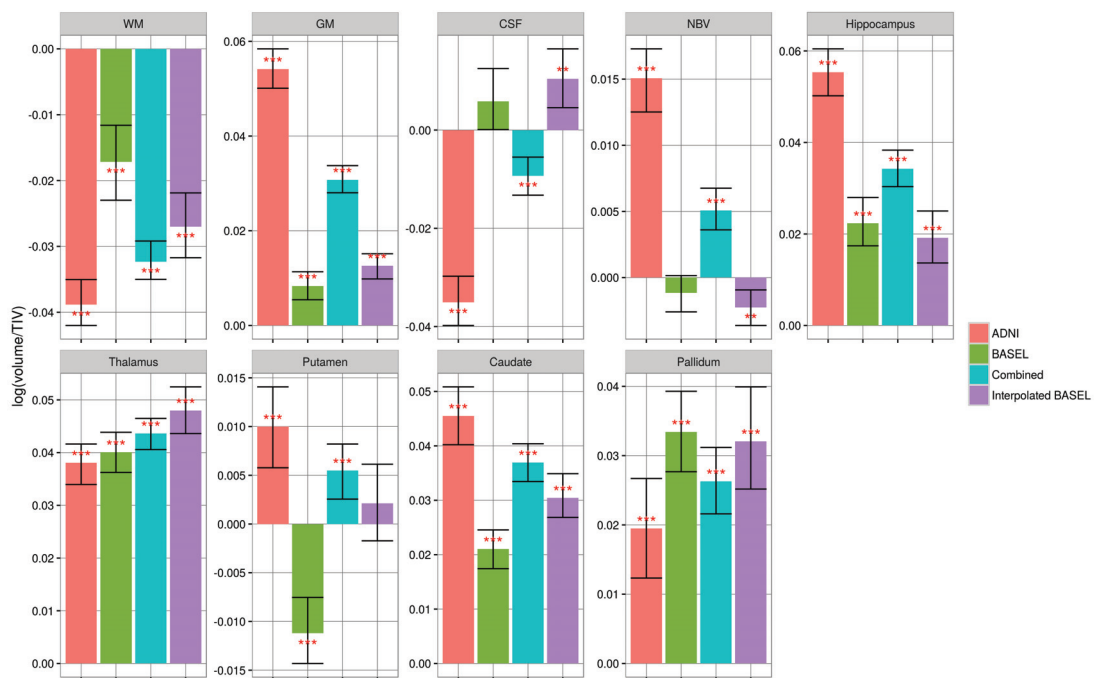


Figure 5.1: Regression coefficient for protocol-based model fit separately for each structure on ADNI, Basel, interpolated Basel, and combined datasets. Error bars indicate 95% confidence interval computed via boot strap with 500 simulations. * $p < 0.05$; ** $p < 10^{-2}$; *** $p < 10^{-3}$;

First, the effects of protocol-based correction strategy were examined. Regression coefficients for ADNI, Basel, combined, and interpolated BASEL datasets are plotted in Figure 5.1. Note that the direction of the field strength effect is consistent for all structures except for CSF and putamen. The magnitude of the effect is notably smaller with the interpolated Basel dataset compared to the ADNI dataset except for thalamus and palladium structures.

The leave-one-out cross validation results for the model trained on the ADNI dataset are illustrated in Figure 5.2, Table A.3. Comparison of distributions of errors for corrected/uncorrected 1.5T and 3T protocols showed a notable decrease in mean RVDs for all of the evaluated structures except for the putamen after applying the protocol-based correction strategy. However, note that for all of the investigated brain structures, RVDs were still larger on the average than the scan-rescan variability (illustrated in green) after the application of the protocol-based

Chapter 5. Correction and optimization strategies towards reliable and robust estimation of data elements

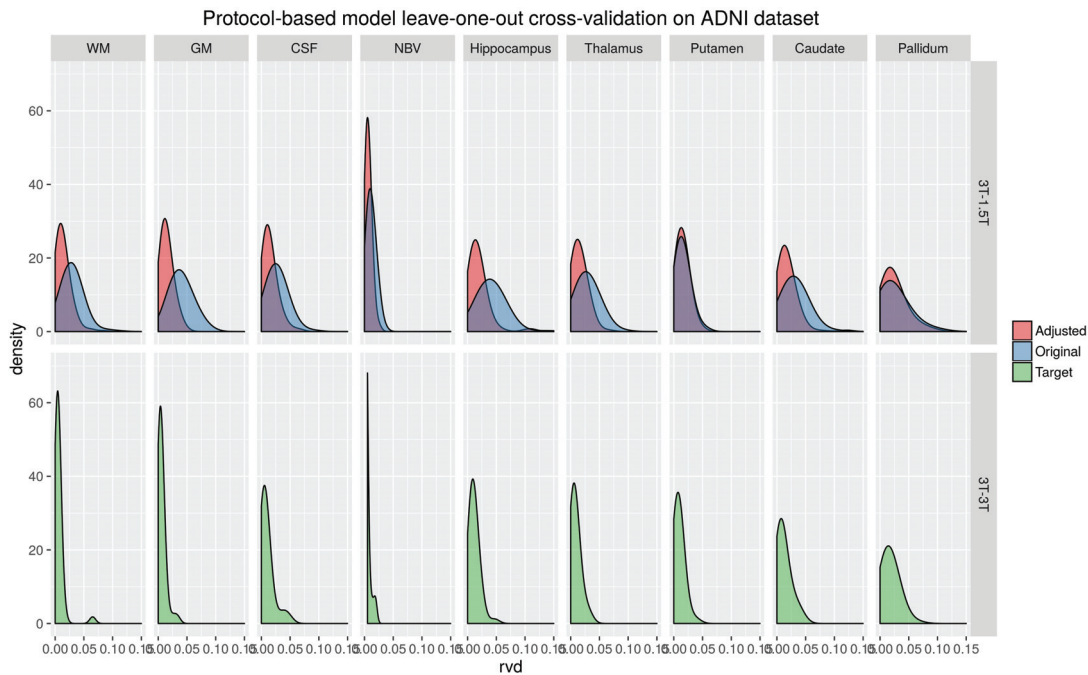


Figure 5.2: Distributions of relative volume differences 1.5T/3T (Adjusted, Original) and 3T/3T (pure scan-rescan variability) of ADNI dataset with protocol-based compensation strategy computed in leave-one-out cross-validation settings

correction strategy.

When a leave-one-out cross validation was performed on the interpolated Basel dataset (Figure 5.3, Table A.4), there was a decrease in mean RVDs of: WM, thalamus, pallidum, caudate, and hippocampus structures when volumetric data were scaled to the same reference. There were only small changes to the distribution of errors in volumes of GM, CSF, NBV, and putamen structures compared to unadjusted data. It is noteworthy that the mean unadjusted RVDs are smaller with the interpolated Basel dataset (Table A.4) compared to ADNI dataset (Table A.3) and are very close to the distribution of scan-rescan RVDs for CSF, NBV, and putamen structures.

Figure 5.4, and Table A.5 illustrate the results of leave-one-out cross validation of the model fit on Bonn dataset. There was a notable decrease in mean RVDs of WM structure when accelerated protocols are scaled to match the ADNI-2 reference protocol (Table A.5). The distribution of errors in volumes for other structures did not change significantly. Note that due to the study design, the scan-rescan errors are not available.

When the model used to correct for protocol differences was trained on the ADNI dataset and the volumes from the interpolated Basel dataset were scaled to the same reference value, there was a noticeable increase in protocol discrepancies for GM (Mean RVD 0.01 (original) vs 0.03 (adjusted)), CSF (Mean RVD 0.02 vs 0.04), and NBV (Mean RVD 0.005 vs 0.01) volumes.



Figure 5.3: Distributions of relative volume differences 1.5T/3T (Adjusted, Original) and 3T/3T (pure scan-rescan variability) of interpolated Basel dataset with protocol-based compensation strategy computed in leave-one-out cross-validation settings

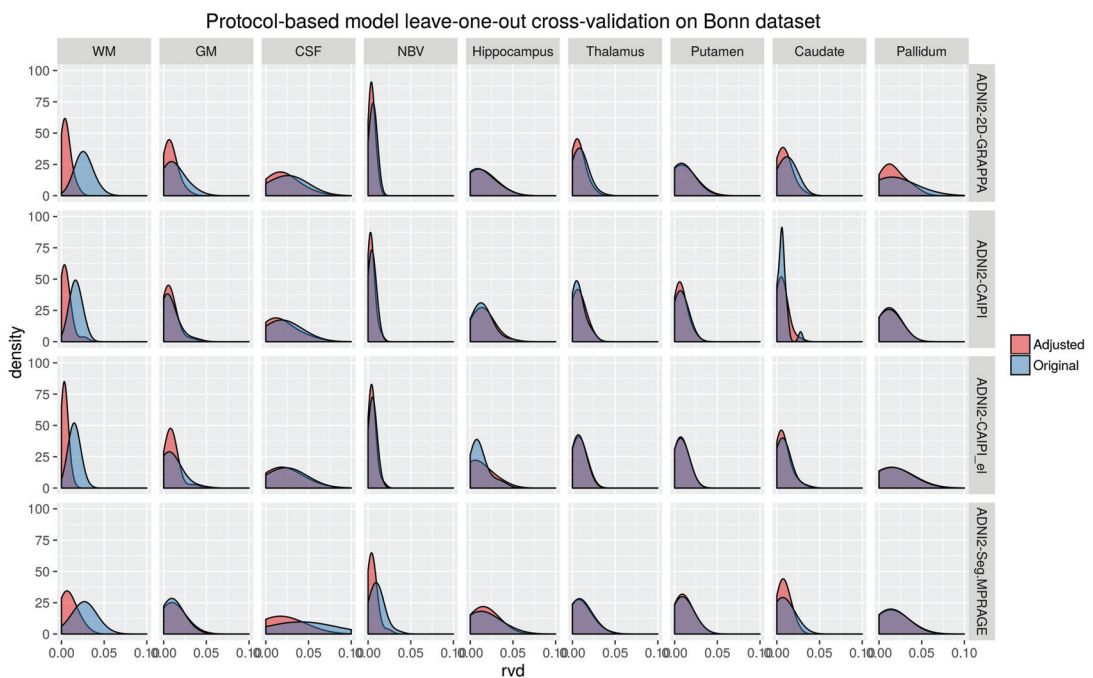


Figure 5.4: Distributions of relative volume differences ADNI-2/accelerated protocols (Adjusted, Original) of Bonn dataset with protocol-based compensation strategy computed in leave-one-out cross-validation settings

Chapter 5. Correction and optimization strategies towards reliable and robust estimation of data elements

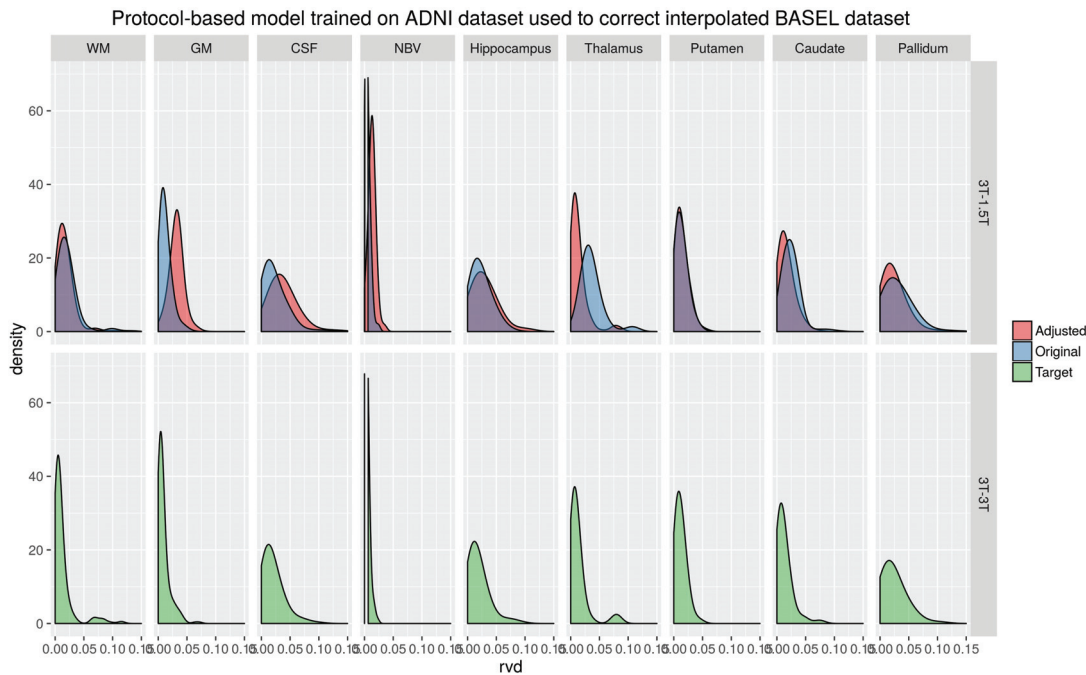


Figure 5.5: Distributions of relative volume differences 1.5T/3T (Adjusted, Original) and 3T/3T (pure scan-rescan variability) of interpolated Basel dataset with protocol-based compensation strategy computed when the model is trained on ADNI dataset and applied to scale interpolated Basel dataset

However, mean RVDs between 1.5T and 3T protocols of WM, thalamus, caudate, pallidum improved compared to unadjusted data (Figure 5.5, Table A.6). The improvement in mean RVDs for thalamus is the most striking. No changes were observed in mean RVDs of putamen.

Training correction model on the interpolated Basel dataset and scaling the volumes from the ADNI dataset to the same reference value resulted in a noticeable improvement in mean RVDs (Figure 5.6, Table A.7) of all structures except the CSF and putamen. However, these improvements were rather modest compared to the improvements observed during leave-one-out cross validation and the scan-rescan variability of 3T data.

When both ADNI and interpolated Basel datasets were combined into a combined dataset and leave-one-out cross-validation procedure was performed (Figure 5.7), there was an improvement in discrepancies in volumes for all structures. On a subset of interpolated Basel dataset, the discrepancies in both CSF and GM were smaller than when the same model was trained on ADNI database and interpolated Basel data was scaled to reference value. However overall, it should be noted that the improvements were rather modest in comparison with the scan-rescan variability of 3T-3T data for both datasets with an exception of thalamus structure.

