UNIVERSIDADE DE LISBOA

FACULDADE DE CIÊNCIAS

DEPARTAMENTO DE FÍSICA

Quantitative comparison of multi-centre MRI data for mild to severe Traumatic Brain Injury

Mestrado Integrado em Engenharia Biomédica e Biofísica
Perfil em Engenharia Clínica e Instrumentação Médica

Liane dos Santos Canas

Dissertação orientada por:
Dr. Marta Correia, MRC Cognition and Brain Sciences Unit, University of Cambridge, United Kingdom
Dr. Rita Nunes, Instituto de Biofísica e Engenharia Biomédica, Departamento de Física da Faculdade de Ciências da Universidade de Lisboa, Portugal

2015

Faculdade de Ciências da Universidade de Lisboa

Departamento de Física

Liane dos Santos Canas

Quantitative comparison of multi-centre MRI data for mild to severe Traumatic Brain Injury

MESTRADO INTEGRADO EM ENGENHARIA BIOMÉDICA E BIOFÍSICA

Perfil em Engenharia Clínica e Instrumentação Médica

Dissertação orientada por:

Dr. Marta Correia, MRC Cognition and Brain Sciences Unit, University of Cambridge, United Kingdom
Dr. Rita Nunes, Instituto de Biofísica e Engenharia Biomédica, Departamento de Física da Faculdade de Ciências da Universidade de Lisboa, Portugal

2015

To the ones that always had the courage to dream.

But mainly to the ones that always have had the strength to support the dreamers.

RESUMO

O trauma cerebral (frequentemente denominado de TBI – *Traumatic Brain Injury*) é uma das principais causas de morte e incapacidade em jovens adultos, afectando cerca de 2,5 milhões de sujeitos, por ano, só na Europa, sendo que 75 000 acabam por morrer em consequência do mesmo. Actualmente, TBI é classificada pela Organização Mundial de Saúde como uma epidemia silenciosa e um grave problema de saúde pública.

O diagnóstico do trauma cerebral assenta em parâmetros como a escala de coma de Glasgow ou o nível de perda de consciência, o que condiciona um diagnóstico exacto, já que por vezes os sintomas associados ao trauma cerebral não se manifestam de imediato. Assim, as técnicas de imagem médicas derivadas da Ressonância Magnética, como é o caso das imagens obtidas usando o tensor de difusão - DTI (*Diffusion Tensor Imaging*) – e imagens de Ressonância Magnética funcional – fMRI (*Functional Magnetic Resonance Imaging*) - têm-se mostrado bastante relevantes para um diagnóstico mais eficaz do TBI.

TBI trata-se de um conjunto de reacções a um agressão externa, que dependem de diversos factores e que por esse motivo tornam difícil definir TBI, assim como a melhor abordagem para o seu tratamento. Por conseguinte, uma das formas mais eficazes de estudar TBI, tentando definir um tratamento adequado, é através de estudos longitudinais, que permitam, abrangendo um número alargado de pacientes, melhor caracterizar esta patologia. É neste sentido que surge o projecto CENTER-TBI.

O CENTER-TBI é um estudo multicentro que visa, por meio da aquisição de dados em 60 centros e abrangendo um total de 5400 indivíduos, uma caracterização mais eficaz do trauma cerebral, assim como a identificação da intervenção mais eficaz para o tratamento de TBI.

Considerando as mais-valias que um estudo multicentro oferece, entre as quais se destaca o aumento da população em estudo o que permite um aumento de poder estatístico dos testes, assim como a garantia de que a população é o mais heterógena possível permitindo a análise de diferentes características e sintomas associados ao TBI, será possível definir uma abordagem que esclareça a comunidade médica sobre como proceder perante um paciente com trauma cerebral.

No entanto, a viabilidade deste tipo de estudos, e do projecto CENTER-TBI em particular, está fortemente dependente da reprodutibilidade dos dados. Para tal, foi definido qual o procedimento a adoptar para a aquisição de imagens médicas, estabelecendo-se protocolos que definem qual a sequência de aquisição de MRI. Porém, seja por incapacidade de implementação da sequência tal como está definida, seja por características intrínsecas ao *scanner* utilizado, existe uma fracção de variabilidade em cada imagem adquirida que é inerente ao scanner. Tal facto introduz um viés nos dados que impossibilita que os mesmos sejam exactamente reprodutíveis e comparáveis entre scanners.

Este projecto tem como objectivo principal a redução da variabilidade entre dados provenientes de *scanners* diferentes, assegurando a reprodutibilidade dos mesmos.

Assim, numa primeira fase, após uma extensa pesquisa sobre o estado da arte relativo a estudos multicentro, procedeu-se ao desenvolvimento de algoritmos que permitam a quantificação da variabilidade que é originada pelas singularidades do *hardware*. De seguida, procedeu-se à comparação não só dados provenientes de sujeitos saudáveis, definidos como grupo de controlo, mas também dados provenientes de pacientes com TBI. Desse modo, foi possível estimar o quanto a variabilidade introduzida pelo *scanner* afecta o diagnóstico de pacientes.

Posteriormente, após terminada a fase de quantificação, procedeu-se à aplicação de dois métodos distintos de correcção de variabilidade pelo *hardware,* sendo que no segundo caso, ao método testado foram introduzidas várias variações nas quais se tentou obter uma melhor performance do algoritmo em análise. Esperou-se que com a aplicação de ambos os métodos de correcção, as diferenças encontradas entre os vários *scans,* adquiridos em diferentes centros, se ficam a dever exclusivamente a

diferenças anatómicas e/ou fisiológicas entre pacientes, permitindo desse modo a comparação dos diferentes indivíduos em análise. Assim, as conclusões e pressupostos assumidos tendo por base estudos e análises de dados multicentro terão a sua fiabilidade assegurada.

Em suma, este projecto teve como objectivo a quantificação e correcção da variabilidade entre *scanners*, dado que esta se pode tornar um factor de erro com ênfase suficiente para colocar em causa a fiabilidade de estudos multicentro.

Palavras-chave: Estudos Multicentro, Técnicas de imagem médica, Variabilidade, Correcção da variabilidade introduzida pelo *scanner*, Trauma Cerebral.

ABSTRACT

Introduction: Multicentre studies have proven themselves very useful to collect data from subjects with interesting characteristics for *Traumatic Brain Injury* (TBI) research, such as heterogeneous approaches for TBI treatment. Multicentre projects have been contributing to increase the statistical power of the TBI studies and to improve the reliability of the assumptions held about this illness.

It is necessary to ensure the data reproducibility in order to guarantee the studies reliability. In this way, it is crucial to remove any source of error in these projects, such as variability and bias present in the data, introduced by the hardware. The present dissertation presents a project whose main goal is to quantify and correct for the variability introduced by the hardware. Therefore, the first step to guarantee the viability of the results is to measure the variability present in the data. Second, the data needed to be corrected so as to eliminate the sources of variability, considering the several approaches suggested in the literature. It was also important to determine the best approach to achieve the lowest level of variability in the data, without removing relevant features regarding pathology.

Materials and Methods: In an initial phase of the project, a quantification of the variability across scanners and within centre was performed. For that, the Coefficient of Variation (COV) was calculated for each type of maps in analysis – Fractional Anisotropy (FA), Mean Diffusivity (MD), Grey Matter (GM) and White Matter (WM). A voxel based analysis and Regions of Interest (ROI) analysis was performed in order to characterize the variability present in the data and to confirm the need for the use of a correction model to remove the variability introduced by the hardware in multicentre studies. After that, two methodologies to correct the variability were tested. In the case of the second methodology applied, variations of the initial model were also tested in order to improve the performance of this model. Finally, a comparison of the effectiveness of the methods tested was performed. For that, a Support Vector Machine (SVM) algorithm was applied to obtain an indirect measure of the accuracy of the methods tested.

Results: The results obtained in this initial phase were as expected, in agreement with previous studies described in the literature. The models tested were effective in the correction of the variability introduced by the scanner. The Regression Models showed the best performance in the correction of the variability. However, the Spatial Filtering models were simpler and quicker to apply, and the effectiveness of their performance suggested that these kind of models could be applied in the context of CENTER-TBI and the variability would be corrected.

Conclusions: As expected, the scanners introduced a significant level of variability in the data. The variability introduced by the hardware can be quantified within scanners, analysing the data from the same device, or across scanners and comparing the data from different devices. In both cases, the results suggest that the correction and elimination of variability introduced by the hardware are needed, before proceeding with further analysis using data from multicentre studies, in order to ensure the reliability of the results. The methods tested in this dissertation showed to be effective in the elimination of the variability introduced by the scanner.

Keywords: Multicentre studies, Medical Imaging Techniques, Variability across scanner, Correction of the bias introduced by the scanner, Traumatic Brain Injury.

ACKNOWLEDGEMENTS

Primary my gratitude goes to my family, who always support my dreams and always support the decision of to go abroad in order to follow my wish of learn more and more. Without them the difficulties of the work and the difficulty of be alone in the foreign country would not be surpass. The key for a good work is not only the effort and the devotion to the subject in study. This key for the success is, in fact, the faith that people have in us, their support and our love for the work we do. I learnt that not between the halls of the University but with my family, thus, without them I never would conclude my dissertation and my project.

I would like to thanks Dr Marta Correia to give the opportunity to work in her group. I appreciate all her contributions of time, ideas, and all the energy that she focuses on my work. With her it was possible learned so much about themes that, until now, I just known superficially. Without her sympathy and patient that cannot be possible.

I am also thankful to Prof. Dr. Rita Nunes for her availability to be my supervisor in this project, helping me in all of this process and this project. She always was so supportive and read any of the ideas, lines and pages of this dissertation and all of the works and reports developed on context of this dissertation. I am sure that any success of the work that I developed it was also her success, since even the best results without a convenient presentation and a good argumentation are not properly recognized.

I am very thankful to Prof. Dr. Eduardo Ducla-Soares that since the very begging, in the first class of a young student of 18 years old, believe in her passion about the biomedical field. He open her eyes to the research field and the importance of work not for the grade but for the continuous growing as a scientist and as a person. If that young girl are now the young researcher as I am today is due to the inspiration that he was in her course.

I also would like to express my thanks to Dr Virginia Newcombe and Dr Guy Williams, who were really attentive and spend their time to discuss results, present suggestions and guide me in this project. I would like to thank the staff of Division of Anaesthetics of Addenbrookes Hospital, for providing me with the conditions to learn about TBI and MRI for all their sympathy during my stay. I am very thankful as well to Dr. David Menon for his kindness in receive me. Besides, I always will be thankful for the clever discussions in which he participated during my internship. This meetings contributed so much for my dissertation and work.