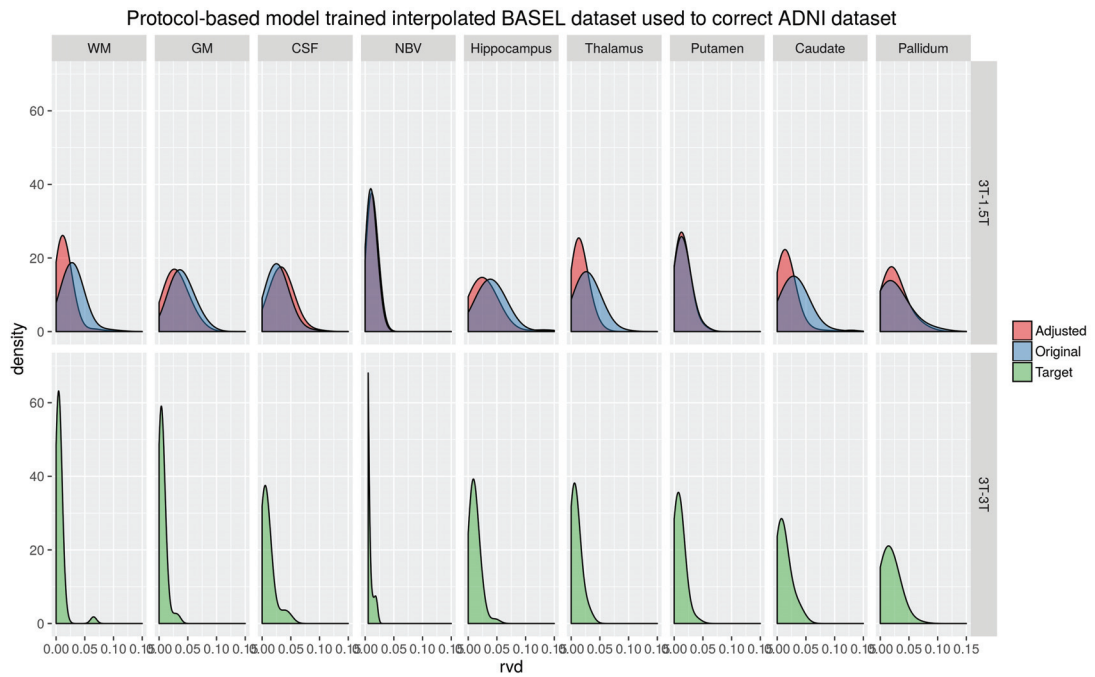


Figure 5.6: Distributions of relative volume differences 1.5T/3T (Adjusted, Original) and 3T/3T (pure scan-rescan variability) of ADNI dataset with protocol-based compensation strategy computed when the model is trained on interpolated Basel dataset and applied to scale ADNI dataset

Chapter 5. Correction and optimization strategies towards reliable and robust estimation of data elements

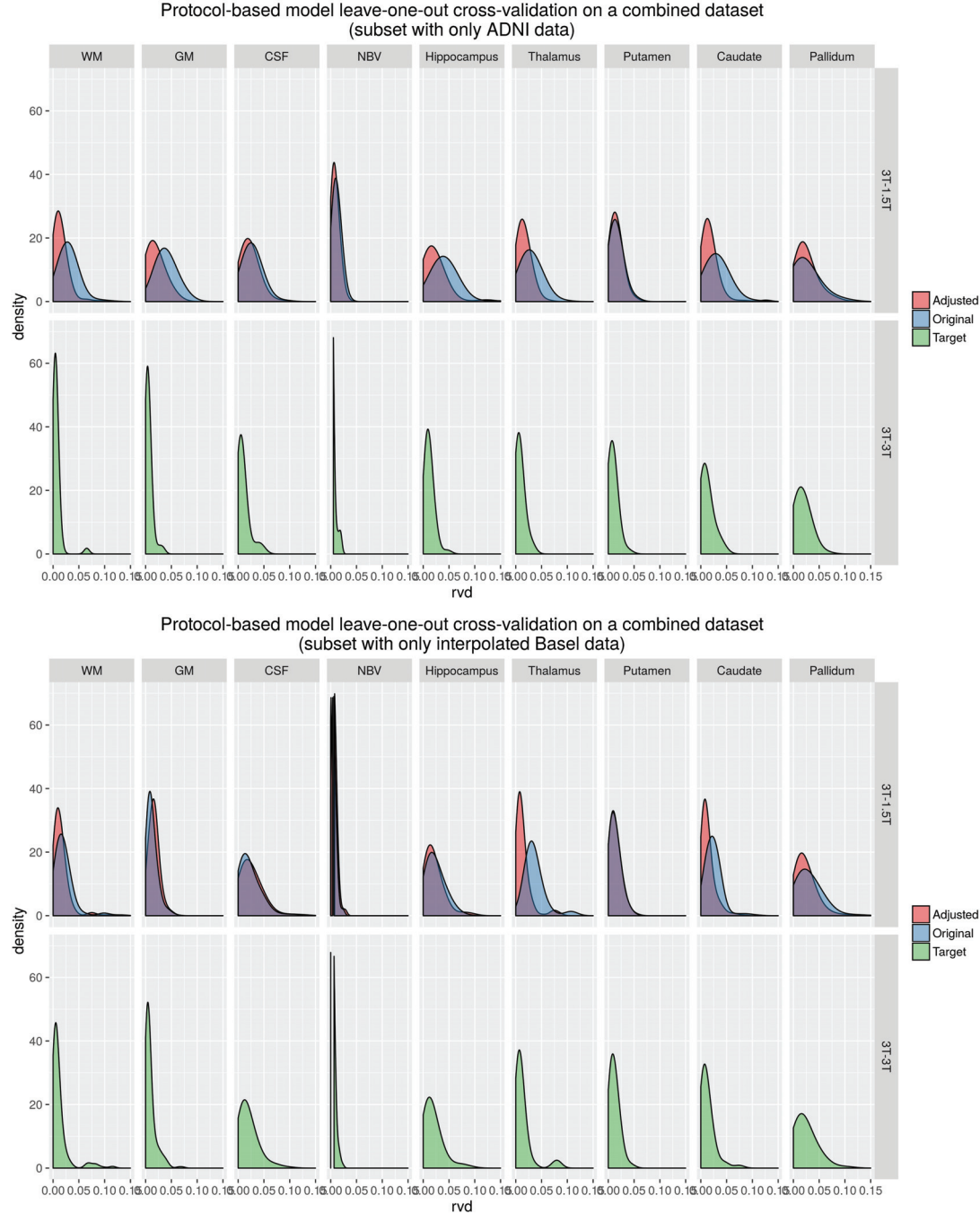


Figure 5.7: Distributions of relative volume differences 1.5T/3T (Adjusted, Original) and 3T/3T (pure scan-rescan variability) of combined dataset with protocol-based compensation strategy computed in leave-one-out cross-validation settings. ADNI and Interpolated Basel datasets are shown separately.

5.3.2 Image-property-based compensation strategy

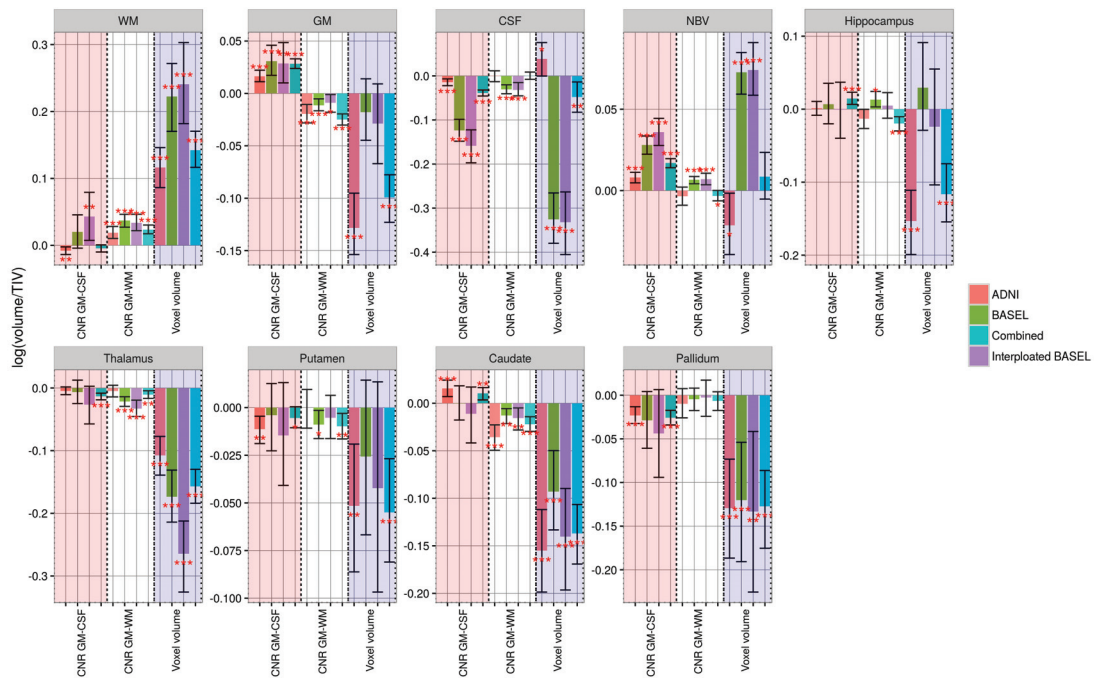


Figure 5.8: Regression coefficient for CNR, and voxel volume effects fit separately for each structure on ADNI, Basel, interpolated Basel, and combined datasets. Error bars indicate 95% confidence interval computed via boot strap with 500 simulations. * $p<0.05$; ** $p<10^{-2}$; *** $p<10^{-3}$;

Figure 5.8 shows regression coefficients of the model that considered CNR and voxel volume as factors defining the outcome of automated brain volume segmentation. The model was trained on the ADNI, BASEL, combined, and interpolated Basel datasets. Note that due to the presence of correlation between CNR and voxel volume, the confidence intervals were rather large.

The results of leave-one-out cross-validation were similar to the protocol-based correction strategy for both ADNI (Figure 5.9, Table A.10) and interpolated Basel (Figure 5.10, Table A.11) datasets. It is important to note that the 3T scan-rescan distributions of errors did not change significantly after the application of the image-property-based compensation strategy. The outcome of leave-one-out cross validation performed on a Bonn dataset was similar to the one of protocol-based compensation strategy resulting in a notable decrease in discrepancies in volumes of WM structure (Figure 5.11, Table 1 A.12).

When the image-property-based model was trained on ADNI dataset and volumes from interpolated Basel dataset were adjusted to match a 3T reference (Figure 5.12, Table A.13), there was a notable decrease in mean RVDs for WM, thalamus, caudate and pallidum (Table A.13). No changes in mean RVD of hippocampus were observed. However, discrepancies in volumes of GM, and CSF structures were increased compared to original values. Note that this

Chapter 5. Correction and optimization strategies towards reliable and robust estimation of data elements

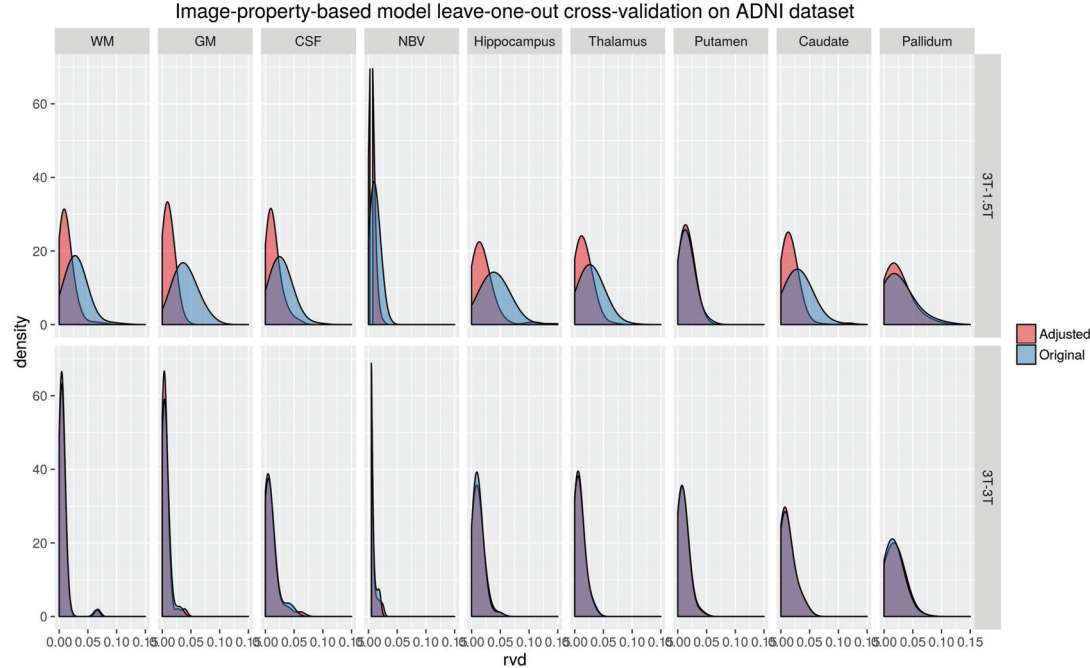


Figure 5.9: Distributions of relative volume differences(Adjusted, Original) 1.5T/3T and 3T/3T of ADNI dataset with image-property-based compensation strategy computed in leave-one-out cross-validation settings

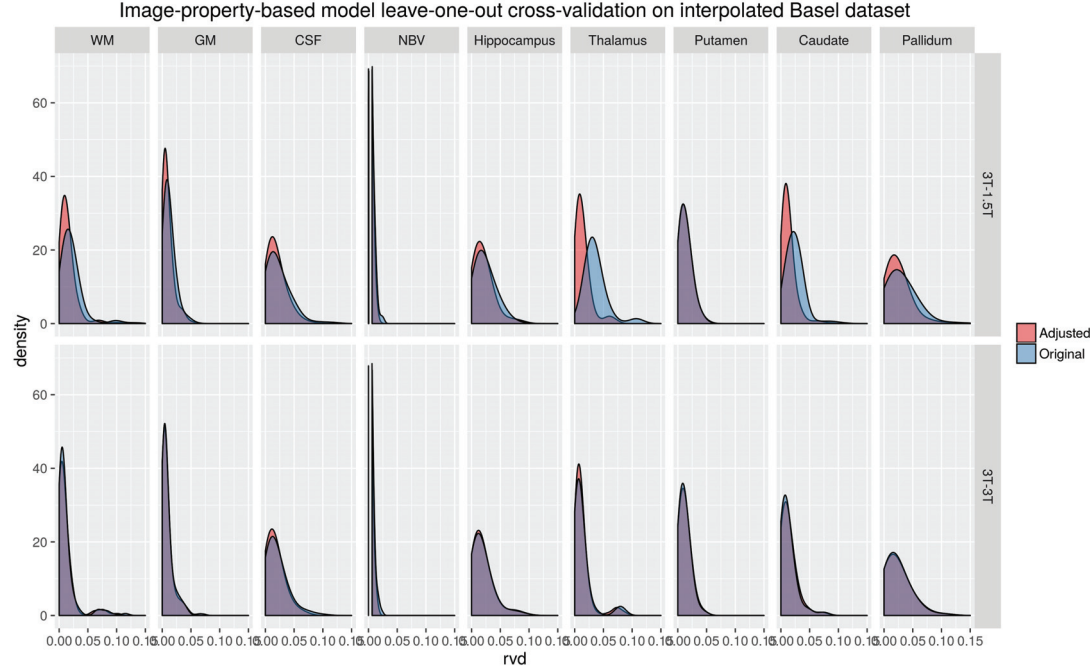


Figure 5.10: Distributions of relative volume differences(Adjusted, Original) 1.5T/3T and 3T/3T of interpolated Basel dataset with image-property-based compensation strategy computed in leave-one-out cross-validation settings



Figure 5.11: Distributions of relative volume differences ADNI-2/accelerated protocols (Adjusted, Original) of Bonn dataset with image-property-based compensation strategy computed in leave-one-out cross-validation settings

increase is smaller than the one of a purely protocol-based compensation strategy.

Figure 5.13, and Table A.14 illustrate the case when the image-property-based model was trained on interpolated Basel dataset and volumes from ADNI dataset were adjusted to the same reference. There was an improvement for all structures except putamen and pallidum. The reduction in discrepancies of volume estimates between 1.5T - 3T protocols when image-property-based correction strategy was applied was greater compared to purely protocol-based approach most notably for volumes of WM, and GM structures.

When both ADNI and interpolated Basel datasets were combined into a combined dataset and a leave-one-out cross-validation procedure was performed, there was an improvement in correcting both the interpolated Basel and the ADNI datasets (Figure 5.14, Table A.15, Table A.16). The discrepancies in both CSF and GM were smaller than when the image-property-based model was trained on ADNI database.

Overall, the image-property-based correction strategy performed considerably better than the protocol-based strategy most notable reducing the discrepancies between 1.5T and 3T protocols in the case of WM, GM, and CSF structures as illustrated by Figure 5.7 and Figure 5.14.

Chapter 5. Correction and optimization strategies towards reliable and robust estimation of data elements

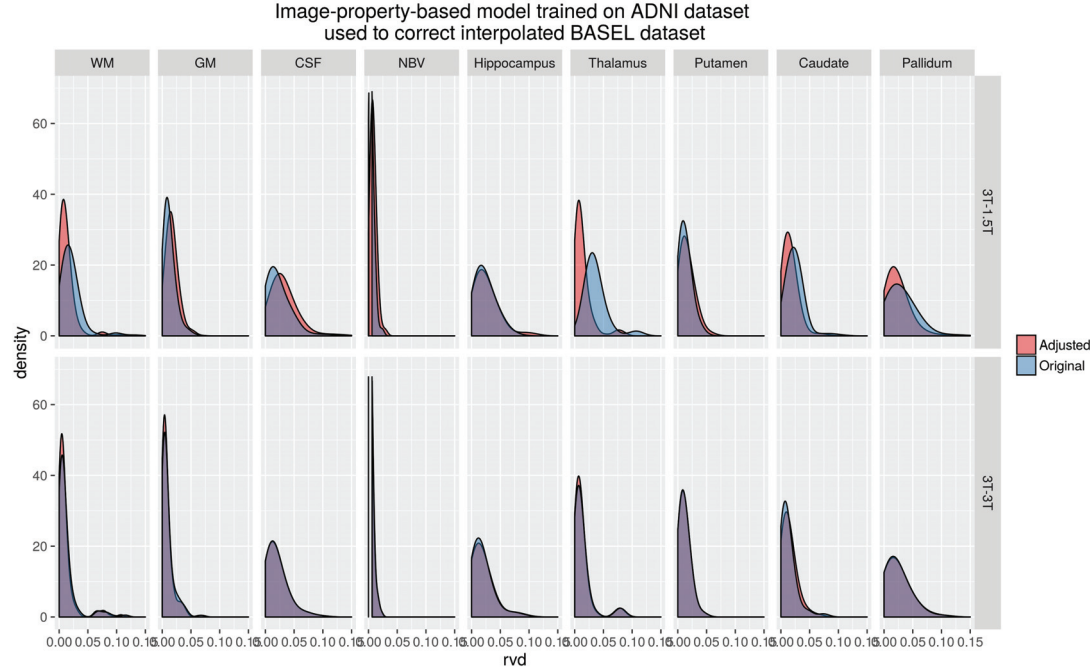


Figure 5.12: Distributions of relative volume differences (Adjusted, Original) between 1.5T/3T and 3T/3T of interpolated Basel dataset with image-property-based compensation strategy trained on ADNI dataset

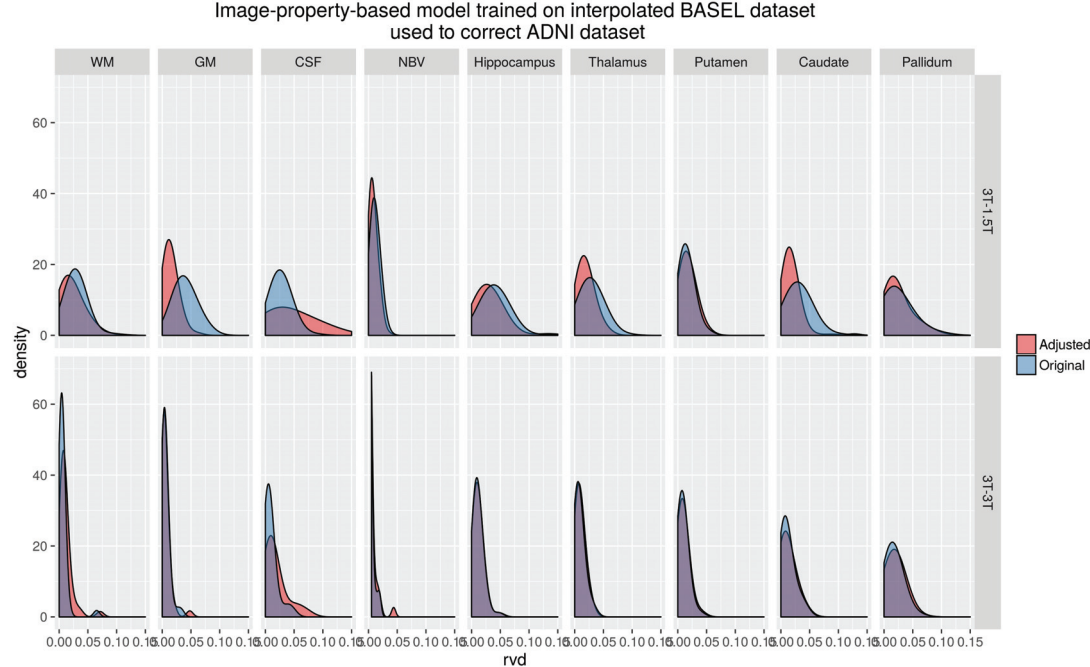


Figure 5.13: Distributions of relative volume differences (Adjusted, Original) between 1.5T/3T and 3T/3T of ADNI dataset with image-property-based compensation strategy trained on interpolated Basel dataset

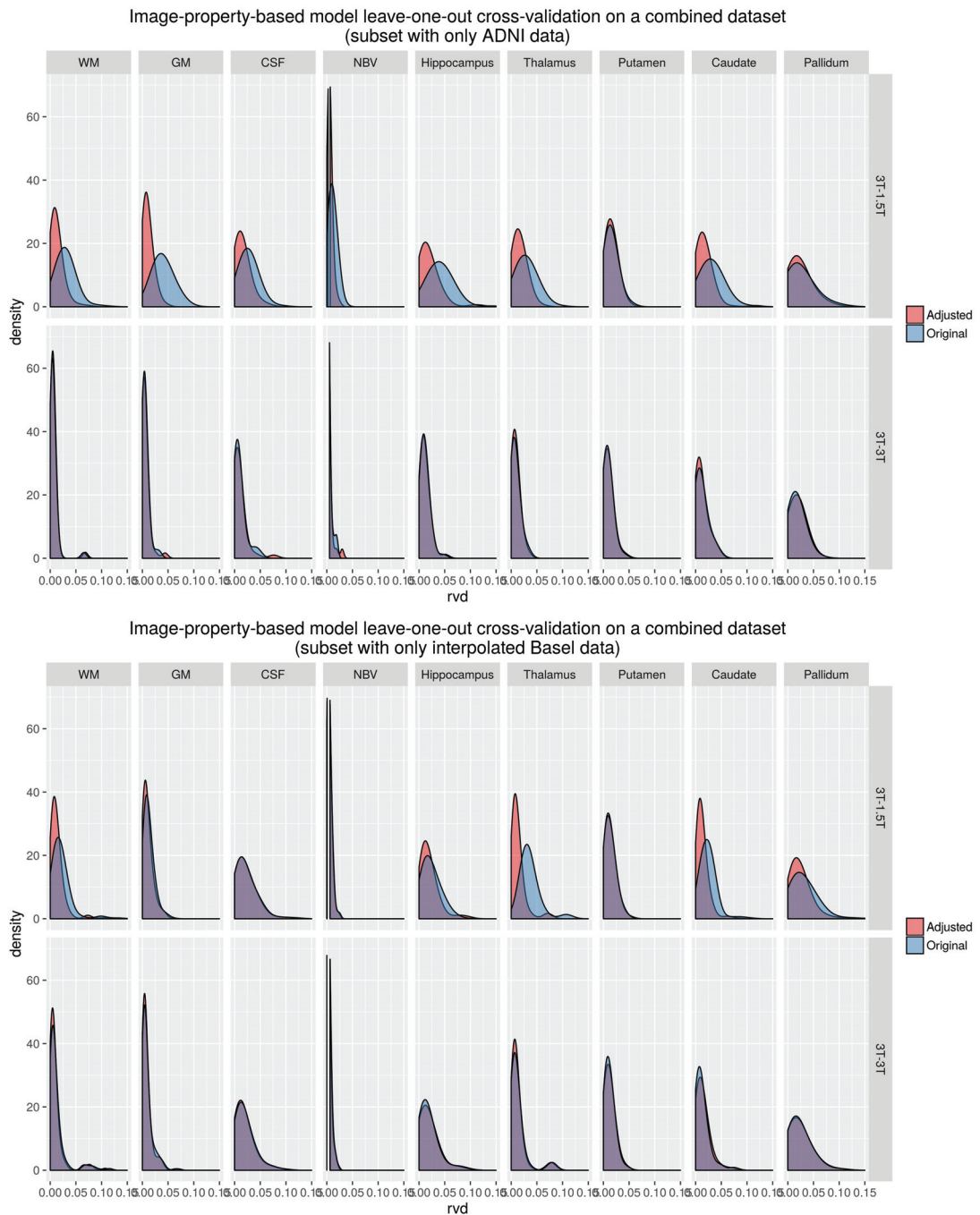


Figure 5.14: Distributions of relative volume differences (Adjusted, Original) 1.5T/3T and 3T/3T of combined dataset with image-property-based compensation strategy computed in leave-one-out cross-validation settings. ADNI and Interpolated Basel datasets are shown separately.

Chapter 5. Correction and optimization strategies towards reliable and robust estimation of data elements

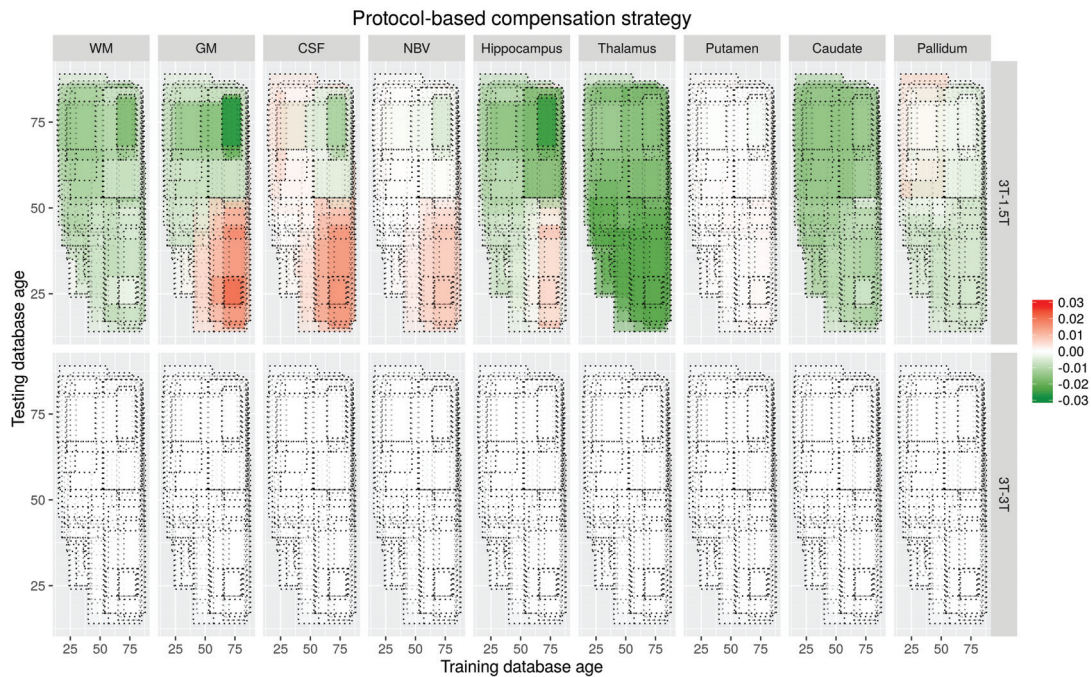


Figure 5.15: Bootstrap validation of protocol-based compensation strategy. Testing database age is plotted on the y-axis and training database age is plotted on the x-axis. Colour represents the difference in mean RVDs of adjusted and unadjusted data.