Finally, I would like to thank to my friends, especially to the ones that remained close despite of the distance, and made the distance just a concept and not a reality. Their love and encouragement were enough to keep me going. Without their help I would not have been able to start or even finished this project.

INDEX

Resumo viii
Abstract x
Acknowledgements xii
Index xiv
List of Figures xvii
List of Tables xxii
List of Abreviations xxiii
Introduction 25
1.1. Context 25
1.2. Objectives of the Dissertation 26
1.3. Outline of the Dissertation 27
Background 28
2.1. Introduction 28
2.2. Medical Imaging 29
2.2.1. Magnetic Resonance Imaging 29
2.2.2. Diffusion Tensor Imaging 33
Multicentre Studies: State of the Art 39
3.1. Introduction 39
3.2. Advantages and Limitations of Multicentre studies 39
3.3. Quantification of Intra- and Inter-scanner Variability 41
3.3.1. Magnetic Resonance Imaging – MPRAGE 41
3.3.2. Diffusion Tensor Imaging 42
3.3.3. Quantification methods for MPRAGE and DTI images 45
3.4. Approaches to Variability Correction 45
3.4.1. Structural Magnetic Resonance Imaging – MPRAGE 45
3.4.2. Diffusion Tensor Imaging 47
3.5. Summary 48
Quantification of variability Intra- and Inter-scanner 49
4.1. Introduction 49
4.1. Methodology 50
4.1.1. Group definition and Data pre-processing 50
4.1.1. Quantification of variability – Coefficient of Variation 51
4.2. Results 53

4.2.1. Variability quantification within centre 53
4.2.2. Quantification of inter-scanner variability 57
4.2. Discussion 60
4.2.1. Quantification of intra-scanner variability 60
4.2.3. Quantification of inter-scanner variability 61
4.3. Conclusion 61
Spatial Filtering Model 62
5.1. Introduction 62
5.2. Methodology 63
5.2.1. Group definition and Data pre-processing 63
5.2.2. Spatial Filtering Method 63
5.3. Results 65
5.3.1. Spatial Filtering Model 65
5.4. Discussion 77
5.4.1. Spatial Filtering Method 77
5.4.2. Limitations of the Method and Future Improvements 79
5.5. Conclusion 80
Regression Models 81
6.1. Introduction 81
6.2. Methodology 82
6.2.1. Group definition and Data pre-processing 82
6.2.2. Regression Model: Ordinary Least Squares (OLS) 82
6.2.3. Support Vector Machine (SVM) 84
6.3. Results 84
6.3.1. Regression Model: Ordinary Least Squares (OLS) 84
6.4. Discussion 85
6.4.1. Regression Model: Ordinary Least Squares (OLS) 85
6.4.2. Limitations and Future Improvements 86
6.5. Conclusion 87
Models Comparison 88
7.1. Introduction 88
7.2. Methodology 88
7.2.1. Group definition and Data pre-processing 88
7.2.2. Support Vector Machine (SVM) 88
7.3. Results 89
7.3.1. Comparison of the effectiveness of correction methods 89
7.4. Discussion 90
7.5. Conclusion 90

Conclusion and Future Work....91
References....94
Regions of Interest....xcviii
A.1. White Matter Regions....xcviii
A.2. Grey Matter Regions....xcix
Quantification of Variability....c
B.1. Voxel Based Analysis....c
B.2. ROI Analysis....cii
Spatial Filtering Model....cv
DTI data pre-processing....cix
D.1. Pre-processing and extraction of DTI measures....cix
D.2.1. Spatial filtering of DWI volumes....cx
D.2.2. Eddy current and motion corrections....cx
D.2.3. Extraction of DTI measures....cxi
D.2.4. Registration....cxi
D.2. Quantification of Variability intra- and inter-scanner....cxii
GUI for Multicentre study analysis....cxiv

LIST OF FIGURES

Figure 2. 1 – External magnetic field effect on protons. In the left side of the image is possible to see the protons in absence of magnetic field. In consequence, the spin's magnetic moment of the protons are not aligned (represented by the black arrows). On other hand, when a magnetic field is applied (right side) the spin's magnetic moment of the protons aligns in direction of B_0. The orientation of the spins is explained by the Boltzmann statistics, which explains the different orientation of the spin's magnetic momentum of protons represented at this image. Image adapted from (Haacke, E. M; Brown, R.W; Thompson, M. R; Venkatesan, 1999).....29

Figure 2.2 – Proton relaxation and longitudinal magnetization recovery, after a 90° RF pulse is applied at equilibrium. The z component of the net magnetisation, Mz, is reduced to zero. This component recovers gradually back to its equilibrium value if no further RF pulses are applied. The recovery of Mz is an exponential process with a time constant T1. This is the time at which the magnetization has recovered to 63% of its value at equilibrium. Image adapted from (Ridgway, 2010). 30

Figure 2.3 –Transversal magnetization relaxation, after a 90° RF pulse is applied at equilibrium. Initially the transverse magnetisation (red arrow) has a maximum amplitude as the population of proton's spin magnetic moments rotate in phase. The amplitude of the net transverse magnetisation decays as the proton's spin magnetic moments move out of phase with one another (shown by the small black arrows). The resultant decaying signal is known as the Free Induction Decay (FID). The overall term for the observed loss of phase coherence (de-phasing) is T2* relaxation, which combines the effect of T2 relaxation and additional de-phasing caused by local variations (inhomogeneities) in the applied magnetic field. T2 relaxation is the result of spin-spin interactions and due to the random nature of molecular motion, this process is irreversible. T2* relaxation accounts for the more rapid decay of the FID signal, however the additional decay caused by field inhomogeneities can be reversed by the application of a 180° refocusing pulse. Both T2 and T2* are exponential processes with times constants T2 and T2* respectively. This is the time at which the magnetization has decayed to 37% of its initial value immediately after the 90° RF pulse. Image adapted from (Ridgway, 2010). 31

Figure 2.4 – Schematization of Spin-Echo sequence. The presence of magnetic field inhomogeneities causes additional de-phasing of the proton magnetic moments. The Larmor frequency is slower where the magnetic field is reduced and faster where the field is increased resulting in a loss or gain in relative phase respectively. After a period of half the echo time, TE/2, the application of a 180° RF pulse causes an instantaneous change in sign of the phase shifts by rotating the spins (in this example) about the y axis. As the differences in Larmor frequency remain unchanged, the proton magnetic moments the move back into phase over a similar time period, reversing the de-phasing effect of the magnetic field inhomogeneities to generate a spin echo. In addition to the effect of the 180° refocusing pulse, gradients are applied to de-phase and re-phase the signal for imaging purposes. Note that for spin echo pulse sequences, the second gradient has the same sign as the first, as the 180° pulse also changes the sign of the phase shifts caused by the first gradient. Image adapted from (Ridgway, 2010). 32

Figure 2.5 – Schematization of Gradient-Echo sequence. The application of the 1st positive magnetic field gradient causes rapid de-phasing of the transverse magnetisation, Mxy. Therefore the FID signal to zero amplitude. The application of the 2nd negative magnetic field gradient reverses the de-phasing caused by the first gradient pulse, resulting in recovery

of the FID signal to generate a gradient echo at the echo time, TE. Extension of the time duration of the second gradient to twice that of the first gradient causes the FID to then de-phase to zero. The maximum amplitude of the echo depends on both the T2* relaxation rate and the chosen TE. Image adapted from (Ridgway, 2010). 33

Figure 2.6 – Schematization of biological diffusion. Free water diffusion in extracellular mediums (left) and water diffusion in fibrous tissues (right). In first medium (left) in absence of boundaries the movement of the molecules is not constrained; in fact, at the molecular level the diffusion of water is random. On other hand, the boundaries (right) constrain the movement of the molecules and its movement is orientated in the direction of the fibres. Image adapted from (Winston, 2012). 35

Figure 2.7 – Representation of PGSE pulse sequence. G corresponds to the amplitude of the gradient applied during the time δ. The Δ represents the time between pulses and the TE corresponds to the total time – Echo time. Image adapted from (Winston, 2012). 36

Figure 3. 1 – Schematization of the model applied by Moorhead et al., to correct the variability between scanners. Image adapted from (Moorhead et al., 2009). 47

Figure 4.1 – Coefficient of Variation expressed as a percentage value for each type of image analysed. Each bar represents the mean COV value for each centre. The line corresponds to the standard deviation of the COV value. 53

Figure 4.2 – Coefficient of Variation expressed as a percentage value for FA images. Each bar represents the mean COV value for each ROI in the three centres. 54

Figure 4.4 – Coefficient of Variation expressed as a percentage value for MD images. 54

Figure 4.4 – Coefficient of Variation expressed as a percentage value for GM images. 55

Figure 4.5 – Coefficient of Variation expressed as a percentage value for MD images. 55

Figure 4.6 – Summary of the analysis performed for the 48 ROI defined. The red stars correspond to the values defined statistically as outliers. 56

Figure 4.7 – Coefficient of Variation expressed as a percentage value for each type of image analysed. Each bar represents the mean COV value. The line corresponds to the standard error of the COV value. 57

Figure 4.8 – Coefficient of Variation expressed as a percentage value for FA images. Each bar represents the mean COV value for each ROI across centres. 58

Figure 4.9 – Coefficient of Variation expressed as a percentage value for MD images for each ROI across centres. ..58

Figure 4.10 – Coefficient of Variation expressed as a percentage value for GM images across centres. 58

Figure 4.11 – Coefficient of Variation expressed as a percentage value for WM images for each ROI across centres. 59

Figure 4.12 – Summary of the analysis performed for the 48 ROI defined. The red starts correspond to the values defined statistically as outliers. 59

Figure 5.1 –FA maps. Z score of the Controls from Centre 3 – WBIC – when expressed in terms of controls from Centre 2 - Finland. The shadow region corresponds to the region of non-significant differences among subjects. Each circle represents a different subject. 65

Figure 5.2 – FA maps. Z score of the Patients when expressed in terms of a baseline defined by the controls from the two centres under study. 66

Figure 5.3 – MD maps. Z score of the Controls from Centre 3 – WBIC – when expressed in terms of controls from Centre 2 - Finland. 66
Figure 5.4 – MD maps. Z score of the Patients for the Mean Diffusivity data. 66
Figure 5.5 – GM maps. Z score of the Controls from Centre 3 – WBIC – when expressed in terms of controls from Centre 2 – Finland. 67
Figure 5.6 – GM maps. Z score of the Patients when expressed in terms of a baseline defined by the controls from the two centres under study. 67
Figure 5.7 – WM maps. Z score of the Controls from Centre 3 – WBIC – when expressed in terms of controls from Centre 2 - Finland. 67
Figure 5.8 – WM maps. Z score of the Patients when expressed in terms of a baseline defined by the controls from the two centres under study. 68
Figure 5.9 – FA maps. Z score of the Controls from Centre 3 – WBIC – when expressed in terms of controls from Centre 2 - Finland. 68
Figure 5.10 – FA maps. Z score of the Patients when expressed in terms of a baseline defined by the controls from the two centres under study, after correction. 68
Figure 5.11 – MD maps. Z score of the Controls from Centre 3 – WBIC – when expressed in terms of controls from Centre 2 - Finland. 69
Figure 5.12 – MD maps. Z score of the Patients when expressed in terms of a baseline defined by the controls from the two centres under study. 69
Figure 5.13 – GM maps. Z score of the Controls from Centre 3 – WBIC – when expressed in terms of controls from Centre 2 - Finland. The correction was performed using a FWHM=2mm. 69
Figure 5.14 – GM maps. Z score of the Patients when expressed in terms of a baseline defined by the controls from the two centres under study. The FWHM used was 2mm. 70
Figure 5.15 – GM maps. Z score of the Controls from Centre 3 – WBIC – when expressed in terms of controls from Centre 2 - Finland. The correction applied a FWHM=8mm. 70
Figure 5.16 – GM maps. Z score of the Patients when expressed in terms of a baseline defined by the controls from the two centres under study. The factor used in spatial filtering was FWHM= 8mm. 70
Figure 5.17 – WM maps. Z score of the Controls from Centre 3 – WBIC – when expressed in terms of controls from Centre 2 - Finland. The FWHM used was equal to 2 mm. 71
Figure 5.18 – WM maps. Z score of the Patients when expressed in terms of a baseline defined by the controls from the two centres under study. The correction was performed using a FWHM=2mm. 71
Figure 5.19 – WM maps. Z score of the Controls from Centre 3 – WBIC – when expressed in terms of controls from Centre 2 - Finland. The correction was performed using a FWHM=8mm. 71
Figure 5.20 – WM maps. Z score of the Patients when expressed in terms of a baseline defined by the controls from the two centres under study. The FWHM value used was 8mm. 72
Figure 5.21 – Number of controls significantly different for each ROI before and after correction for FA maps. 72
Figure 5.22 – Number of patients significantly different for each ROI before and after correction for FA maps. 73
Figure 5.23 – Number of subjects significantly different for each ROI before and after correction for MD maps. 73

Figure 5.24 – Number of patients significantly different for each ROI before and after correction for MD maps. 73

Figure 5.25 – Number of controls significantly different for each ROI before and after correction, Considering to the two types of correction applied, for GM maps. 74

Figure 5.26 – Number of patients significantly different for each ROI before and after correction, considering the correction A and B applied, for GM maps. 74

Figure 5.27 – Number of controls significantly different for each ROI before and after correction, considering the correction A and B applied, for WM maps. 74

Figure 5.28 – Number of patients significantly different for each ROI before and after correction, considering the correction A and B applied, for WM maps. 75

Figure 5.29 – Evaluation of differences for FA maps displaying voxels with significant differences between centres. The top row corresponds to the FA slices before any correction, and the bottom row corresponds to the FA slices after correction. 75

Figure 5.30 – Evaluation of differences for MD maps. Voxels with significant differences between centres. The top row corresponds to the MD slices before any correction, and the bottom row corresponds to the MD slices after correction. 76

Figure 5.31 – Evaluation of differences for GM maps. Voxels with significant differences between centres. The top row corresponds to the GM slices before any correction, and the bottom row correspond to the GM slices after correction A. 76

Figure 8.1 – Schematization of the mixed model to correct the variability. 93

Figure B.1 – Coefficient of Variation for FA maps. It is shown three different views (coronal, axial and sagittal) of the brain, in which a median value of COV for each voxel, obtained for each Centre, is represented using a colour code: red-yellow. c

Figure B.2 – Coefficient of Variation for MD maps. It is shown three different views (coronal, axial and sagittal) of the brain, in which a median value of COV for each voxel, obtained for each Centre, is represented using a colour code: red-yellow. ci

Figure B.3 – Coefficient of Variation for GM maps. It is shown three different views (coronal, axial and sagittal) of the brain, in which a median value of COV for each voxel, obtained for each Centre, is represented using a colour code: red-yellow. ci

Figure B.4 – Coefficient of Variation for WM maps. It is shown three different views (coronal, axial and sagittal) of the brain, in which a median value of COV for each voxel, obtained for each Centre, is represented using a colour code: red-yellow. cii

Figure B.5 – Coefficient of Variation for FA maps. It is shown three different views (coronal, axial and sagittal) of the brain, in which a median value of COV for each ROI, obtained for each Centre, is represented using a colour code: red-yellow. cii

Figure B.6 – Coefficient of Variation for MD maps. It is shown three different views (coronal, axial and sagittal) of the brain, in which a median value of COV for each ROI, obtained for each Centre, is represented using a colour code: red-yellow. ciii

Figure B.7 – Coefficient of Variation for GM maps. It is shown three different views (coronal, axial and sagittal) of the brain, in which a median value of COV for each ROI, obtained for each Centre, is represented using a colour code: red-yellow. ciii

Figure B.8 – Coefficient of Variation for FA maps. It is shown three different views (coronal, axial and sagittal) of the brain, in which a median value of COV for each ROI, obtained for each Centre, is represented using a colour code: red-yellow. civ

Figure C.1 – Number of Patients significantly different when compared with the baseline, for each ROI, before and after correction. cv

Figure C.2 – Number of Patients significantly different when compared with the baseline from centre 3, before and after correction. cvi

Figure C.3 – Number of Patients significantly different when compared with the baseline from centre 2, before and after correction. cvi

Figure C.4 – Number of Patients significantly different when compared with the baseline from centre 3, before and after correction. cvi

Figure C.5 – Number of Patients significantly different when compared with the baseline from centre 2, before and after the two types of corrections were applied. cvii

Figure C.6 – Number of Patients significantly different when compared with the baseline from centre 3, before and after the two types of corrections were applied. cvii

Figure C.7 – Number of Patients significantly different when compared with the baseline from centre 2, before and after the two types of corrections were applied. cviii

Figure C.8 – Number of Patients significantly different when compared with the baseline from centre 3, before and after the two types of corrections were applied. cviii

Figure D.1 – In both panels a b0 volume is shown in the axial plane, visualized with a window of intensities [3; 1500]. Panel A shows the b0 volume from one of the subjects without spatial filtering. Panel B shows the result of spatial filtering using smoothing by a Gaussian kernel with FWHM=2mm. cx

Figure D.2 – Summary of the different methodologies implemented to process the DTI data. cxii

LIST OF TABLES

Table 4.1 – Summary of the Coefficient of Variation obtained in within centre. 56
Table 4.2 - Summary of the Coefficient of Variation obtained across centres. 60
Table 5.1 – Summary of statistical analysis.. 77
Table 6.1 – SVM results of the data classification of the results of model correction, in terms of the Centre. 85
Table 7.1 – SVM results of the data classification of the results of models of correction, in terms of the Centre. 89

List of Abreviations

ANCOVA	Analysis of Covariance
CBU	Cognition and Brain Sciences Unit
COV	Coefficient of Variation
CT	Computerized Tomography
DTI	Diffusion Tensor Imaging
DWI	Diffusion Weight Imaging
EPI	Echo Planar Imaging
FA	Fractional Anisotropy
FID	Free Induction Decay
FLIRT	FMRIB's Linear Image Registration Tool
fMRI	Functional Magnetic Resonance Imaging
FNIRT	FMRIB's Non Linear Image Registration Tool
FOV	Field of View
FSL	FMRIB Software Library
FWHM	Full Width at Half Maximum
GCS	Glasgow Coma Scale
GE	Gradient Echo Sequence
GLM	General Linear Model
GM	Grey Matter
GPR	Gaussian Process Regression
ICC	Interclass Correlation Coefficient
JHU	Johns Hopkins University
LLS	Linear Least Squares
LOC	Loss of Conscientious
MD	Mean Diffusion
MPRAGE	Magnetization Prepared Rapid Gradient Echo Sequence
MRI	Magnetic Resonance Imaging
MT	Magnetization Transfer
NLLS	Non-Linear Least Squares
PD	Proton Diffusion
PGSE	Pulsed-Gradient Spin-Echo Sequence
RF	Radio Frequency

ROI	Region of Interest
SE	Spin Echo Sequence
SFNR	Signal-to-Fluctuation-Noise-Ratio
sMRI	Structural Magnetic Resonance Imaging
SNR	Signal-to-Noise-Ratio
SVM	Support Vector Machine
TBI	Traumatic Brain Injury
TBSS	*Tract-Based Spatial Statistics*
VBM	Voxel-based morphometry
WBIC	Wolfson Brain Imaging Centre
WM	White Matter

Chapter 1

INTRODUCTION

1.1. CONTEXT

The World Health Organization expects that Traumatic Brain Injury will surpass many diseases as the major cause of death and disability by the year 2020 (Hyder, Wunderlich, Puvanachandra, & Gururaj, 2007). In Europe, TBI affects about 7.7 million individuals, and it is estimated that 30-70% of them will suffer on-going mental illness as well physical and cognitive disabilities (Kostro et al., 2014; Wilson et al., 2014)

Traumatic Brain Injury (TBI) is defined as an alteration in brain function, or other evidence of brain pathology, caused by an external factor (Menon, Schwab, Wright, & Maas, 2010). TBI can be classified as mild, moderate or severe based on acute TBI variables that include the following clinical signs:

- Duration of loss of consciousness (LOC);
- Glasgow Coma Score (GCS);
- Loss of memory for events immediately before or after the injury (Post Traumatic Amnesia - PTA);
- Neurologic deficits, such as weakness, loss of balance, dysphasia paresis/plegia, sensory loss, aphasia, etc.