5.3.3 Bootstrap

The results of bootstrap procedure that investigated the applicability of both image-property-based and protocol-based models are illustrated in Figure 5.15 and Figure 5.16. The y-axis represents the age of the testing datasets and x-axis represents the age of training datasets. Each rectangle represents one bootstrap run. The size of the rectangle in each dimension is 2 standard deviations of the subjects age and the centre is the mean age of subjects for training and testing datasets. The colour represents the difference in the mean RVDs between adjusted and unadjusted data. Red colour indicates the increase in 1.5T-3T discrepancies and green indicates decrease.

The protocol-based compensation strategy amplified the discrepancies between 1.5T and 3T protocols when the age range of training and testing datasets differed (Figure 5.15). The results confirm our initial observations that when the model is trained on ADNI dataset and interpolated Basel data was scaled to a reference value (Figure 5.5) there was a notable increase in differences between 1.5T-3T protocols for GM, CSF structures. The modest improvement in differences between 1.5 T - 3 T data for GM and CSF structures was observed when the model was trained on young subjects and old subjects were used to test the model, as was previously observed when the model was trained on interpolated Basel dataset and the data from ADNI dataset was corrected (Figure 5.6).

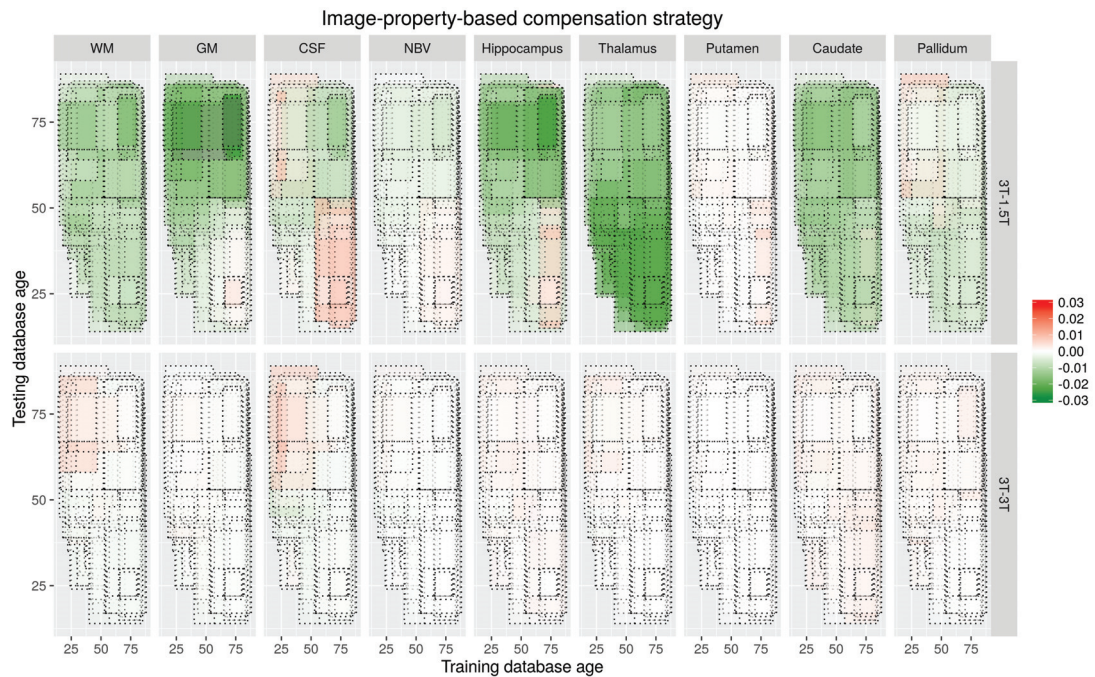


Figure 5.16: Bootstrap validation of image-property-based compensation strategy. Testing database age is plotted on the y-axis and training database age is plotted on the x-axis. Colour represents the difference in mean RVDs of adjusted and unadjusted data.

The image-property-based compensation strategy was more robust with respect to the choice of the age range for both training and testing datasets regarding discrepancies between 1.5T and 3T protocols (Figure 5.16) when compared to the protocol-based compensation strategy (Figure 5.15). The modest increase in 1.5T-3T protocol discrepancies when old subjects were used to train the model and young subjects were used to test confirmed our initial investigation of training the model on ADNI dataset and scaling interpolated Basel data to a reference value (Figure 5.12). There was a noticeable improvement in differences between the protocols for WM and CSF structures when the model was trained on young subjects and the volumes from old subjects were corrected similar to previously investigated scenario when the model is trained on interpolated Basel data and ADNI data is scaled to reference value (Figure 5.13).

5.4 Discussion

In this work, we investigated several approaches to scale multi-protocol and multi-scanner volumetric data to a reference value. A protocol-based compensation strategy was compared with an image-property-based compensation strategy on four datasets: Bonn dataset (the same system with different protocols that modulate CNR keeping voxel volume constant), Basel dataset (different systems, different protocols between field strength), ADNI dataset (different systems, different protocols between field strength), and combined dataset (ADNI

Chapter 5. Correction and optimization strategies towards reliable and robust estimation of data elements

and interpolated Basel dataset).

The protocol-based leave-one-out validations performed well on each individual dataset as illustrated in Figure 5.2, Figure 5.3, and Figure 5.15. Results obtained using the MorphoBox prototype confirmed the feasibility of scaling morphometric data to a reference value purely based on the protocol information as previously reported using both FSL and FreeSurfer segmentation tools[45, 68].

With a protocol-based approach, if we focus on correcting differences between the 1.5T and 3T protocols then when scaling factors were learned using the ADNI dataset and applied to interpolated Basel dataset, the systematic bias in volumes of GM and CSF structures was increased (Figure 5.5).

To obtain further insight into these discrepancies, regression coefficients (Figure 5.1) of the same model fit separately on both ADNI and Basel datasets were examined. The regression coefficients were drastically different for GM and CSF structures. There are several possible sources for these discrepancies:

- (a) Age of the subjects is different (old vs. young)
- (b) 3T interpolated Basel data was acquired using new high-channel head coils while the ADNI data was acquired using older hardware

It is important to note that changes in image contrast associated with ageing have been previously reported [69]. The CNR measured on both ADNI and interpolated Basel datasets confirms previously reported changes in contrast. The decreased contrast makes the tissue segmentation problem more difficult and therefore leads to the higher discrepancies in volume measurements. Therefore, there is an indication that these age-dependent CNR changes can result in age-dependent biases in brain volume estimates.

When leave-one-out cross validation was performed on a combined dataset, the systematic biases were decreased. However, differences were still considerable when compared to the scan-rescan variability. In case of volumes of GM and CSF structures, regression coefficients in Figure 5.1 highlight that when the protocol-based model was fit on a combined dataset it resulted in a compromise between models fit separately on both ADNI and Basel datasets. While this leads to an overall decrease in the systematic biases, it does not yield optimal performance when ADNI and Basel datasets are examined separately (Figure 5.7).

In practice, all these findings indicate that it is difficult to generalize the protocol-based compensation strategies to different age groups and possibly systems. This means that for new age groups, protocols, and systems a separate calibration study would need to be carried out. This may be very costly and often not possible in a real-world scenarios.

Image-property-based compensation strategies have a potential to relax this condition since they would require only one calibration study to be performed. This strategy assumes that the

main factors defining the outcome of morphometric assessments are CNR and voxel volume. For instance, if the calibration study involved a sufficient range of CNR and voxel volume parameter settings than when a new protocol is introduced the effects on morphometry can be extrapolated without the need of performing a separate calibration study. Prior to interpretation of the results, it should be noted that in the current analysis there was a strong correlation between the CNR values and voxel size in the three datasets (ADNI, Basel, Combined) due to the study designs, i.e. the voxel size adjustment between protocols at different field strengths in order to match the SNR.

Due to the presence of this colinearity, we can only perform the following tests:

- (a) Leave-one-out cross validation on each individual dataset (ADNI, Basel, Bonn)
- (b) Leave-one-out cross validation the combined dataset (ADNI + interpolated Basel)
- (c) Train on ADNI dataset and scale volumes of interpolated Basel dataset to a reference value
- (d) Train on Basel dataset and scale volumes of ADNI dataset to a reference value

It is however not possible with the available datasets to extrapolate our findings learned from ADNI or Basel datasets to the Bonn dataset due to the presence of the co-linearity. This is an important limitation of the present investigation. To perform such a task, a dataset that would allow us to disentangle the CNR and voxel size is needed. Such a dataset should consist of a series of measurements performed on several subjects where in each measurement the CNR would be varied across a considerable range (in the context of present study from 2-7 a.u.) and the voxel size would be varied independently.

The image-property-based leave-one-out validations performed well on each individual dataset as illustrated in Figure 5.9, Figure 5.10, and Figure 5.11 and give hope regarding the applicability of the presented approach. There were only very small differences between protocol-based correction strategies and image-property-based correction strategies. When CNR and voxel volume scaling factors were learned from the ADNI dataset and applied to the Basel dataset similarly to the protocol-based correction strategy, we increased the systematic bias of GM and CSF volumes. However, this increase is smaller than with a purely protocol-based compensation strategy. When the image-property-based model was trained on the interpolated Basel dataset and the ADNI data was scaled to a reference value, the systematic biases were reduced. Note that for volumes of GM and CSF structures the bias is still considerable and is higher than both scan-rescan variability and the results of leave-one-out cross validation. This gives an indication that the image-property-based model may be overly simplistic.

When leave-one-out cross validation of the image-property-based model is performed on a combined dataset, systematic biases decrease dramatically. For volumes of GM and CSF

Chapter 5. Correction and optimization strategies towards reliable and robust estimation of data elements

structures this decrease was greater than the one of purely protocol-based correction strategy. These findings indicate that one of the reasons for suboptimal performance in this scenario when the model was trained on ADNI dataset and applied to Basel dataset may be insufficient size of the training database.

The applicability of both image-property-based and protocol-based models was studied using a bootstrap procedure. The results from a leave-one-out cross validation on a combined dataset can be susceptible to overfitting. Therefore, bootstrap procedure gives some further insights into the applicability of our approach. In this context, it is illustrative to compare two compensation strategies: a purely protocol-based and an image-property-based. Figure 5.15 and Figure 5.16 show that the image-property-based model was more generalizable. GM and CSF structures illustrate this effect, and the best performance is achieved when the same age range is used for both training and testing datasets. However, these plots also indicate that the image-property-based compensation strategy is more robust when the range of ages is different on training and testing datasets compared to protocol-based model.

Note that overall, while basic image parameters considered in the compensation strategies investigated in the present thesis can help to mitigate some of the observed differences between the protocols, further research into the metrics that can further parametrize the image for automated volumetric assessment is needed. For instance, in the case of the hippocampus volumes that can serve as a biomarker for Alzheimer's disease, while there was a significant reduction in differences after applying compensation strategies to the sequential volume measurements (reported RVDs need to be multiplied by a factor of 2 to be comparable to results reported by [8]), the differences in volumes were still above the 3% that is necessary to detect Alzheimer's disease from the subject's atrophy rate assessment.

5.5 Conclusion

The analysis of sequential automated brain volume measurements performed on data acquired with different protocols and systems discussed in the previous chapters indicated possible discrepancies between brain volume measurements. The primary objective of this study was to investigate the feasibility of compensation strategies.

The present study confirmed previously reported findings on the feasibility of scaling morphometric data to a reference value purely based on the protocols that were used. In both cases of differences between 1.5T-3T ADNI protocols and between standard and accelerated protocols, the resulting systematic bias can be greatly reduced with an application of appropriate scaling coefficients. However in contrast to the previous studies, the current study demonstrated discrepancies in scaling coefficients between young and old subjects or between different hardware setups. Under these circumstances, caution must be exercised in application of purely protocol-based compensation strategies as they may potentially confine the atrophy detection or the ability to detect differences between groups of subjects. This has very important consequences to a practical application of the protocol-based scaling procedures for both clinical routine and research settings as it means that a separate calibration study would need to be carried out for each new age group, protocol and system.

The results of this study give an indication that the image-property-based compensation strategies can have the potential to be more robust compared to the protocol-based ones when there are technology based variations in image content (SNR,CNR) across the subject groups or hardware setups. In practical terms, this means that only one calibration study would need to be performed and the subsequent data can be corrected even if new groups of subjects that significantly differ in contrast compared to the calibration group , acquisition protocols or systems were to be introduced.

Overall, further research is needed to validate our findings. If we restrict ourselves to the differences between 1.5T and 3T ADNI protocols, in the context of the current study an independent test dataset is needed to further support the findings concerning the image-property-based compensation strategy. This dataset must contain both young and old subjects sampled uniformly across the age range. However, if we want to generalize to the accelerated protocols, then both a separate training and testing datasets are needed. In such datasets, for each subject, both the CNR and voxel volume has to be varied independently which will allow to disentangle the CNR from voxel volume scaling coefficients as well as scan-rescan data needs to be available for each measurement.

6 Conclusion and outlook

6.1 Conclusion

MRI is widely used for imaging both anatomy and physiological processes within the human brain. This thesis is focused on structural brain MRI, which can provide sufficient anatomical detail to perform quantitative volumetric measurements. Quantitative volumetric measurements can serve as sensitive biomarkers to track atrophic brain regions, assess disease progression, differentiate pathology from healthy aging, and monitor therapy response.

However, there is currently still a gap between state-of-the-art brain morphometry techniques and everyday clinical practice. The main objective of this thesis was to understand the factors that affect the adoption of automated quantitative volumetric brain measurements in clinical practice and to examine possible compensation strategies to mitigate the effect of such factors.

The main obstacles that affect the adoption of the automated quantitative volumetric measurements in clinical practice can be summarized as follows:

1. Sensitivity to variations in imaging protocol parameters:

Quantitative volumetric brain measurements from a particular subject need to be comparable over time. For instance, if atrophy rates are of interest, a bias in one of the time points can potentially confound the assessment of the atrophy rate.