The most common manifestation of this pathology is mild TBI, with an incidence of 70-90% of all cases. (Kraus et al., 2007)

Certain brain regions are more susceptible to contusion following TBI, such as frontal and anterior temporal cortices, due to the proximity to the skull. On the other hand, white matter tracts are particularly susceptible to the shearing forces that occur with TBI. (Risdall & Menon, 2011)

The diagnosis of TBI is made when the symptoms and signs are closely related to the insult. However, other clinical manifestations may be delayed. These manifestations include cognitive changes such as decreased mental flexibility, impaired attention, poor planning, impaired judgement, deficits in verbal fluency, problems with working memory as well as increased impulsivity (Kraus et al., 2007). Determining the extent of clinically relevant neuropathology associated with TBI might be problematic. To solve this issue, modern medical imaging techniques have been used, such as Diffusion Tensor Imaging (DTI). These MRI techniques provide pathological information *in vivo*. Besides, the prospect of tracking white matter microstructural changes over time holds the promise of measuring neuroplasticity and repair following TBI, which can provide a way of monitoring the therapeutic response and can improve the efficiency of the treatment.

However, regarding the inherent heterogeneity of TBI, which is determined by multiple factors[1], the study of this pathology is a huge challenge, which requires a diversified dataset. A large dataset ensures the presence of a large spectrum of different symptoms and of therapeutic responses, allowing a more complete understanding of this illness. Thus, the multicentre studies have proven themselves very useful to collect data from subjects with interesting characteristics for TBI research, contributing to increase the statistical power of the TBI studies and to improve the reliability of the assumptions made about this illness.

Considering this fact, a pragmatic approach to TBI research is required. The CENTER-TBI Project is a large European study that aims to improve the care for patients with TBI (Figure 1.1.). It consists in a prospective longitudinal observational study in 60 centres from 20 countries including approximately 5400 patients, and it forms part of the larger global initiative InTBIR: International Initiative for Traumatic Brain Injury Research with projects currently ongoing in Europe, the US and Canada.

The main goals of the CENTER-TBI are the improvement of the characterization of TBI as a disease and the identification of the most effective clinical interventions for managing TBI. For that, the CENTER-TBI study will include works in several areas of research, such as neuroinformatics, genetic associations, MR imaging, biomarkers and comparative effectiveness research (CER), which will give a complete and detailed characterization of TBI (CENTER-TBI, 2014).

The CENTER-TBI purposes to connect the real clinical situation to the approaches studied, in order to translate the research outputs into valid and useful information for physicians and clinicians, allowing a better care of the patients, as well as a more efficient approach for the hospitals and health care centres. Therefore, it will develop and sustain an international TBI knowledge community that integrates results of the project with high quality 'living evidence reviews' of the current state of knowledge.

These goals will only be possible if the reliability of the study is ensured. It is in this topic that the project that foments this thesis is inserted. The main goals of the thesis will be discuss in the next section.

1.2. Objectives of the Dissertation

The principal aim of this thesis is to develop robust methods for quantification and correction of variability across scanners in multicentre studies. Secondly, this dissertation is inserted in the CENTER-TBI Project and, consequently, it aims to show the potential of multicentre studies for the improvement of diagnosis of TBI, and as an advanced method in brain study.

Also, it is expected that this work can be part of the set of published studies in the context of CENTER-TBI, specifically as a study of the impact of the variation introduced by the scanner in multicentre studies and contributing to the definition of which methodologies could be useful to overcome the main issues behind these studies.

[1] The factors which contribute to the heterogeneity of TBI are injury location, physiology, extracranial injuries, and constitutional effects of the patient (CENTER-TBI, 2014; Hyder et al., 2007; Wilson et al., 2014).

1.3. Outline of the Dissertation

This report is organized in 8 chapters as described below.

In the present chapter, Chapter 1, a brief contextualization of the project, in which this dissertation is inserted, is presented. In this chapter, the goals of the project are detailed.

The second chapter introduces the concepts behind medical imaging useful for TBI analysis, namely MRI. Considering the topics presented, the chapter is divided into two sections. In the first one, the concepts on MRI, such as the physical principles and imaging reconstruction principles of MRI, are described.

Chapter 3 introduces the state of the art of multicentre studies. The limitations and advantages of multicentre studies, as well as their usefulness on TBI study, are explained. In this chapter, the methodologies applied in previous studies to quantify the variability introduced by the hardware in medical imaging are also detailed, and the impact this variability on the studies reliability and data reproducibility is also discussed. A set of strategies for data calibration and error correction presented in the literature are described in the last section of this chapter.

In Chapter 4 the dataset used in study developed on context of this dissertation and its main characteristics are detailed. The methods used to quantify the variability between scanners and intra-scanner are also described in this chapter. In addition, the variation measured intra- and inter-centre is presented as well as a comment about the values achieved for the three centres, independently, and for the inter-centre analysis.

The following chapters, Chapters 5 and 6, a comprehensive and extensive description of the methods used for the elimination of the variability are presented as well as the main results obtained using the models employed. Thereby, in Chapter 5 an evaluation of the Spatial Filtering Model is presented, in which an ROI analysis and a voxel based analysis are compared and discussed. In this chapter, the effectiveness of a default pipeline to correct for the errors in Magnetization Prepared Rapid Gradient Echo (MPRAGE) images, namely in Grey Matter (GM) and White Matter (WM) segmentations is also discussed. In Chapter 6, the usefulness of using covariates to absorb the error introduced by nuisance variables in the data is discussed.

In Chapter 7 a comparison between models is presented.

Finally, Chapter 8 presents a conclusion and evaluation of the project. Moreover, in this chapter, an overview of the work developed as well as a reflection on future work in the context of multicentre studies and variability correction in MRI are presented.

Chapter 2

BACKGROUND

2.1. INTRODUCTION

In medical science when we refer to medical imaging, we are referring to a wide range of techniques and technologies which are used to analyse the human body for purposes of diagnosis, monitoring or treating pathological medical conditions.

Medical imaging includes a range of non-invasive techniques which have been proven to be an essential tool for current medical practice. These techniques are part of the current approach to diagnose and evaluate patients when they arrive at a medical facility or hospital. In fact, medical imaging techniques enable higher viability and accuracy of diagnosis, which result in greater effectiveness of medical treatment. The use of medical imaging has enabled physicians to study the symptoms and progress of a specific pathology, without the need for exploratory surgery. In modern medicine, there are several medical imaging methods with different features, which register different types of information that may be more targeted to functional or anatomical details, depending on the specific symptoms or pathology. For instance, techniques like structural Magnetic Resonance Imaging (sMRI), Computerised Tomography (CT) or Diffusion Tensor Imaging (DTI) are more directed to anatomical studies, whereas functional Magnetic Resonance Imaging (fMRI) is useful in functional evaluations.

Nowadays, medical imaging techniques are an important procedure in the employed methodology to study the patho-physicological mechanisms of several diseases, and at the same time ensuring the comfort and safety of the individuals being analysed. Thereby, taking all characteristics into account it is obvious why these techniques have achieved such an important role in fields such as Clinical Neuroscience. In particular, medical imaging methods like standard MRI structural images can readily demonstrate large focal contusions or bleeds, diffuse axonal injury may be detected indirectly by brain volume loss (volumetric analysis) (Boven et al., 2011). On the other hand, studies using functional MRI (fMRI) in patients with TBI show abnormal patterns of brain activation in patients compared with healthy control subjects (Boven et al., 2011). DTI also demonstrated potential in TBI diagnosis, since DTI allows for the specific examination of the integrity of white matter tracts, which are especially vulnerable to the mechanical trauma of TBI (Kraus et al., 2007; Shenton et al., 2012). Thereby, there has been increasing interest in using MRI techniques to diagnose and monitor TBI patients (Risdall & Menon, 2011; Shenton et al., 2012).

In this chapter the basic principles of MRI and DTI are presented, and the advantages and limitations of these techniques for the purpose of medical diagnostics are discussed.

2.2. Medical Imaging

2.2.1. Magnetic Resonance Imaging

Magnetic Resonance Imaging (MRI) is a non-invasive technique which allows *in-vivo* imaging of the human body, without requiring the use of radiation. MRI is a high resolution technique and it is frequently used to acquire brain images, since it allows acquisition of images with high level of anatomical detail and, consequently, makes possible the detection and monitoring of many lesions.

(a) Physical principles behind MRI

MRI uses magnetic fields and electromagnetic energy to generate signals from atomic nuclei, which can be used to create an image. The human body is composed of 60% of water, whereby the hydrogen nuclei the third most abundant in the human body. MRI uses the signal from hydrogen nuclei to generate the image; however, it is possible to image other nuclei such as sodium.

When a volume of water, in this case a body area, is exposed to an external and static magnetic field, B_0, the spin's magnetic moment ($\vec{\mu}$) of the protons aligns in the direction of B_0 (Figure 2.1).

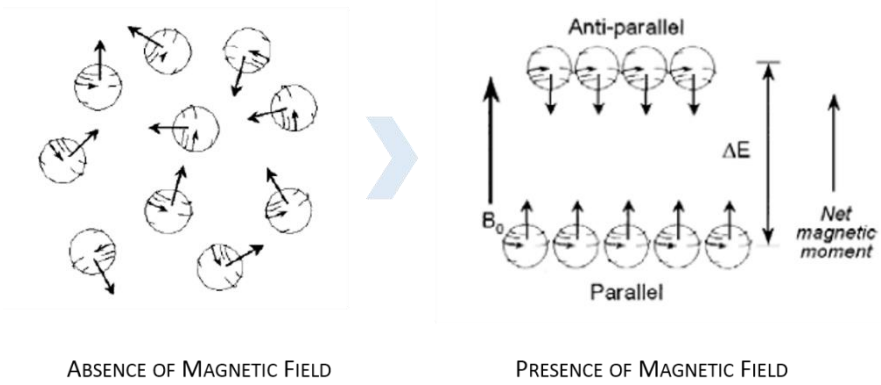

Figure 2. 1 – External magnetic field effect on protons. In the left side of the image is possible to see the protons in absence of magnetic field. In consequence, the spin's magnetic moment of the protons are not aligned (represented by the black arrows). On other hand, when a magnetic field is applied (right side) the spin's magnetic moment of the protons aligns in direction of B_0. The orientation of the spins is explained by the Boltzmann statistics, which explains the different orientation of the spin's magnetic momentum of protons represented at this image. Image adapted from (Haacke, E. M; Brown, R.W; Thompson, M. R; Venkatesan, 1999).

This phenomenon is explained by Boltzmann statistics, which state that there are more spins in the lower energy level (orientated parallel to B_0), than spins in the higher energy level (orientated anti-parallel to B_0). In this situation, the spin's magnetic moment starts precessing at a specific frequency, known as the Larmor frequency – ω_0:

$$\omega_0 = \gamma B_0 \qquad \text{Equation 2.1.}$$

where γ is the gyromagnetic ratio.