2. Availability of normative ranges:

One of the clinical use cases can be a comparison of the subject's volume of a particular brain structure to a database of age and sex-matched healthy controls with an aim to get some insight into disease progression. However, this comparison can be challenging if the data were acquired with a protocol that differs from the protocol used in the construction of the normative range. It can be costly and not practical to construct a normative database for every possible T1-weighted protocol. Moreover, if the patient is not cooperative, it can be challenging to acquire the imaging data with the exact parameters used in the construction of the normative range.

Chapter 6. Conclusion and outlook

Quantitative volumetric measurements can serve as a decision support tool for both clinicians and researchers. Moving towards this goal and keeping in mind the hindrances that slow down the adoption of quantitative volumetric measurements in clinical practice, the main contributions of this thesis can be summarized as follows.

First, factors that can affect the automated volumetric measurements were examined.

In the context of Alzheimer's and dementia, there has been a substantial effort to standardize the protocols across vendors and field strength that led to the development of the ADNI protocol. In this thesis, the reproducibility of ADNI protocol was examined across different systems and contributions of technological (acquisition system, repositioning) and physiological (2 week gap between rescan) effects were investigated. Our main result was that the compatibility of serial and cross-platform scans acquired with the ADNI protocol can be best reached by minimizing differences in field, protocol, hardware, software, and physiology. We have observed significant differences (biases) in volumetric measurements between 1.5T and 3T protocols. Within a field strength, an important finding is that most of the variability in the measurements was in back-to-back differences and neither repositioning nor 2 week gap between a rescan did significantly increase the scan-rescan variability.

More generally, in the long-term perspective with the improvements in the MRI hardware it is desirable to enable acquisition of data with the new advanced accelerated protocols. The reduced acquisition time can reduce motion artifacts, improve patient comfort, and increase the overall patient throughput. From a radiology department's management point of view, the most efficient scanner usage is desirable to allow all available scanners to contribute to serial exams rather than restricting the follow-up exam to the previous scanner hardware. In the present thesis, four variants of a four-fold-accelerated MPRAGE protocol (2D-GRAPPA, CAIPIRINHA, CAIPIRINHA elliptical, and segmented MPRAGE) were examined and compared to the ADNI protocol with respect to clinical readings, basic image quality metrics (SNR, CNR), and automated brain tissue segmentation.

The main outcomes of this analysis were the following:

1. Brain scans with the four-fold-accelerated protocols provided very similar information for qualitative readings in the 3T settings.
2. There were significant differences in brain structure's volume estimates between conventional and highly accelerated protocols.
3. Our results suggest that volumetric biases (at least under the given study conditions) were largely affected by the noise properties of the images.

Second, scaling procedures were examined as means for compensating for the observed discrepancies in automated brain volume measurements between different systems or protocols.

The current state-of-the-art protocol-based and new proposed image-property-based compensation strategies were examined and compared to each other. The results of this work can be summarised as follows:

Current state-of-the-art protocol-based approaches can be used to correct for differences in automated brain volume measurements between systems and protocols. However, caution must be exercised when the results are extrapolated to a new group of subjects that significantly differ in CNR compared to the training dataset. Moreover, whenever a new protocol, system or subject subgroup with significantly different CNR properties is introduced, a separate calibration study needs to be carried out to learn the scaling coefficients for this particular protocol, system or subject subgroup.

The new proposed image-property-based correction strategy can be applied to correct for differences in automated brain volume estimates using heterogeneous systems and/or protocols, without knowledge of the said systems or protocols. There can be significant benefits from applying image-property-based correction strategies as opposed to protocol-based approaches since they are potentially more generalizable to unseen systems and protocols. In practical terms, this means that only one calibration study needs to be performed and the subsequent data can be corrected even if a new acquisition protocol, system, or group of subjects with substantially different CNR were to be introduced later into the study.

In summary, this thesis identified the factors that can affect the reproducibility of the sequential automated brain tissue segmentations and investigated the applicability of compensation strategies to facilitate adoption of automated quantitative volumetric measurements in clinical practice.

6.2 Future Research

Further research is needed to validate the image-property-based approach of scaling volumetric data to a fixed reference. The validation of the image-property-based approach can be divided into two parts:

1. Differences between 1.5T and 3T ADNI protocols: In the context of the present thesis, an independent testing dataset is needed to further support our findings. This dataset must contain both young and old subjects sampled uniformly across the age range.
2. Accelerated protocols: In addition to the data collected during the present thesis, separate training and testing datasets are needed. In such datasets, for each subject, both the CNR and voxel volume would need to be varied independently. This would allow to better disentangle the CNR from voxel volume scaling coefficients than using protocols in which CNR and voxel size were roughly matched.

Overall, the image-property-based model considered that CNR and voxel volume can explain the differences between consecutive acquisitions. However, this model may be over simplistic and additional image quality metrics could be considered.

In the present thesis, we aimed to gain some understanding of the effects that acquisition parameters of T1-weighted structural brain imaging have on automated brain volume segmentation and volumetric biomarker extraction. Further research is needed to refine the correction strategies proposed in this thesis as well as to determine the effect of protocol parameters on other automated post processing algorithms. An example of this is automated lesion segmentation in the context of multiple sclerosis, where lesion volumes and total lesion load are biomarkers of interest, which are likely to be system and protocol-dependent. Protocol-based or image-property based approaches could be considered if discrepancies in the lesion volume measurements are observed between different protocols or systems.

A Appendix

A.1 Reproducibility of volumetric data acquired with a standardized T1-weighted brain-imaging protocol

	%TIV		%TIV
Between Subjects Std.	1.2585	Between Subjects Std.	0.9101
Scan - Rescan Std.	0.8960	Scan - Rescan Std.	0.8432
Std. due to R2	0.2106	Std. due to R2	0.3386
Std. due to R3	0.2068	Std. due to R3	0.4414
ESPREE - AVANTO	-0.7854 ***	ESPREE - AVANTO	0.7087 ***
SKYRA - AVANTO	-1.0864 ***	SKYRA - AVANTO	1.1430 ***
PRISMA - AVANTO	-1.0268 ***	PRISMA - AVANTO	0.7417 ***
Age(years)	0.0852	Age(years)	-0.1247 *
SEX(MALE)	2.6138 ***	SEX(MALE)	-2.3068 ***
(a) WM		(b) GM	
	%TIV		%TIV
Between Subjects Std.	1.4844	Between Subjects Std.	1.4840
Scan - Rescan Std.	0.6421	Scan - Rescan Std.	0.6425
Std. due to R2	0.1959	Std. due to R2	0.1962
Std. due to R3	0.2824	Std. due to R3	0.2822
ESPREE - AVANTO	0.0767	ESPREE - AVANTO	-0.0770
SKYRA - AVANTO	-0.0561	SKYRA - AVANTO	0.0566
PRISMA - AVANTO	0.2853 **	PRISMA - AVANTO	-0.2857 **
Age(years)	0.0155	Age(years)	-0.0154
SEX(MALE)	-0.0451	SEX(MALE)	0.0454
(c) CSF		(d) NBV	

Appendix A. Appendix

	%TIV		%TIV
Between Subjects Std.	0.0331	Between Subjects Std.	0.0745
Scan - Rescan Std.	0.0137	Scan - Rescan Std.	0.0238
Std. due to R2	0.0054	Std. due to R2	0.0045
Std. due to R3	0.0033	Std. due to R3	0.0052
ESPREE - AVANTO	-0.0011	ESPREE - AVANTO	0.0101 *
SKYRA - AVANTO	0.0136 ***	SKYRA - AVANTO	0.0608 ***
PRISMA - AVANTO	0.0148 ***	PRISMA - AVANTO	0.0630 ***
Age(years)	-0.0005	Age(years)	-0.0058
SEX(MALE)	-0.0093	SEX(MALE)	0.0055
(a) Hippocampus		(b) Thalamus	
	%TIV		%TIV
Between Subjects Std.	0.0407	Between Subjects Std.	0.0455
Scan - Rescan Std.	0.0218	Scan - Rescan Std.	0.0140
Std. due to R2	0.0043	Std. due to R2	0.0001
Std. due to R3	0.0082	Std. due to R3	0.0016
ESPREE - AVANTO	0.0018	ESPREE - AVANTO	0.0018
SKYRA - AVANTO	-0.0206 ***	SKYRA - AVANTO	0.0141 ***
PRISMA - AVANTO	-0.0143 ***	PRISMA - AVANTO	0.0236 ***
Age(years)	0.0028	Age(years)	-0.0002
SEX(MALE)	-0.0272	SEX(MALE)	-0.0409 *
(c) Putamen		(d) Caudate	
	%TIV		%TIV
Between Subjects Std.	0.0118		
Scan - Rescan Std.	0.0107		
Std. due to R2	0.0004		
Std. due to R3	0.0021		
ESPREE - AVANTO	0.0010		
SKYRA - AVANTO	0.0125 ***		
PRISMA - AVANTO	0.0154 ***		
Age(years)	0.0010		
SEX(MALE)	0.0009		
(e) Pallidum			

Table A.1: Regression coefficients as fit by the linear mixed-effect model. * p<0.05; ** p<10⁻²; *** p<10⁻³

A.1. Reproducibility of volumetric data acquired with a standardized T1-weighted brain-imaging protocol

%TIV			%TIV		
ESPREE - AVANTO	-0.7854	***	ESPREE - AVANTO	0.7087	***
PRISMA - AVANTO	-1.0268	***	PRISMA - AVANTO	0.7417	***
SKYRA - AVANTO	-1.0864	***	SKYRA - AVANTO	1.1430	***
PRISMA - ESPREE	-0.2414		PRISMA - ESPREE	0.0330	
SKYRA - ESPREE	-0.3010		SKYRA - ESPREE	0.4343	
SKYRA - PRISMA	-0.0595		SKYRA - PRISMA	0.4013	
(a) WM			(b) GM		
%TIV			%TIV		
ESPREE - AVANTO	0.0767		ESPREE - AVANTO	-0.0770	
PRISMA - AVANTO	0.2853	*	PRISMA - AVANTO	-0.2857	*
SKYRA - AVANTO	-0.0561		SKYRA - AVANTO	0.0566	
PRISMA - ESPREE	0.2086		PRISMA - ESPREE	-0.2087	
SKYRA - ESPREE	-0.1328		SKYRA - ESPREE	0.1336	
SKYRA - PRISMA	-0.3414	*	SKYRA - PRISMA	0.3423	*
(c) CSF			(d) NBV		
%TIV			%TIV		
ESPREE - AVANTO	-0.0011		ESPREE - AVANTO	0.0101	
PRISMA - AVANTO	0.0148	***	PRISMA - AVANTO	0.0630	***
SKYRA - AVANTO	0.0136	***	SKYRA - AVANTO	0.0608	***
PRISMA - ESPREE	0.0159	***	PRISMA - ESPREE	0.0529	***
SKYRA - ESPREE	0.0147	***	SKYRA - ESPREE	0.0508	***
SKYRA - PRISMA	-0.0013		SKYRA - PRISMA	-0.0022	
(e) Hippocampus			(f) Thalamus		
%TIV			%TIV		
ESPREE - AVANTO	0.0018		ESPREE - AVANTO	0.0018	
PRISMA - AVANTO	-0.0143	***	PRISMA - AVANTO	0.0236	***
SKYRA - AVANTO	-0.0206	***	SKYRA - AVANTO	0.0141	***
PRISMA - ESPREE	-0.0161	***	PRISMA - ESPREE	0.0217	***
SKYRA - ESPREE	-0.0224	***	SKYRA - ESPREE	0.0123	***
SKYRA - PRISMA	-0.0063		SKYRA - PRISMA	-0.0094	**
(g) Putamen			(h) Caudate		
%TIV			%TIV		
ESPREE - AVANTO	0.0010				
PRISMA - AVANTO	0.0154	***			
SKYRA - AVANTO	0.0125	***			
PRISMA - ESPREE	0.0144	***			
SKYRA - ESPREE	0.0115	***			
SKYRA - PRISMA	-0.0029				
(i) Pallidum					

Table A.2: Volume differences for all possible scanner combinations. * p<0.05; ** p<10-2; *** p<10-3

A.2 Correction and optimization strategies towards reliable and robust estimation of data elements

A.2.1 Protocol-based compensation strategy

	Adjusted Mean RVD	Original Mean RVD	Target Mean RVD
WM	0.0136	0.0296	0.0062
GM	0.0144	0.0398	0.0058
CSF	0.0146	0.0272	0.0094
NBV	0.0066	0.0115	0.0036
Hippocampus	0.0183	0.0424	0.0113
Thalamus	0.0149	0.0293	0.0082
Putamen	0.0150	0.0156	0.0087
Caudate	0.0197	0.0347	0.0117
Pallidum	0.0241	0.0255	0.0171

Table A.3: Mean relative volume differences 1.5T/3T (Adjusted, Original) and 3T/3T (Target, pure scan-rescan variability) of ADNI dataset with protocol-based compensation strategy computed in leave-one-out cross-validation settings

A.2. Correction and optimization strategies towards reliable and robust estimation of data elements

	Adjusted Mean RVD	Original Mean RVD	Target Mean RVD
WM	0.0137	0.0212	0.0120
GM	0.0094	0.0124	0.0089
CSF	0.0220	0.0224	0.0195
NBV	0.0050	0.0051	0.0045
Hippocampus	0.0193	0.0234	0.0184
Thalamus	0.0140	0.0361	0.0132
Putamen	0.0133	0.0132	0.0116
Caudate	0.0124	0.0241	0.0126
Pallidum	0.0235	0.0304	0.0235

Table A.4: Mean relative volume differences 1.5T/3T (Adjusted, Original) and 3T/3T (Target, pure scan-rescan variability) of interpolated Basel dataset with protocol-based compensation strategy computed in leave-one-out cross-validation settings

Appendix A. Appendix

	Adjusted Mean RVD	Original Mean RVD
WM	0.0053	0.0266
GM	0.0081	0.0122
CSF	0.0198	0.0244
NBV	0.0045	0.0055
Hippocampus	0.0127	0.0120
Thalamus	0.0070	0.0090
Putamen	0.0100	0.0096
Caudate	0.0084	0.0117
Pallidum	0.0168	0.0183

(a) ADNI2-2D-GRAPPA

	Adjusted Mean RVD	Original Mean RVD
WM	0.0052	0.0170
GM	0.0073	0.0076
CSF	0.0169	0.0204
NBV	0.0039	0.0047
Hippocampus	0.0149	0.0144
Thalamus	0.0075	0.0071
Putamen	0.0075	0.0075
Caudate	0.0063	0.0066
Pallidum	0.0127	0.0118

(b) ADNI2-CAIPI

	Adjusted Mean RVD	Original Mean RVD
WM	0.0038	0.0152
GM	0.0092	0.0087
CSF	0.0212	0.0236
NBV	0.0048	0.0053
Hippocampus	0.0099	0.0115
Thalamus	0.0080	0.0076
Putamen	0.0078	0.0076
Caudate	0.0074	0.0078
Pallidum	0.0174	0.0163

(c) ADNI2-CAIPI el

	Adjusted Mean RVD	Original Mean RVD
WM	0.0085	0.0266
GM	0.0113	0.0111
CSF	0.0229	0.0467
NBV	0.0053	0.0103
Hippocampus	0.0165	0.0156
Thalamus	0.0094	0.0102
Putamen	0.0108	0.0103
Caudate	0.0084	0.0083
Pallidum	0.0158	0.0155

(d) ADNI2-Seg.MPRAGE

Table A.5: Mean relative volume differences ADNI-2/accelerated protocols (Original, Adjusted) of Bonn dataset with protocol-based compensation strategy computed in leave-one-out cross-validation settings

A.2. Correction and optimization strategies towards reliable and robust estimation of data elements

	Adjusted Mean RVD	Original Mean RVD	Target Mean RVD
WM	0.0170	0.0212	0.0120
GM	0.0317	0.0124	0.0089
CSF	0.0379	0.0224	0.0195
NBV	0.0131	0.0051	0.0045
Hippocampus	0.0296	0.0234	0.0184
Thalamus	0.0137	0.0361	0.0132
Putamen	0.0140	0.0132	0.0116
Caudate	0.0160	0.0241	0.0126
Pallidum	0.0242	0.0304	0.0235

Table A.6: Mean relative volume differences 1.5T/3T (Original, Adjusted) and 3T/3T (Target pure scan-rescan variability) of interpolated Basel dataset with protocol-based compensation strategy computed when the model is trained on ADNI dataset and applied to scale interpolated Basel dataset

Appendix A. Appendix

	Adjusted Mean RVD	Original Mean RVD	Target Mean RVD
WM	0.0151	0.0296	0.0062
GM	0.0305	0.0398	0.0058
CSF	0.0342	0.0272	0.0094
NBV	0.0130	0.0115	0.0036
Hippocampus	0.0293	0.0424	0.0113
Thalamus	0.0164	0.0293	0.0082
Putamen	0.0153	0.0156	0.0087
Caudate	0.0199	0.0347	0.0117
Pallidum	0.0257	0.0255	0.0171

Table A.7: Mean relative volume differences 1.5T/3T (original/corrected) and 3T/3T (Target, pure scan-rescan variability) of ADNI dataset with protocol-based compensation strategy computed when the model is trained on interpolated Basel dataset and applied to scale ADNI dataset

A.2. Correction and optimization strategies towards reliable and robust estimation of data elements

	Adjusted Mean RVD	Original Mean RVD	Target Mean RVD
WM	0.0139	0.0296	0.0062
GM	0.0188	0.0398	0.0058
CSF	0.0218	0.0272	0.0094
NBV	0.0087	0.0115	0.0036
Hippocampus	0.0216	0.0424	0.0113
Thalamus	0.0154	0.0293	0.0082
Putamen	0.0150	0.0156	0.0087
Caudate	0.0193	0.0347	0.0117
Pallidum	0.0247	0.0255	0.0171

Table A.8: Mean relative volume differences of ADNI dataset 1.5T/3T (Adjusted, Original) and 3T/3T (Target, pure scan-rescan variability). Combined dataset with protocol-based compensation strategy computed in leave-one-out cross-validation settings.

Appendix A. Appendix

	Adjusted Mean RVD	Original Mean RVD	Target Mean RVD
WM	0.0148	0.0212	0.0120
GM	0.0164	0.0124	0.0089
CSF	0.0256	0.0224	0.0195
NBV	0.0071	0.0051	0.0045
Hippocampus	0.0209	0.0234	0.0184
Thalamus	0.0135	0.0361	0.0132
Putamen	0.0134	0.0132	0.0116
Caudate	0.0132	0.0241	0.0126
Pallidum	0.0236	0.0304	0.0235

Table A.9: Mean relative volume differences of interpolated Basel dataset 1.5T/3T (Adjusted, Original) and 3T/3T (Target, pure scan-rescan variability). Combined dataset with protocol-based compensation strategy computed in leave-one-out cross-validation settings.

A.2. Correction and optimization strategies towards reliable and robust estimation of data elements

A.2.2 Image-property-based compensation strategy

	Adjusted Mean RVD	Original Mean RVD
WM	0.0128	0.0296
GM	0.0113	0.0398
CSF	0.0139	0.0272
NBV	0.0061	0.0115
Hippocampus	0.0189	0.0424
Thalamus	0.0153	0.0293
Putamen	0.0157	0.0156
Caudate	0.0183	0.0347
Pallidum	0.0237	0.0255

(a) 3T-1.5T

	Adjusted Mean RVD	Original Mean RVD
WM	0.0065	0.0062
GM	0.0055	0.0058
CSF	0.0089	0.0094
NBV	0.0033	0.0036
Hippocampus	0.0112	0.0113
Thalamus	0.0080	0.0082
Putamen	0.0092	0.0087
Caudate	0.0115	0.0117
Pallidum	0.0181	0.0171

(b) 3T-3T

Table A.10: Mean relative volume differences (Adjusted, Original) 1.5T/3T and 3T/3T of ADNI dataset with image-property-based compensation strategy computed in leave-one-out cross-validation settings

Appendix A. Appendix

	Adjusted Mean RVD	Original Mean RVD
WM	0.0146	0.0212
GM	0.0093	0.0124
CSF	0.0189	0.0224
NBV	0.0043	0.0051
Hippocampus	0.0200	0.0234
Thalamus	0.0140	0.0361
Putamen	0.0134	0.0132
Caudate	0.0131	0.0241
Pallidum	0.0243	0.0304

(a) 3T-1.5T

	Adjusted Mean RVD	Original Mean RVD
WM	0.0111	0.0120
GM	0.0085	0.0089
CSF	0.0165	0.0195
NBV	0.0037	0.0045
Hippocampus	0.0189	0.0184
Thalamus	0.0131	0.0132
Putamen	0.0119	0.0116
Caudate	0.0134	0.0126
Pallidum	0.0236	0.0235

(b) 3T-3T

Table A.11: Mean relative volume differences (Adjusted, Original) 1.5T/3T and 3T/3T of interpolated Basel dataset with image-property-based compensation strategy computed in leave-one-out cross-validation settings

A.2. Correction and optimization strategies towards reliable and robust estimation of data elements

	Adjusted Mean RVD	Original Mean RVD
WM	0.0166	0.0266
GM	0.0121	0.0122
CSF	0.0179	0.0244
NBV	0.0037	0.0055
Hippocampus	0.0126	0.0120
Thalamus	0.0091	0.0090
Putamen	0.0099	0.0096
Caudate	0.0101	0.0117
Pallidum	0.0185	0.0183

(a) ADNI2-2D-GRAPPA

	Adjusted Mean RVD	Original Mean RVD
WM	0.0137	0.0170
GM	0.0065	0.0076
CSF	0.0175	0.0204
NBV	0.0036	0.0047
Hippocampus	0.0140	0.0144
Thalamus	0.0073	0.0071
Putamen	0.0072	0.0075
Caudate	0.0063	0.0066
Pallidum	0.0124	0.0118

(b) ADNI2-CAIPI

	Adjusted Mean RVD	Original Mean RVD
WM	0.0119	0.0152
GM	0.0073	0.0087
CSF	0.0184	0.0236
NBV	0.0039	0.0053
Hippocampus	0.0113	0.0115
Thalamus	0.0078	0.0076
Putamen	0.0075	0.0076
Caudate	0.0084	0.0078
Pallidum	0.0176	0.0163

(c) ADNI2-CAIPI el

	Adjusted Mean RVD	Original Mean RVD
WM	0.0211	0.0266
GM	0.0094	0.0111
CSF	0.0324	0.0467
NBV	0.0068	0.0103
Hippocampus	0.0155	0.0156
Thalamus	0.0105	0.0102
Putamen	0.0102	0.0103
Caudate	0.0086	0.0083
Pallidum	0.0161	0.0155

(d) ADNI2-Seg.MPRAGE

Table A.12: Mean relative volume differences ADNI-2/accelerated protocols (Original, Adjusted) of Bonn dataset with image-property-based compensation strategy computed in leave-one-out cross-validation settings

Appendix A. Appendix

	Adjusted Mean RVD	Original Mean RVD
WM	0.0130	0.0212
GM	0.0169	0.0124
CSF	0.0318	0.0224
NBV	0.0084	0.0051
Hippocampus	0.0251	0.0234
Thalamus	0.0135	0.0361
Putamen	0.0160	0.0132
Caudate	0.0159	0.0241
Pallidum	0.0234	0.0304

(a) 3T-1.5T

	Adjusted Mean RVD	Original Mean RVD
WM	0.0110	0.0120
GM	0.0082	0.0089
CSF	0.0190	0.0195
NBV	0.0042	0.0045
Hippocampus	0.0186	0.0184
Thalamus	0.0131	0.0132
Putamen	0.0117	0.0116
Caudate	0.0142	0.0126
Pallidum	0.0233	0.0235

(b) 3T-3T

Table A.13: Mean relative volume differences (Adjusted, Original) between 1.5T/3T and 3T/3T of interpolated Basel dataset with image-property-based compensation strategy trained on ADNI dataset

A.2. Correction and optimization strategies towards reliable and robust estimation of data elements

	Adjusted Mean RVD	Original Mean RVD
WM	0.0217	0.0296
GM	0.0147	0.0398
CSF	0.0440	0.0272
NBV	0.0083	0.0115
Hippocampus	0.0311	0.0424
Thalamus	0.0188	0.0293
Putamen	0.0175	0.0156
Caudate	0.0204	0.0347
Pallidum	0.0252	0.0255

(a) 3T-1.5T

	Adjusted Mean RVD	Original Mean RVD
WM	0.0111	0.0062
GM	0.0055	0.0058
CSF	0.0206	0.0094
NBV	0.0052	0.0036
Hippocampus	0.0114	0.0113
Thalamus	0.0093	0.0082
Putamen	0.0095	0.0087
Caudate	0.0117	0.0117
Pallidum	0.0188	0.0171

(b) 3T-3T

Table A.14: Mean relative volume differences (Adjusted, Original) between 1.5T/3T and 3T/3T of ADNI dataset with image-property-based compensation strategy trained on interpolated Basel dataset

Appendix A. Appendix

	Adjusted Mean RVD	Original Mean RVD
WM	0.0127	0.0296
GM	0.0106	0.0398
CSF	0.0149	0.0272
NBV	0.0060	0.0115
Hippocampus	0.0191	0.0424
Thalamus	0.0160	0.0293
Putamen	0.0155	0.0156
Caudate	0.0183	0.0347
Pallidum	0.0236	0.0255

(a) 3T-1.5T

	Adjusted Mean RVD	Original Mean RVD
WM	0.0066	0.0062
GM	0.0053	0.0058
CSF	0.0091	0.0094
NBV	0.0034	0.0036
Hippocampus	0.0111	0.0113
Thalamus	0.0082	0.0082
Putamen	0.0091	0.0087
Caudate	0.0115	0.0117
Pallidum	0.0181	0.0171

(b) 3T-3T

Table A.15: Mean relative volume differences of interpolated Basel dataset (Adjusted, Original) between 1.5T/3T and 3T/3T. Combined dataset with image-property-based compensation strategy computed in leave-one-out cross-validation settings.

A.2. Correction and optimization strategies towards reliable and robust estimation of data elements

	Adjusted Mean RVD	Original Mean RVD
WM	0.0134	0.0212
GM	0.0101	0.0124
CSF	0.0216	0.0224
NBV	0.0049	0.0051
Hippocampus	0.0195	0.0234
Thalamus	0.0132	0.0361
Putamen	0.0132	0.0132
Caudate	0.0129	0.0241
Pallidum	0.0236	0.0304

(a) 3T-1.5T

	Adjusted Mean RVD	Original Mean RVD
WM	0.0109	0.0120
GM	0.0083	0.0089
CSF	0.0183	0.0195
NBV	0.0040	0.0045
Hippocampus	0.0190	0.0184
Thalamus	0.0130	0.0132
Putamen	0.0118	0.0116
Caudate	0.0135	0.0126
Pallidum	0.0234	0.0235

(b) 3T-3T

Table A.16: Mean relative volume differences of ADNI dataset (Adjusted, Original) between 1.5T/3T and 3T/3T. Combined dataset with image property based compensation strategy computed in leave-one-out cross-validation settings.