The sum of magnetic moments of all the protons exposed to the external magnetic field is given by the Net Magnetic Moment ($\vec{M}$), which is aligned with B_0. This vector has a component in the B_0 direction, designated as longitudinal magnetization (M_Z), which is the larger component at equilibrium state, since the transversal component of B_0, designated

as transversal magnetization (M_{xy}) is close to zero at this state. Note that, in this situation, the protons are not precessing in phase and the $\vec{\mu}$ component transversal to B_0 is randomly distributed.

By applying a radio frequency pulse, with the Larmor frequency, it is possible to force the protons to precess in phase, which will increase the M_{xy}. The angle of rotation relative to the main magnetic field direction is called flip angle. When the radio frequency pulse is interrupted the $\vec{M}$ starts losing its transverse component, and returns to the equilibrium state. This effect of relaxation can be induced by two different processes: the spin-lattice relaxation (T_1 relaxation) or a spin-spin interaction (T_2 relaxation). In the first case (Figure 2.2), the protons transfer their energy to the surrounding macro-molecules and realign with B_0. T_1 is the time constant that represents the time required for the protons to recover 63% of initial M_z component, after a 90 degrees pulse has been applied. This phenomenon can be described by the following expression:

$$M_z(t) = M_0\left[1 - \exp(\frac{-t}{T_1})\right] \quad \text{Equation 2.2.}$$

where M_0 refers to the magnetization in the equilibrium state.

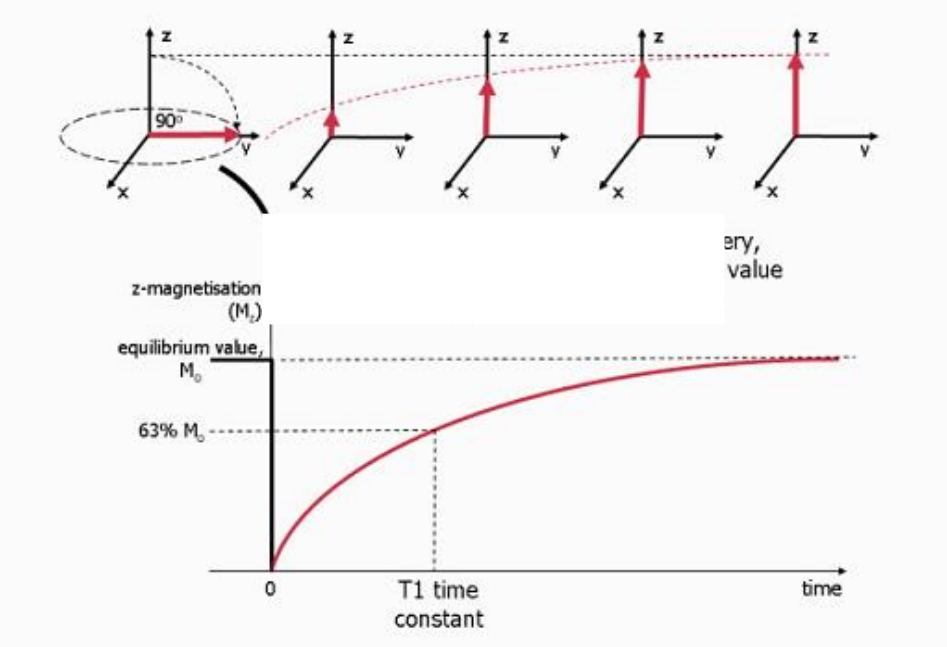

Figure 2.2 – Proton relaxation and longitudinal magnetization recovery, after a 90° RF pulse is applied at equilibrium. The z component of the net magnetisation, Mz, is reduced to zero. This component recovers gradually back to its equilibrium value if no further RF pulses are applied. The recovery of Mz is an exponential process with a time constant T1. This is the time at which the magnetization has recovered to 63% of its value at equilibrium. Image adapted from (Ridgway, 2010).

In the second case, (Figure 2.3), the reduction of M_{xy} is a consequence of loss of phase coherence between the protons, which can be expressed by Equation 2.3, where M_{xy0} refers to the initial transversal magnetization and T_2 corresponds to the time constant of spin-spin relaxation.

$$M_{xy0}(t) = M_{xy0}\left[\exp(\frac{-t}{T_2})\right] \quad \text{Equation 2.3}$$

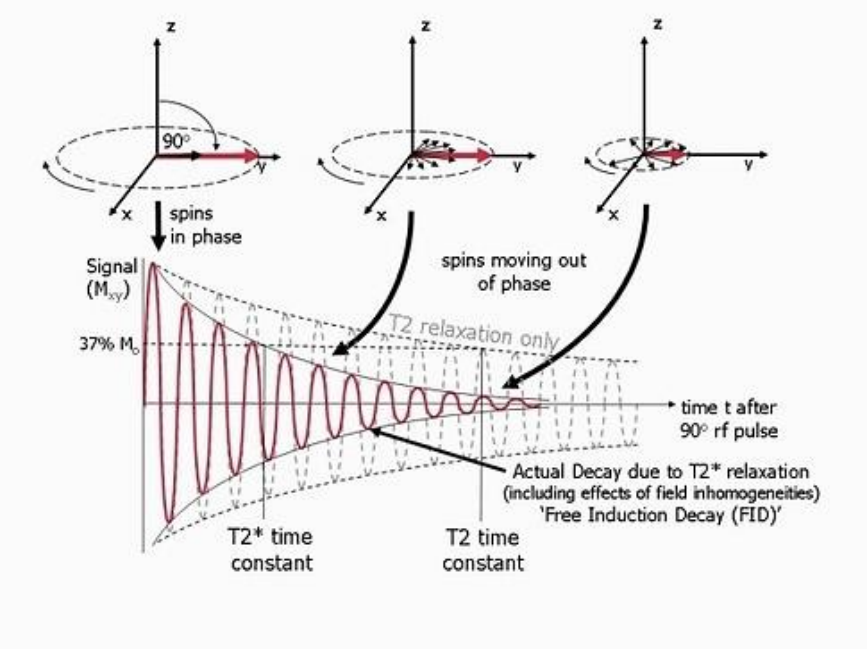

Figure 2.3 –Transverse magnetization relaxation, after a 90° RF pulse is applied at equilibrium. Initially the transverse magnetisation (red arrow) has a maximum amplitude as the population of proton's spin magnetic moments rotate in phase. The amplitude of the net transverse magnetisation decays as the proton's spin magnetic moments move out of phase with one another (shown by the small black arrows). The resultant decaying signal is known as the Free Induction Decay (FID). The overall term for the observed loss of phase coherence (de-phasing) is T2* relaxation, which combines the effect of T2 relaxation and additional de-phasing caused by local variations (inhomogeneities) in the applied magnetic field. T2 relaxation is the result of spin-spin interactions and due to the random nature of molecular motion, this process is irreversible. T2* relaxation accounts for the more rapid decay of the FID signal, however the additional decay caused by field inhomogeneities can be reversed by the application of a 180° refocusing pulse. Both T2 and T2* are exponential processes with times constants T2 and T2* respectively. This is the time at which the magnetization has decayed to 37% of its initial value immediately after the 90° RF pulse. Image adapted from (Ridgway, 2010).

(b) Image Principles and Imaging sequences

During both relaxation processes, the protons emit a radio frequency with the Larmor frequency, which is designated as free induction decay (FID). Considering that the FID does not provide a spatial discrimination, for image generation it is necessary to apply a position encoding process.

When a radio frequency pulse is applied with a specific bandwidth, only the protons with Larmor frequency coinciding with the RF pulse frequency profile are excited, which allows for a single slice to be selected. However, this procedure does not differentiate the protons within the same slice. For that, it is necessary to apply two more gradients in the orthogonal plane to the slices selected. By applying these gradients, and measuring the echoes produced, it is possible to fill the k-space[2], with information that codifies the spatial positions in the image according to their frequency. The spatial information can be obtained applying a Fourier transform to the k-space.

[2] The *k*-space refers to a matrix composed by complex values, which are sampled during an MR acquisition, in a premeditated scheme controlled by a pulse sequence, i.e. an accurately timed sequence of radiofrequency and gradient pulses. Therefore, the k-space can be described as a temporary image space, in which data from digitized MR signals are stored during data acquisition. In fact, the k-space holds raw data before reconstruction. The *k*-space corresponds to the spatial frequency domain, and can be described using the frequency encoding component (k_{FE}) and the phase encoding component (k_{PE}) components, which can be expressed using the following expressions: $k_{FE=}\gamma G_{FE} m\Delta t$ and $k_{PE=}\gamma n G_{PE}\tau$. The Δt corresponds to the interval of time during the sample is acquired, τ is the duration of GPE, γ is the gyromagnetic ratio, m is the sample number in the FE direction and n is the sample number in the PE direction

MRI allows the acquisition of images with different contrast by changing the RF pulse and gradient design throughout time. Each combination of RF and gradient steps is called an MR sequence. The Spin-Echo sequence is one of the most often used to acquire MRI images (Figure 2.4). The gradient-echo sequence is another kind of sequence commonly used (Figure 2.5), which is characterized by shorter acquisition times compared to Spin-Echo. In this project, images acquired using a magnetization prepared rapid gradient echo sequence (MPRAGE) were analysed. In this case, the sequence starts with magnetization preparation to introduce a T_1 contrast. This is achieved using an RF pulse with a flip angle of 180 degrees, also denominated as inversion pulse. M_z is described now by the Equation 2.4, instead of the Equation 2.2.

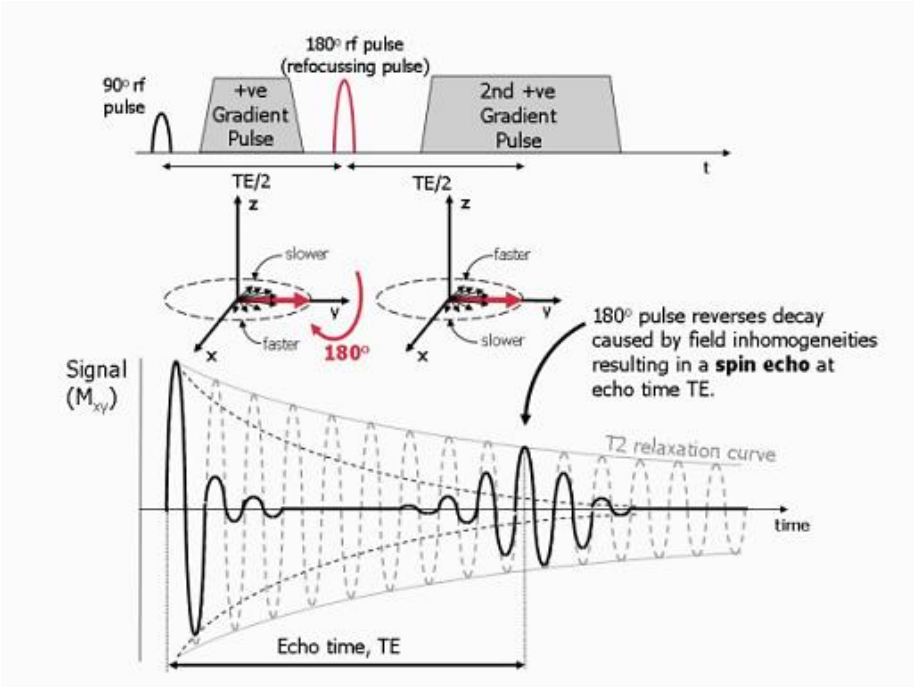

Figure 2.4 – Schematization of Spin-Echo sequence. The presence of magnetic field inhomogeneities causes additional de-phasing of the proton magnetic moments. The Larmor frequency is slower where the magnetic field is reduced and faster where the field is increased resulting in a loss or gain in relative phase respectively. After a period of half the echo time, TE/2, the application of a 180° RF pulse causes an instantaneous change in sign of the phase shifts by rotating the spins (in this example) about the y axis. As the differences in Larmor frequency remain unchanged, the proton magnetic moments move back into phase over a similar time period, reversing the de-phasing effect of the magnetic field inhomogeneities to generate a spin echo. In addition to the effect of the 180° refocusing pulse, gradients are applied to de-phase and re-phase the signal for imaging purposes. Note that for spin echo pulse sequences, the second gradient has the same sign as the first, as the 180° pulse also changes the sign of the phase shifts caused by the first gradient. Image adapted from (Ridgway, 2010).

the 2D-Fourier Transform of this encoded signal results in a representation of the spin density distribution in two dimensions. Thus position (x,y) and spatial frequency (k_{FE}, k_{PE}) constitute a Fourier transform pair (Twieg, 1983)

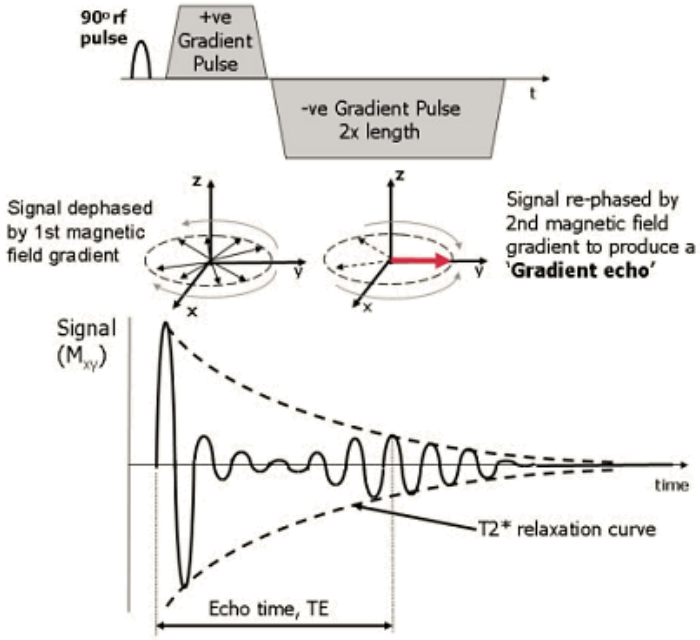

Figure 2.5 – Schematization of Gradient-Echo sequence. The application of the 1st positive magnetic field gradient causes rapid de-phasing of the transverse magnetisation, Mxy. Therefore the FID signal tends to zero amplitude. The application of the 2nd negative magnetic field gradient reverses the de-phasing caused by the first gradient pulse, resulting in recovery of the FID signal to generate a gradient echo at the echo time, TE. Extension of the time duration of the second gradient to twice that of the first gradient causes the FID to then de-phase to zero. The maximum amplitude of the echo depends on both the T2* relaxation rate and the chosen TE. Image adapted from (Ridgway, 2010).

The T_1 contrast will dominate the sequence, depending on the time between the inversion pulse and the gradient echo sequence. The period of time between both sequence periods is called inversion time TI. Then, when the Gradient Echo (GE) sequence is applied the signal is acquired. This acquisition period is followed by a magnetization recovery period where the magnetization recovers during a delay time (t) before the next inversion pulse, in order to prevent saturation effects.

$$M_z(t) = M_0\left[1 - 2\exp(\frac{-t}{T_1})\right]$$ Equation 2.4.

2.2.2. Diffusion Tensor Imaging

Diffusion Tensor Imaging (DTI) is a medical imaging technique that exploits the exquisite sensitivity of magnetic resonance imaging to diffusion processes to measure microscopic tissue orientation characteristics *in vivo* (Williams, Paul, Clark, & Gordon, 2007).

DTI characterizes the three-dimensional diffusion of water as a function of spatial localization. This technique is highly sensitive to the microstructural architecture of brain tissues, allowing the assessment of changes in the diffusion pattern in the presence of pathological symptoms (Alexander, Lee, Lazar, & Field, 2008).

The methods used to acquire and analyse DTI have evolved significantly in the last years. Among these are new pulse sequences and diffusion tensor encoding schemes that improve the accuracy and spatial resolution, by reducing the artefacts in tensor measurements.

(c) Physical principles of Diffusion

Diffusion is a kinetic process, in particular a mass process, which occurs in a fluid that leads to the homogenization, or uniform mixing, of the chemical components in a phase (Ceric, 2005).

The process of diffusion depends on many factors such as size and form of particles diffused, temperature and viscosity of the fluid. The Fick Law describes the flux, *J*, (Equation 2.5) of mass or energy in a medium where there isn't an initial thermic and chemical balance (Equation 2.6). Note that *D* denotes the diffusion coefficient and ∇C is the concentration gradient of the diffusing component.

$$J = -D\nabla C \quad \text{Equation 2.5.}$$

$$J_x = -D\frac{\partial C}{\partial x}\,;\, J_y = -D\frac{\partial C}{\partial y};\, J_z = -D\frac{\partial C}{\partial z} \quad \text{Equation 2.6.}$$

The movement of particles between two compartments due to diffusion processes results in a decrease of the concentration gradient, reaching a steady state, in which a uniform concentration across the medium is achieved. Considering the principle of mass conservation, described by Equation 2.7 where *r* represent a space vector (*x,y,z*) and *t* the temporal component, and using Fick's first law (Equation 2.5), it is possible to obtain Fick's second law, that describes the variation of concentration in a medium (Equation 2.8) (Ceric, 2005).

$$\nabla J(r,t) = -\frac{\partial C(r,t)}{\partial t} \quad \text{Equation 2.8.}$$

$$D\nabla^2 C(r,t) = \frac{\partial C(r,t)}{\partial t} \quad \text{Equation 2.9.}$$

On a molecular level, diffusion refers to a random displacement of the molecules agitated by thermal energy, known as Brownian motion (Hagmann et al., 2006).

Einstein described the diffusion coefficient, *D*, as a proportion of the mean squared-displacement divided by the number of dimensions, *n*, and the interval of time needed to the diffusion process, Δt (Equation 2.9). (Einstein, 1956)

$$D = \frac{\langle \Delta r^2 \rangle}{2n\Delta t} \quad \text{Equation 2.9.}$$

In the absence of boundaries, the diffusion of water, on a molecular level, is described by a Gaussian probability density, as shown in the following equation:

$$P(\Delta r, \Delta t) = \frac{1}{\sqrt{(4\pi D\Delta t)^3}} \times \exp\left(\frac{-\Delta r^2}{4D\Delta t}\right) \quad \text{Equation 2.10.}$$

(d) Biological Diffusion

The water diffusion in biological tissues occurs inside, outside and trough cellular structures, and it is modulated by the interactions with cellular membranes and organelles. The cellular membranes show a behaviour similar to boundaries dividing compartments. Due to this fact, the intracellular water tends to be more restricted when compared with extracellular water and, consequently, has a restricted diffusion. In the case of fibrous tissues, such as white matter, the water diffusion is highly restricted in the directions perpendicular to the fibres and, on the contrary, shows high diffusion in the directions parallel to the fibre orientation, as is schematized in Figure 2.6 (Alexander et al., 2008; Winston, 2012). Therefore, the diffusion in fibrous tissues can be considered to be anisotropic.

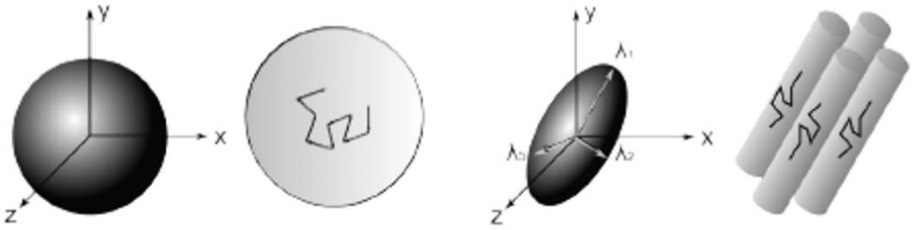

Figure 2.6 – Schematization of biological diffusion. Free water diffusion in extracellular mediums (left) and water diffusion in fibrous tissues (right). In first medium (left) in absence of boundaries the movement of the molecules is not constrained; in fact, at the molecular level the diffusion of water is random. On other hand, the boundaries (right) constrain the movement of the molecules and its movement is orientated in the direction of the fibres. Image adapted from (Winston, 2012).

Basser et al, (Basser, PJ; Matitello, J; LeBihan, 1994), described the anisotropic diffusion behaviour using the diffusion tensor (Equations 2.11), which is a 3x3 covariance matrix (Equation 2.12) that describes the covariance of diffusion displacements in three directions.

$$P(\Delta \boldsymbol{r}, \Delta t) = \frac{1}{\sqrt{(4\pi\Delta t)^3 |\boldsymbol{D}|}} \times \exp\left(\frac{-\Delta \boldsymbol{r}^{\boldsymbol{T}} \boldsymbol{D}^{-\mathbf{1}} \Delta \boldsymbol{r}}{4\Delta t}\right) \qquad \text{Equation 2.11.}$$

$$D = \begin{bmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{yx} & D_{yy} & D_{yz} \\ D_{zx} & D_{zy} & D_{zz} \end{bmatrix} \qquad \text{Equation 2.12.}$$

Note, that the off-diagonal elements are symmetric about the diagonal ($D_{ij}=D_{ji}$), thus $D_{xy} = D_{yx}$, $D_{xz}=D_{zx}$, *and* $D_{yz}=D_{zy}$ and only six non-collinear diffusion encoding directions are needed. The diagonal elements correspond to diffusion variances along the *x*, *y* and *z* directions.