Bibliography

- [1] J. C. Baron, G. Chételat, B. Desgranges, G. Perchey, B. Landeau, V. de la Sayette, and F. Eustache, “In vivo mapping of gray matter loss with voxel-based morphometry in mild Alzheimer’s disease,” *NeuroImage*, vol. 14, no. 2, pp. 298–309, 2001.
- [2] N. Bernasconi, S. Duchesne, A. Janke, J. Lerch, D. L. Collins, and A. Bernasconi, “Whole-brain voxel-based statistical analysis of gray matter and white matter in temporal lobe epilepsy,” *NeuroImage*, vol. 23, no. 2, pp. 717–723, 2004.
- [3] L. Bonilha, C. Rorden, J. J. Halford, M. Eckert, S. Appenzeller, F. Cendes, and L. M. Li, “Asymmetrical extra-hippocampal grey matter loss related to hippocampal atrophy in patients with medial temporal lobe epilepsy,” *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 78, no. 3, pp. 286–294, 2006.
- [4] E. J. Burton, I. G. McKeith, D. J. Burn, E. D. Williams, and J. T. O’Brien, “Cerebral atrophy in Parkinson’s disease with and without dementia: A comparison with Alzheimer’s disease, dementia with Lewy bodies and controls,” *Brain*, vol. 127, no. 4, pp. 791–800, 2004.
- [5] R. Camicioli, M. Gee, T. P. Bouchard, N. J. Fisher, C. C. Hanstock, D. J. Emery, and W. R. W. Martin, “Voxel-based morphometry reveals extra-nigral atrophy patterns associated with dopamine refractory cognitive and motor impairment in parkinsonism,” *Parkinsonism & related disorders*, vol. 15, pp. 187–95, mar 2009.
- [6] G. Chételat, B. Landeau, F. Eustache, F. Mézange, F. Viader, V. De La Sayette, B. Desgranges, and J. C. Baron, “Using voxel-based morphometry to map the structural changes associated with rapid conversion in MCI: A longitudinal MRI study,” *NeuroImage*, vol. 27, no. 4, pp. 934–946, 2005.
- [7] A. Cifelli, M. Arridge, P. Jezzard, M. M. Esiri, J. Palace, and P. M. Matthews, “Thalamic neurodegeneration in multiple sclerosis,” *Annals of Neurology*, vol. 52, no. 5, pp. 650–653, 2002.
- [8] G. B. Frisoni, N. C. Fox, C. R. Jack, P. Scheltens, and P. M. Thompson, “The clinical use of structural MRI in Alzheimer disease,” *Nat Rev Neurol*, vol. 6, pp. 67–77, feb 2010.
- [9] C. R. Jack, V. J. Lowe, S. D. Weigand, H. J. Wiste, M. L. Senjem, D. S. Knopman, M. M. Shiung, J. L. Gunter, B. F. Boeve, B. J. Kemp, M. Weiner, and R. C. Petersen, “Serial PIB

Bibliography

- and MRI in normal, mild cognitive impairment and Alzheimers disease: Implications for sequence of pathological events in Alzheimers disease,” *Brain*, vol. 132, no. 5, pp. 1355–1365, 2009.
- [10] R. Kikinis, C. R. Guttmann, D. Metcalf, W. M. Wells, G. J. Ettinger, H. L. Weiner, and F. A. Jolesz, “Quantitative follow-up of patients with multiple sclerosis using MRI: technical aspects.,” *Journal of Magnetic Resonance Imaging*, vol. 9, no. 4, pp. 519–30, 1999.
- [11] M. P. Sanfilipo, R. H. B. Benedict, B. Weinstock-Guttman, and R. Bakshi, “Gray and white matter brain atrophy and neuropsychological impairment in multiple sclerosis,” *Neurology*, vol. 66, pp. 685–692, mar 2006.
- [12] D. Schmitter, A. Roche, B. Maréchal, D. Ribes, A. Abdulkadir, M. Bach-Cuadra, A. Dadducci, C. Granziera, S. Klöppel, P. Maeder, R. Meuli, and G. Krueger, “An evaluation of volume-based morphometry for prediction of mild cognitive impairment and Alzheimer’s disease,” *NeuroImage. Clinical*, vol. 7, pp. 7–17, jan 2015.
- [13] G. Tao, S. Datta, R. He, F. Nelson, J. S. Wolinsky, and P. A. Narayana, “Deep Gray Matter Atrophy in Multiple Sclerosis: A Tensor Based Morphometry,” jul 2009.
- [14] J. L. Whitwell, S. D. Weigand, M. M. Shiung, B. F. Boeve, T. J. Ferman, G. E. Smith, D. S. Knopman, R. C. Petersen, E. E. Benarroch, K. A. Josephs, and C. R. Jack, “Focal atrophy in Dementia with Lewy Bodies on MRI: a distinct pattern from Alzheimer’s disease,” mar 2007.
- [15] I. C. Wright, Z. R. Ellison, T. Sharma, K. J. Friston, R. M. Murray, and P. K. McGuire, “Mapping of grey matter changes in schizophrenia,” *Schizophrenia Research*, vol. 35, no. 1, pp. 1–14, 1999.
- [16] C. R. Jack, M. A. Bernstein, N. C. Fox, P. Thompson, G. Alexander, D. Harvey, B. Borowski, P. J. Britson, J. L. Whitwell, C. Ward, A. M. Dale, J. P. Felmlee, J. L. Gunter, D. L. Hill, R. Killiany, N. Schuff, S. Fox-Bosetti, C. Lin, C. Studholme, C. S. DeCarli, G. Krueger, H. A. Ward, G. J. Metzger, K. T. Scott, R. Mallozzi, D. Blezek, J. Levy, J. P. Debbins, A. S. Fleisher, M. Albert, R. Green, G. Bartzokis, G. Glover, J. Mugler, and M. W. Weiner, “The Alzheimer’s Disease Neuroimaging Initiative (ADNI): MRI methods,” *Journal of magnetic resonance imaging: JMRI*, vol. 27, pp. 685–91, apr 2008.
- [17] C. R. Jack, M. A. Bernstein, B. J. Borowski, J. L. Gunter, N. C. Fox, P. M. Thompson, N. Schuff, G. Krueger, R. J. Killiany, C. S. Decarli, A. M. Dale, O. W. Carmichael, D. Tosun, and M. W. Weiner, “Update on the magnetic resonance imaging core of the Alzheimer’s disease neuroimaging initiative,” *Alzheimer’s & dementia: the journal of the Alzheimer’s Association*, vol. 6, pp. 212–20, may 2010.
- [18] C. R. K. Ching, X. Hua, D. P. Hibar, C. P. Ward, J. L. Gunter, M. A. Bernstein, C. R. Jack, M. W. Weiner, and P. M. Thompson, “Does MRI scan acceleration affect power to track brain change?,” *Neurobiology of aging*, vol. 36 Suppl 1, pp. S167–77, jan 2015.

- [19] J. Jovicich, S. Czanner, D. Greve, E. Haley, A. van der Kouwe, R. Gollub, D. Kennedy, F. Schmitt, G. Brown, J. Macfall, B. Fischl, and A. Dale, "Reliability in multi-site structural MRI studies: effects of gradient non-linearity correction on phantom and human data," *NeuroImage*, vol. 30, pp. 436–43, apr 2006.
- [20] J. Jovicich, S. Czanner, X. Han, D. Salat, A. van der Kouwe, B. Quinn, J. Pacheco, M. Albert, R. Killiany, D. Blacker, P. Maguire, D. Rosas, N. Makris, R. Gollub, A. Dale, B. C. Dickerson, and B. Fischl, "MRI-derived measurements of human subcortical, ventricular and intracranial brain volumes: Reliability effects of scan sessions, acquisition sequences, data analyses, scanner upgrade, scanner vendors and field strengths," *NeuroImage*, vol. 46, pp. 177–92, may 2009.
- [21] J. Jovicich, M. Marizzoni, R. Sala-Llonch, B. Bosch, D. Bartrés-Faz, J. Arnold, J. Benninghoff, J. Wiltfang, L. Roccatagliata, F. Nobili, T. Hensch, A. Tränkner, P. Schönknecht, M. Leroy, R. Lopes, R. Bordet, V. Chanoine, J.-P. Ranjeva, M. Didic, H. Gros-Dagnac, P. Payoux, G. Zoccatelli, F. Alessandrini, A. Beltramello, N. Bargalló, O. Blin, and G. B. Frisoni, "Brain morphometry reproducibility in multi-center 3T MRI studies: a comparison of cross-sectional and longitudinal segmentations," *NeuroImage*, vol. 83, pp. 472–84, dec 2013.
- [22] G. Krueger, C. Granziera, C. R. Jack, J. L. Gunter, A. Littmann, B. Mortamet, S. Kanengiesser, A. G. Sorensen, C. P. Ward, D. A. Reyes, P. J. Britson, H. Fischer, and M. A. Bernstein, "Effects of MRI scan acceleration on brain volume measurement consistency," *Journal of magnetic resonance imaging : JMRI*, vol. 36, pp. 1234–40, nov 2012.
- [23] F. Kruggel, J. Turner, and L. T. Muftuler, "Impact of scanner hardware and imaging protocol on image quality and compartment volume precision in the ADNI cohort," *NeuroImage*, vol. 49, pp. 2123–33, feb 2010.
- [24] K. K. Leung, I. M. Malone, S. Ourselin, J. L. Gunter, M. a. Bernstein, P. M. Thompson, C. R. Jack, M. W. Weiner, and N. C. Fox, "Effects of changing from non-accelerated to accelerated MRI for follow-up in brain atrophy measurement," *NeuroImage*, vol. 107C, pp. 46–53, dec 2014.
- [25] R. A. Morey, E. S. Selgrade, H. R. Wagner, S. A. Huettel, L. Wang, and G. McCarthy, "Scan-rescan reliability of subcortical brain volumes derived from automated segmentation," *Human brain mapping*, vol. 31, pp. 1751–62, nov 2010.
- [26] M. Reuter, N. J. Schmansky, H. D. Rosas, and B. Fischl, "Within-subject template estimation for unbiased longitudinal image analysis," *NeuroImage*, vol. 61, pp. 1402–18, jul 2012.
- [27] J. S. Wonderlick, D. A. Ziegler, P. Hosseini-Varnamkhasti, J. J. Locascio, A. Bakkour, A. van der Kouwe, C. Triantafyllou, S. Corkin, and B. C. Dickerson, "Reliability of MRI-derived cortical and subcortical morphometric measures: effects of pulse sequence, voxel geometry, and parallel imaging," *NeuroImage*, vol. 44, pp. 1324–33, feb 2009.

Bibliography

- [28] J. P. Mugler and J. R. Brookeman, "Three-dimensional magnetization-prepared rapid gradient-echo imaging (3D MP RAGE)," *Magnetic Resonance in Medicine*, vol. 15, no. 1, pp. 152–157, 1990.
- [29] M. A. Griswold, P. M. Jakob, R. M. Heidemann, M. Nittka, V. Jellus, J. Wang, B. Kiefer, and A. Haase, "Generalized autocalibrating partially parallel acquisitions (GRAPPA)," *Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine*, vol. 47, pp. 1202–10, jun 2002.
- [30] K. P. Pruessmann, M. Weiger, M. B. Scheidegger, and P. Boesiger, "SENSE: Sensitivity encoding for fast MRI," *Magnetic Resonance in Medicine*, vol. 42, no. 5, pp. 952–962, 1999.
- [31] C. Studholme, D. L. G. Hill, and D. J. Hawkes, "An overlap invariant entropy measure of 3D medical image alignment," *Pattern Recognition*, vol. 32, pp. 71–86, 1999.
- [32] C. Chef d'hotel, G. Hermosillo, and O. Faugeras, "Flows of diffeomorphisms for multi-modal image registration," in *Biomedical Imaging, 2002. Proceedings. 2002 IEEE International Symposium on*, pp. 753–756, 2002.
- [33] J. Ashburner and K. J. Friston, "Unified segmentation," *NeuroImage*, vol. 26, no. 3, pp. 839–851, 2005.
- [34] M. B. Cuadra, L. Cammoun, T. Butz, O. Cuisenaire, and J. P. Thiran, "Comparison and validation of tissue modelization and statistical classification methods in T1-weighted MR brain images," *IEEE Transactions on Medical Imaging*, vol. 24, no. 12, pp. 1548–1565, 2005.
- [35] K. Van Leemput, F. Maes, D. Vandermeulen, and P. Suetens, "Automated model-based tissue classification of MR images of the brain.," *IEEE transactions on medical imaging*, vol. 18, no. 10, pp. 897–908, 1999.
- [36] M. Albert, S. DeKosky, D. Dickson, B. Dubois, H. Feldman, N. Fox, a. Gamst, D. Holtzman, W. Jagust, R. C. Petersen, P. Snyder, M. Carrillo, B. Thies, and C. Phelps, "The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on," *Alzheimer's & . . .*, vol. 7, no. 3, pp. 270–279, 2011.
- [37] G. M. McKhann, D. S. Knopman, H. Chertkow, B. T. Hyman, C. R. Jack, C. H. Kawas, W. E. Klunk, W. J. Koroshetz, J. J. Manly, R. Mayeux, R. C. Mohs, J. C. Morris, M. N. Rossor, P. Scheltens, M. C. Carrillo, B. Thies, S. Weintraub, and C. H. Phelps, "The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease," *Alzheimer's & Dementia*, vol. 7, no. 3, pp. 263–269, 2011.
- [38] D. O. Walsh, A. F. Gmitro, and M. W. Marcellin, "Adaptive reconstruction of phased array MR imagery," *Magnetic Resonance in Medicine*, vol. 43, no. 5, pp. 682–690, 2000.

-
- [39] A. Roche, D. Ribes, M. Bach-Cuadra, and G. Krüger, “On the convergence of EM-like algorithms for image segmentation using Markov random fields,” *Medical Image Analysis*, vol. 15, no. 6, pp. 830–9, 2011.
- [40] F. Wilcoxon, “Individual Comparisons by Ranking Methods,” *Biometrics Bulletin*, vol. 1, no. 6, pp. 80–83, 1945.
- [41] P. M. Robson, A. K. Grant, A. J. Madhuranthakam, R. Lattanzi, D. K. Sodickson, and C. a. McKenzie, “Comprehensive quantification of signal-to-noise ratio and g-factor for image-based and k-space-based parallel imaging reconstructions,” *Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine*, vol. 60, pp. 895–907, oct 2008.
- [42] C. N. Wiens, S. J. Kisch, J. D. Willig-Onwuachi, and C. A. McKenzie, “Computationally rapid method of estimating signal-to-noise ratio for phased array image reconstructions,” *Magnetic Resonance in Medicine*, vol. 66, pp. 1192–1197, oct 2011.
- [43] G. C. Wiggins, C. Triantafyllou, A. Potthast, A. Reykowski, M. Nittka, and L. L. Wald, “32-Channel 3 tesla receive-only phased-array head coil with soccer-ball element geometry,” *Magnetic Resonance in Medicine*, vol. 56, no. 1, pp. 216–223, 2006.
- [44] D. Bates, M. Mächler, B. Bolker, and S. Walker, “Fitting Linear Mixed-Effects Models Using {lme4},” *Journal of Statistical Software*, vol. 67, no. 1, pp. 1–48, 2015.
- [45] J. L. Bernal-Rusiel, D. N. Greve, M. Reuter, B. Fischl, and M. R. Sabuncu, “Statistical analysis of longitudinal neuroimage data with Linear Mixed Effects models,” *NeuroImage*, vol. 66, pp. 249–260, 2013.
- [46] A. Kuznetsova, P. B. Brockhoff, and R. H. B. Christensen, “lmerTest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package),” *R package version*, vol. 2, no. 6, 2013.
- [47] O. J. Dunn, “Multiple Comparisons Among Means,” *Journal of the American Statistical Association*, vol. 56, no. 293, pp. 52–64, 1961.
- [48] K. P. Cosgrove, C. M. Mazure, and J. K. Staley, “Evolving Knowledge of Sex Differences in Brain Structure, Function, and Chemistry,” *Biological Psychiatry*, vol. 62, pp. 847–855, jul 2007.
- [49] R. C. Gur, B. I. Turetsky, M. Matsui, M. Yan, W. Bilker, P. Hughett, and R. E. Gur, “Sex Differences in Brain Gray and White Matter in Healthy Young Adults: Correlations with Cognitive Performance,” *The Journal of Neuroscience*, vol. 19, no. 10, pp. 4065–4072, 1999.
- [50] P. Falkovskiy, D. Brenner, T. Feiweier, S. Kannengiesser, B. Maréchal, T. Kober, A. Roche, K. Thostenson, R. Meuli, D. Reyes, T. Stoecker, M. A. Bernstein, J.-P. Thiran, and G. Krueger, “Comparison of accelerated T1-weighted whole-brain structural-imaging protocols,” *NeuroImage*, vol. 124, Part, pp. 157–167, 2016.