The diagonalization of the diffusion tensor yields the eigenvalues (λ_1, λ_2, and λ_3) and the respective eigenvectors ($\hat{e}_1$, $\hat{e}_2$ and $\hat{e}_3$), which describe the directions and apparent diffusivities along the axes of principal diffusion. If the diffusion is isotropic, then $\lambda_1 = \lambda_2 = \lambda_3$, and the tensor can be visually represented by a sphere. On the other hand, if the diffusion is anisotropic then $\lambda_1 > \lambda_2 > \lambda_3$, and the tensor assumes an ellipsoid format (Alexander et al., 2008).

(e) Diffusion-Weighted Imaging

The MRI sequence most commonly used to quantify the diffusion in brain is a pulsed-gradient spin-echo (PGSE) with a single shot, echo planar imaging (EPI) readout (Figure 2.7). In this sequence, the first gradient pulse de-phases the magnetization across the sample, whilst the second pulse is responsible for re-phasing the magnetization. In the case of static molecules (molecules that are not diffusing), the phases induced by both gradient pulses will completely cancel, the magnetization will be maximally coherent and, consequently, there will be no signal attenuation due to diffusion. On the other hand, in the case of diffusion motion in the gradient direction, motion will cause the signal phase to change by different amounts for each pulse and for each spin; in other words, there will be a phase difference, which is proportional to the displacement, the area of diffusion gradient pulses defined by the amplitude, *G*, and the duration, δ, and the spacing between pulses, Δ. Therefore, the phase dispersion due to diffusion will cause signal attenuation.

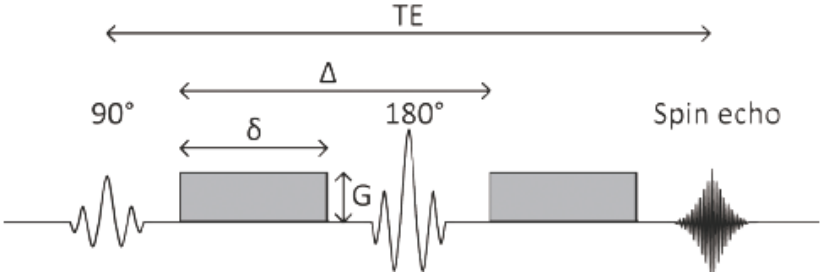

Figure 2.7 – Representation of PGSE pulse sequence. *G* corresponds to the amplitude of the gradient applied during the time δ. The Δ represents the time between pulses and the TE corresponds to the Echo time. Image adapted from (Winston, 2012).

Assuming the Gaussian displacement distribution mentioned before, the amplitude of the signal attenuation can be calculated with the following expression:

$$S = S_0 \exp\left\{-(\gamma G\delta)^2 \times (\Delta - \frac{\delta}{3}) \times D\right\}$$ Equation 2.13.

where S_0 is the signal without attenuation (situation where no diffusion gradient is applied), γ is the gyromagnetic ratio and *D* is the diffusion coefficient. This expression can be rewritten (Equation 2.14 and 2.15), and the *D* coefficient can be estimated using Equation 2.16, in the particular case in which the image only has one b-value.

$$b = (\gamma G\delta)^2 \times (\Delta - \frac{\delta}{3})$$ Equation 2.14.

$$S = S_0 \exp\{-b \times D\}$$ Equation 2.15.

$$D = -\frac{1}{b}\ln(\frac{S}{S_0})$$ Equation 2.16.

The Diffusion-weighted image (DWI) is extremely sensitive to subject motion, and even small motions can lead to phase and amplitude modulations in the acquired data and significant ghosting artefacts in the reconstructed images. Considering this fact, an EPI sequence could be useful to acquire images with a low presence of artefacts, since the fast acquisition speed of EPI turns this sequence highly efficient and maximizes the image signal-to-noise ratio. This fact reduces

the time that subjects need to be immobilized, reducing the probability of bulk motion. Thus, the EPI sequence improves the accuracy of the diffusion measurements, thereby this sequence is the most commonly used sequence to acquire DTI images. However, the EPI sequence shows some limitations, such as the image distortion caused by magnetic field inhomogeneities (Alexander et al., 2008).

(f) Diffusion Tensor Imaging

As detailed above, the anisotropic diffusion can be fully described using a diffusion tensor. Diffusion Tensor Imaging (DTI) uses the raw DWI images to estimate the diffusion tensor for all voxels in the brain (Alexander et al., 2008; Basser, PJ; Matitello, J; LeBihan, 1994).

Considering Equation 2.15, which describes the anisotropic diffusion in a medium, it is possible to define a matrix **B** which only depends on the parameters of the diffusion gradients (Equation 2.17), and where each entry of matrix **B**, b_{ij}, is estimated using Equation 2.14, taking into account the G value for position ij.

$$\boldsymbol{B} = \begin{bmatrix} b_{xx} & b_{xy} & b_{xz} \\ b_{yx} & b_{yy} & b_{yz} \\ b_{zx} & b_{zy} & b_{zz} \end{bmatrix} \quad \text{Equation 2.17.}$$

Thus, the Equation 2.15 can be rewritten as:

$$S = S_0 \exp\{-B.D\} \quad \text{Equation 2.18.}$$

A minimum of six non-collinear diffusion encoding directions are required to estimate the diffusion tensor, since the elements of the matrix **D** are symmetric about the diagonal. The estimation of diffusion parameters from raw DWI (using the general expression shown in Equation 2.6) can be done using a linear least square approach, denominated linear least squares (LLS), or a non-linear approach – non-linear least squares (NLLS) (Alexander et al., 2008; Winston, 2012). Note that the LLS allows to obtain a quicker solution and it is not so demanding from the computational point of view, when compared with NLLS. However, the solution obtained with LLS might be not so satisfactory, due to the propagation of errors, since in LLS the noise enters in the system as a random variable statistically dependent on the signal (Ozcan, 2011). Both approaches were implemented in this project, and their performance will be analysed in the next chapters.

The diffusion tensor can provide several DTI invariant measures, such as the mean diffusivity (MD), fractional anisotropy (FA), axial diffusivity ($\lambda_{||}$) and radial diffusivity (λ_r).

The mean diffusivity corresponds to the trace of the diffusion tensor divided by three (Equation 2.19). This measure is proportional to the orientationally-averaged apparent diffusivity, and its values are remarkably similar across grey matter and white matter. MD is an inverse measure of the membrane density. This measure is sensitive to cellularity, edema, and necrosis. (Andrew L. Alexander, 2011). Thus, MD has an important role in the context of TBI diagnosis, since allows to evaluate the edema characteristics, which is a physical manifestation of TBI. The FA value quantifies the fraction of the whole ''magnitude'' of the diffusion tensor that can be ascribed to anisotropic diffusion (Equation 2.20, the λ corresponds to the eigenvalues in the three directions defined). Finally, the axial diffusivity refers to the magnitude of diffusion along the principal ellipsoid component, so it is equal to λ_1. On the other hand, radial diffusivity is the mean diffusivity along orthogonal components relative to the principal ellipsoid axis (Equation 2.21).

$$MD = \frac{Trace\ (D)}{3} = \frac{\lambda_1 + \lambda_2 + \lambda_3}{3}$$ Equation 2.19.

$$FA = \sqrt{\frac{3}{2}} \times \frac{\sqrt{(\lambda_1 - \langle\lambda\rangle)^2 + (\lambda_2 - \langle\lambda\rangle)^2 + (\lambda_3 - \langle\lambda\rangle)^2}}{\sqrt{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}}$$ Equation 2.20.

$$\lambda_r = \frac{\lambda_2 + \lambda_3}{3}$$ Equation 2.21.

Chapter 3

Multicentre Studies: State of the Art

3.1. Introduction

Medical Imaging is an attractive and non-invasive technique to diagnose, evaluate and monitor the therapeutic response of several diseases which affect the brain structures. Therefore, medical imaging techniques, such as DTI, have been proposed as a sensitive biomarker for traumatic white matter injury, and might be a useful tool for evaluating patient prognostics, tracking the microstructural changes during recovery from TBI (Sidaros et al., 2008).

However, determining the extent of clinically relevant neuropathology associated with TBI is problematic. Thus, further studies on its clinical manifestation and the therapeutic response are needed, in order to improve the treatment of this disease (Kraus et al., 2007). To improve characterization and classification of TBI and to identify the most effective clinical care, a multicentre study is being developed (CENTER-TBI, 2014). Multicentre studies allow for the analysis of larger populations, allowing for the inclusion of subjects with diverse demographic origins and with different pathological manifestations and therapeutic responses. Thus, it is expected that this project will achieve an impact in terms of treatment provision, health care policy and improve also health, quality of life, and social integration for patients (CENTER-TBI, 2014).

In the literature it is possible to find several advantages associated to multicentre studies, however a few limitations are also pointed out which might be problematic and question the reliability of these kind of projects.

In this chapter a brief description of the advantages and limitations of multicentre studies is presented. Considering the main problems of multicentre studies, the state of the art on quantification of the limitations of these studies is also presented, as well as several methodologies that might be useful to solve these issues. Note that most of the methodologies used to address the limitations of multicentre studies exhibit, so far, an empiric character and must be performed before or during data acquisition.

3.2. Advantages and Limitations of Multicentre studies

Multicentre studies are performed in multiple centres located in different sites and conducted by different professionals that have different resources at their disposal. This implies the use of different medical-machines with different values of calibration as well as different patients. So, multicentre studies allow the study of a large number of patients and healthy subjects, comparing subjects from different demographic origins. By doing so, thanks to multicentre studies it is possible to study different pathologies on a wider scale, something which would be somewhat more difficult to study using a conventional approach (using a single equipment installed in a unique centre).

One of the most relevant characteristic of multicentre studies is their capability to increase the statistic power of the researches performed with the data acquired in this context (Suckling et al., 2012), which is achieved by the large number of subjects present in the sample (Friedman, Glover, & The FBIRN Consortium, 2006a). The large number of subjects is also useful to eliminate the confounding effects of inter-subject heterogeneity inherent to cross-sectional studies (Deoni et al., 2008).