Bibliography

- [51] M. Blaimer, F. A. Breuer, M. Mueller, N. Seiberlich, D. Ebel, R. M. Heidemann, M. A. Griswold, and P. M. Jakob, "2D-GRAPPA-operator for faster 3D parallel MRI," *Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine*, vol. 56, pp. 1359–64, dec 2006.
- [52] D. Brenner, R. Stirnberg, E. D. Pracht, and T. Stöcker, "Two-dimensional accelerated MP-RAGE imaging with flexible linear reordering," *Magma (New York, N.Y.)*, vol. 27, pp. 455–62, oct 2014.
- [53] F. A. Breuer, M. Blaimer, M. F. Mueller, N. Seiberlich, R. M. Heidemann, M. A. Griswold, and P. M. Jakob, "Controlled aliasing in volumetric parallel imaging (2D CAIPIRINHA)," *Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine*, vol. 55, pp. 549–56, mar 2006.
- [54] P. Falkovskiy, T. Kober, D. Reyes, K. Steinert, M. Seeger, M. Bernstein, and G. Krueger, "Segmented Multi-Echo MPRAGE Acquisition for Accelerated T1-weighted Brain Imaging," in *ISMRM*, vol. 27, 2013.
- [55] C. R. Jack, "Alliance for aging research AD biomarkers work group: structural MRI," *Neurobiology of aging*, vol. 32 Suppl 1, pp. S48–57, dec 2011.
- [56] M. I. Miller, C. E. Priebe, A. Qiu, B. Fischl, A. Kolasny, T. Brown, Y. Park, J. T. Ratnanather, E. Busa, J. Jovicich, P. Yu, B. C. Dickerson, and R. L. Buckner, "Collaborative computational anatomy: an MRI morphometry study of the human brain via diffeomorphic metric mapping," *Human brain mapping*, vol. 30, pp. 2132–41, jul 2009.
- [57] K. L. Mills and C. K. Tamnes, "Methods and considerations for longitudinal structural brain imaging analysis across development," *Developmental cognitive neuroscience*, vol. 9C, pp. 172–190, jul 2014.
- [58] M. Bernstein, "Effect of windowing and zero-filled reconstruction of MRI data on spatial resolution and acquisition strategy," *Journal of Magnetic . . .*, vol. 280, pp. 270–280, 2001.
- [59] P. Shrout and J. Fleiss, "Intraclass correlations: Uses in assessing rater reliability," *Psychological Bulletin*, vol. 86, no. 2, pp. 420–428, 1979.
- [60] A. Roche, G. Malandain, and N. Ayache, "Unifying maximum likelihood approaches in medical image registration," *International Journal of Imaging Systems and Technology*, vol. 11, no. 1, pp. 71–80, 2000.
- [61] R. F. Busse, A. C. S. Brau, A. Vu, C. R. Michelich, E. Bayram, R. Kijowski, S. B. Reeder, and H. A. Rowley, "Effects of refocusing flip angle modulation and view ordering in 3D fast spin echo," *Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine*, vol. 60, pp. 640–9, sep 2008.

-
- [62] A. M. Dale, B. Fischl, and M. I. Sereno, "Cortical Surface-Based Analysis: I. Segmentation and Surface Reconstruction," *NeuroImage*, vol. 9, no. 2, pp. 179–194, 1999.
- [63] B. Fischl, M. I. Sereno, and A. M. Dale, "Cortical Surface-Based Analysis: II: Inflation, Flattening, and a Surface-Based Coordinate System," *NeuroImage*, vol. 9, no. 2, pp. 195–207, 1999.
- [64] A. J. W. van der Kouwe, T. Benner, D. H. Salat, and B. Fischl, "Brain morphometry with multiecho MPRAGE," *NeuroImage*, vol. 40, pp. 559–69, apr 2008.
- [65] A. Macovski, "Noise in MRI," *Magnetic Resonance in Medicine*, vol. 36, no. 3, pp. 494–497, 1996.
- [66] J. P. Fortin, E. M. Sweeney, J. Muschelli, C. M. Crainiceanu, and R. T. Shinohara, "Removing inter-subject technical variability in magnetic resonance imaging studies," *NeuroImage*, vol. 132, pp. 198–212, 2016.
- [67] R. T. Shinohara, E. M. Sweeney, J. Goldsmith, N. Shiee, F. J. Mateen, P. A. Calabresi, S. Jarso, D. L. Pham, D. S. Reich, and C. M. Crainiceanu, "Statistical normalization techniques for magnetic resonance imaging," *NeuroImage: Clinical*, vol. 6, pp. 9–19, 2014.
- [68] A. Keshavan, F. Paul, M. K. Beyer, A. H. Zhu, N. Papinutto, R. T. Shinohara, W. Stern, M. Amann, R. Bakshi, A. Bischof, A. Carriero, M. Comabella, J. C. Crane, S. D'Alfonso, P. Demaerel, B. Dubois, M. Filippi, V. Fleischer, B. Fontaine, L. Gaetano, A. Goris, C. Graetz, A. Gröger, S. Groppa, D. A. Hafler, H. F. Harbo, B. Hemmer, K. Jordan, L. Kappos, G. Kirkish, S. Llufriu, S. Magon, F. Martinelli-Boneschi, J. McCauley, X. Montalban, M. Muhlau, D. Pelletier, P. M. Pattany, M. Pericak-Vance, I. Rebeix, M. Rocca, A. Rovira, R. Schlaeger, A. Saiz, T. Sprenger, A. Stecco, B. M. J. Uitdehaag, P. Villoslada, M. P. Wattjes, H. Weiner, J. Wuerfel, C. Zimmer, F. Zipp, International Multiple Sclerosis Genetics Consortium. Electronic address: AIVINSON@PARTNERS.ORG, S. Hauser, J. R. Oksenberg, and R. G. Henry, "Power estimation for non-standardized multisite studies.," *NeuroImage*, vol. 134, pp. 281–294, 2016.
- [69] D. H. Salat, S. Y. Lee, A. J. van der Kouwe, D. N. Greve, B. Fischl, and H. D. Rosas, "Age-associated alterations in cortical gray and white matter signal intensity and gray to white matter contrast," *NeuroImage*, vol. 48, no. 1, pp. 21–28, 2009.
- [70] B. Mortamet, M. a. Bernstein, C. R. Jack, J. L. Gunter, C. Ward, P. J. Britson, R. Meuli, J.-P. Thiran, and G. Krueger, "Automatic quality assessment in structural brain magnetic resonance imaging.," *Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine*, vol. 62, pp. 365–72, aug 2009.

Pavel Falkovskiy

pfalkovskiy@gmail.com

Ave. Louis-Ruchonnet 35, 1003, Lausanne, Switzerland

+41-78-915-78-91

+1-403-288-2076

SUMMARY OF QUALIFICATIONS

- Very strong background in physics and numerical methods
- Strong knowledge of C\C++, CUDA\OpenCL, Python, Matlab, R
- Experienced in MR image acquisition\reconstruction
- Languages: English, Russian, French

EDUCATION

Candidate for Ph.D. in Electrical Engineering

École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, 2012-present

Thesis: Assessment of Image Quality Requirements in Magnetic Resonance Imaging for Quantitative Brain Morphometry

Master of Science in Physics and Applications (Complex Systems)

Université Pierre-et-Marie-Curie, Paris, France, 2011-2012

Honours Bachelor of Science

Honours Physics, Biophysics specialization

University of Waterloo, Waterloo, Ontario, Canada, 2004-2009

WORK EXPERIENCE

2012-present Research Assistant

Siemens Healthcare, Renens, Switzerland

- Designed and implemented pulse sequences for accelerated magnetic resonance brain imaging(MRI)
- Designed and implemented a navigator for motion correction framework
- Developed automated image quality assessment routines

2011 Research Assistant

General Electric Healthcare, Buc, France

- Developed CT image reconstruction algorithm for a GPU using C++\CUDA
- Ported image preprocessing routines to a GPU
- Ported legacy linear algebra library to a GPU

2009-2010 Computer Programmer in Research and Development Department

Athias Geophysical, Calgary, Alberta, Canada

- Developed applications in C++ for seismic data processing and image reconstruction

2007-2008 Research Assistant

Sunnybrook Health Sciences Centre, Toronto, Ontario, Canada

2006 Computer Programmer in Research and Development Department

Sensors & Software Inc., Mississauga, Ontario, Canada

Publications

Journals

P. Falkovskiy, D. Brenner, T. Feiweier, S. Kannengiesser, B. Maréchal, T. Kober, A. Roche, K. Thostenson, R. Meuli, D. Reyes, T. Stoecker, M. A. Bernstein, J.-P. Thiran, and G. Krueger, "Comparison of accelerated T1-weighted whole-brain structural-imaging protocols," *Neuroimage*, vol. 124, pp. 157–167, 2016.

M. Waszak*, P. Falkovskiy*, T. Kober, B. Marechal, R. Mueli, R. Gruetter, and G. Krueger, "Prospective Head Motion Correction with FID-Triggered Image Navigators," *Magnetic resonance in medicine*, 2016.

S. Haller*, P. Falkovskiy*, R. Meuli, J.-P. Thiran, G. Krueger, K.-O. Lovblad, T. Kober, A. Roche, and B. Marechal, "Basic MR sequence parameters systematically bias automated brain volume estimation," *Neuroradiology*, 2016.

Conferences

P. Falkovskiy, T. Kober, D. Reyes, K. Steinert, M. Seeger, M. Bernstein, and G. Krueger, "Segmented Multi-Echo MPRAGE Acquisition for Accelerated T1-weighted Brain Imaging," in *Proc. Intl. Soc. Mag. Reson. Med.* 21, 2013.

P. Falkovskiy, D. Brenner, T. Feiweier, S. Kannengiesser, B. Maréchal, T. Kober, A. Roche, K. Thostenson, M. Seeger, T. Stoecker, M. Bernstein, and G. Krueger, "Comparison of accelerated T1-weighted whole-brain structural imaging protocols," in *Proc. Intl. Soc. Mag. Reson. Med.* 22, 2014.

M. Babayeva, A. Loktyushin, P. Falkovskiy, T. Kober, R. Mueli, R. Gruetter, and G. Krueger, "FID navigator triggered acquisition of imaging navigators for retrospective head motion correction," in *ISMRM workshop on motion correction*, in Tromso, Norway, 2014.

T. Hilbert, T. Kober, T. J. Sumpf, Z. Tan, J. Frahm, P. Falkovskiy, H. Meyer, R. Bendl, J. Thiran, R. Meuli, and G. Krueger, "MARTINI and GRAPPA - When Speed is Taste," in *Proc. Intl. Soc. Mag. Reson. Med.* 22, 2014.

Appendix A. Publications

M. Babayeva, P. Falkovskiy, T. Hilbert, G. Bonnier, B. Maréchal, R. Meuli, J. Thiran, R. Gruetter, G. Krueger, and T. Kober, “Prospective motion correction with FID-triggered image navigators,” in *Proc. Intl. Soc. Mag. Reson. Med.* 23, 2015, vol. 66, no. 1

B. Maréchal, S. Kannengiesser, K. Thostenson, P. Kollasch, P. Falkovskiy, J. Thiran, R. Meuli, and M. A. Bernstein, “A generalized method for automated quality assessment in brain MRI Introduction;,” in *Proc. Intl. Soc. Mag. Reson. Med.* 23, 2015.

P. Falkovskiy, B. Maréchal, S. Yan , Z. Jin , T. Qian , K. O’Brien , R. Meuli, J. Thiran , G. Krueger , T. Kober , and A. Roche , “Quantitative comparison of MP2RAGE skull-stripping strategies,” in *Proc. Intl. Soc. Mag. Reson. Med.* 24, 2016.

P. Falkovskiy, B. Maréchal, T. Kober, P. Maeder, R. Meuli, J. Thiran , and A. Roche , “Impact of image acquisition systems on Alzheimer’s disease-related atrophy detection,” in *Proc. Intl. Soc. Mag. Reson. Med.* 24, 2016.

M. Amann, P. Falkovskiy, A. Thoeni, T. Kober, A. Roche, B. Maréchal, P. Cattin, T. Heye, O. Bieri, T. Sprenger, C. Stippich, G. Krueger, E. Radue, and J. Wuerfel, “Impact of image acquisition systems on Alzheimer’s disease-related atrophy detection,” in *Proc. Intl. Soc. Mag. Reson. Med.* 24, 2016.

S. Haller, P. Falkovskiy, R. Meuli, J. Thiran, G. Krueger, K. Lovblad, A. Roche, T. Kober, and B. Maréchal, “Basic MR sequence parameters systematically bias automated brain volume estimation,” in *Proc. Intl. Soc. Mag. Reson. Med.* 24, 2016.