However, despite the advantages that multicentre studies offer, their usefulness depends on the reproducibility of the data (Friedman, Glover, & The FBIRN Consortium, 2006b). In fact, although data from different sites can be pooled, studies revealed that systematic inter-scanner differences can occur (Abdulkadir et al., 2011). In the case of specific multicentre studies in which the data consists of medical imaging, in particular MRI, the data reproducibility is not always possible. This fact is due to the different calibration performed in the acquired data and the singular interaction between the subject and the magnetic field. Even two scanners from the same manufacturer and model will not produce the exact same images, which sometimes poses difficulties in ensuring the reproducibility of the data and, consequently, can reduce the reliability of the study (Kostro et al., 2014; Suckling et al., 2012). In fact, the variability introduced by the scanner is a potential source of errors in machine learning algorithms used to analyse the data, since the bias affects the learning function during the training phase and affects the choices of the algorithm during the test phase (in other words, scanner bias favours the prediction of one of the classes) (Kostro et al., 2014).

The literature suggests that the principal factors that might be in the origin of scanner noise and, consequently, influence the quality of multicentre studies, are (Deoni et al., 2008; Glover et al., 2012)

- Definition of scan parameters across scanners/vendors;
- The field strength;
- Field of view (FOV), spatial and temporal resolution;
- Slice orientation;
- Acquisition sequence;
- Coil sensitivity;
- Electronic amplifier gain.

In order to correct for the sources of error introduced by the factors mentioned above, it is important to:

- Monitor longitudinal stability in scanner performance, especially after hardware and software upgrades;
- Create a compliant database infrastructure that can store, retrieve, and monitor the variety of data collected.
- Make sure that the imaging parameters are as close as possible between sessions and centres.

The measures described above have an empiric character and can only be applied before the data acquisition. In other words, in order to avoid the bias induced by the scanner it is necessary to carefully plan the design of the process to be used in the image acquisition and for analysing the data shared among institutions. That allows to correct for scanner

noise by empirically adapting parameters, such as the width of window of acquisition[3] or the type of k-space filter[4] used during acquisition (Kostro et al., 2014; Sasaki, Ida, Yamada, Watanabe, & Matsui, 2007). Even with the best planning, some discrepancies between different institutions can occur. The main reason for that is related to the inability of a centre to adequately implement the plan for the multicentre study, due to insufficient resources or time to follow all the necessary requirements.

Despite this, even when controlling all the pre-acquisition sources of error, it is important to control also the variations in any step of the pre-processing pipeline such as segmentation, resampling or spatial normalization, which might introduce discrepancies among images (Abdulkadir et al., 2011).

To reduce the differences introduced by scanner variability, after image acquisition, algorithms can be used for error correction, such as correction by segmentation and intensity normalization algorithms (Kostro et al., 2014). However, it is important to note that the correction algorithms might not remove systematic differences. Besides, the use of these algorithms has to be done carefully since the goal is to remove the variability introduced by the noise of the scanner, without removing the features that can be associated to disease symptoms.

3.3. Quantification of Intra- and Inter-scanner Variability

As detailed before, the main issue for the viability of multicentre studies is the variability between scans introduced by the scanner. Therefore, to ensure the reliability of the study it is crucial to evaluate and quantify this variability.

3.3.1. Magnetic Resonance Imaging – MPRAGE

Ewers et al., (2006), in a multicentre study assessment of reliability of cranial MRI, namely using the MPRAGE sequence to acquire MRI images of a phantom, used the coefficient of variation (Equation 3.1.) to quantify the variability across scans, in which the σ corresponds to the standard deviation of the sample, μ is the mean of the sample and X is the sample in analysis.

$$CoV(X)_i = \frac{\sigma(X_i)}{\mu(X_i)} \times 100 \qquad \text{Equation 3.1.}$$

(Deoni et al., 2008), suggested that individual and group reproducibility for MRI (T1 and T2 images), might be assessed via whole-brain parameter histograms and intra- and inter-site coefficients of variation (COV[5]) (Equation 3.1). This parameter is calculated for each voxel, and it allows the comparison of scans acquired by the same scanner (in this case the matrix X is composed by the scans from the different subjects), or can be used for the comparison between scans acquired in different places (in this case, the X matrix is composed by scans from different subjects, acquired with different

[3] Determining DWI display conditions by using b0 images is a simple and practical method to standardize the window width and level for evaluating diffusions abnormalities and decreasing variation among institutions and operators (Sasaki et al., 2007).

[4] For example, the GE devices provide images which show higher spatial smoothing when compared with Siemens devices. The reason for that is Fermi k-space filter applied during image reconstruction in GE devices (Glover et al., 2012).

[5] The Coefficient of Variation (COV) is a statistical measure of the dispersion of data points in a data series around the mean. It represents the ratio of the standard deviation to the mean, and it is a useful statistic to compare the degree of variation from one data series to another, even if the means are drastically different from each other. (Lehmann, 1986)

devices). In both cases, it is possible to construct visual maps, which express the variability introduced by the scanner in the different structures of the brain, through the differences in colour intensities representing the value of COV obtained for that voxel.

(Huppertz, Kröll-Seger, Klöppel, Ganz, & Kassubek, 2010) achieved a mean COV, for intra-scanner analysis, ranging from 0.50% to 4.4% (median, 0.89%), while the inter-scanner COV varied between 0.66% and 14.7%, after pre-processing the images using a protocol similar to the one applied to the data analysed in this dissertation.

A different way to determine the variability in an MRI image is using the ANOVA – Analysis of variance. (Suckling et al., 2012), used this statistical model to detect the main sources of variation at each intracerebral voxel in standard Montreal Neurological Institute space (MNI) [6] for the image variables derived from each data acquisition type: grey matter tissue segmentations from structural Magnetic Resonance Imaging (sMRI). The statistical analysis was based on a fixed effects model, in which the factor modelling centre broadly accounted for up to 15% of the total variance for all the imaging variables tested in this treatment.

(Weiskopf et al., 2013), used also the COV to quantify the variability across scanners in images such as T1w, PDw and MTw, using just a few regions of interest, defined previously, and by calculating the COV and then comparing the data for these regions. This study suggested a normalization of standard deviation to N-1 sample size, in order to avoid the bias. Note that *N* refers to the sample size.

(Cannon et al., 2014), used the intra-class correlation to average the variability within and inter-scanner. In this study, a voxel based analysis and volumetric analysis were performed. In both types of analysis the results achieved demonstrated a high level of reproducibility intra and inter-site (between-site reliability was good, 0.75– 0.89, for overall grey, white and CSF volumes; the within-site ICCs were excellent, 0.90– 0.99).

3.3.2. Diffusion Tensor Imaging

In the case of DTI measures, such as FA and MD, the COV could also be used as measure of reproducibility (Teipel et al., 2011).

(Zhu et al., 2011) presented a study where the main goal was the quantification of accuracy of multicentre DTI measurements, and it suggests an improvement to the COV used before. In this case, nonparametric bootstrap techniques were used, such as wild bootstrap and residual bootstrap, as well as robust estimators of precision for DTI measurements. Wild bootstrap (Y_r^*) (Equation 3.3.), is a data driven procedure for estimating the properties of a given statistical value with minimum assumptions on the population's distribution. Thus, the value of COV can be estimated by the following expression:

[6] The MNI space defines a standard brain by using a large series of MRI scans on normal controls. The MNI wanted to define a brain that is more representative of the population. For that, a new template was created that was approximately matched to the Talairach brain in a two-stage procedure. First, it took 241 normal MRI scans, and manually defined various landmarks, in order to identify a line very similar to the AC-PC line, and the edges of the brain. Each brain was scaled to match the landmarks to equivalent positions on the Talairach atlas. Then, it acquired 305 normal MRI scans (all right handed, 239 M, 66 F, age 23.4 +/- 4.1), and used an automated 9 parameter linear algorithm to match the brains to the average of the 241 brains that had been matched to the Talairach atlas.(John C. Mazziotta, Arthur W. Toga, Alan Evans, Peter Fox, 1995)

$$CoV(X)_i = \frac{\sigma(Y)}{\mu(\overline{Y_r^*})} \times 100$$ Equation 3.2.

$$\overline{(Y_r^*)} = \frac{1}{R}\sum_{r=1}^{R}[Y_r^*]$$ Equation 3.3.

$$\sigma(Y) = \sqrt{\frac{1}{R-1}\sum_{r=1}^{R}\left[Y_r^* - (\overline{Y_r^*})\right]}$$ Equation 3.4.

where Y is the matrix with DTI measures (i.e, FA or MD measures) and R represents the sample size.

In this study, Zhu et al., (2011), suggested also other parameters for the measurement errors of multicentre DTI studies at multiple levels. The parameters suggested were:

- Measurement of Accuracy: the measurement of accuracy corresponds to the values of bias between actual MD/FA from each dataset and the corresponding values from the gold standard dataset constructed for phantom and human data separately.
- Measurement Precision (within-scan variability): the within-scan variability quantified by wild bootstrap analysis reflects the intrinsic uncertainty associated with each DTI acquisition and was consequently selected as the measure for precision.
- Inter-site Variance Component: it describes the reproducibility of DTI for repeated scans of the same subject or phantom among different sites.
- Intra-site Variance Component: it describes the repeatability of DTI measurements for repeated/longitudinal scans of the same subject/phantom at the same site.

The COV value was used as well by (Magnotta et al., 2012). In this particular study, using DTI invariant measures, the results suggests that subtle changes in the white matter architecture can be studied longitudinally if the subjects are scanned on the same scanner, since the COV value achieved ranged between 1-3%.

Regarding the COV values achieved by (Sakaie et al., 2012), they concluded that the variability within-scanner is lower when compared with the variability inter-scanner. Furthermore, the mean diffusivity maps (MD) had shown a lower value of COV when compared with FA images (the COV value for MD is 4.1% *versus* 9.6% for FA images, when an inter-centre analysis was performed). In this study, the results seem to suggest that the COV value is generally lower when evaluating white matter only.

In case of fMRI and DTI, a useful tool to estimate the variability across scanners is an indirect measure obtained by the comparison of the effect size expected if the images were acquired in one centre with the value obtained for data acquired in a multicentre study (Teipel et al., 2011). The effect size represents the difference in the FA measures of the regional fibre tracts between patients and healthy controls (Equation 3.5). Thus, when comparing the images from the same subjects acquired in different centres, the same value for effect size would be expected. If that is not true, it can be concluded that the difference is introduced by the scanners, and the variability across scanners can be estimated by the difference found in the effect size, it is represented by δ_c, as is indicated in Equation 3.6:

$$d_m = \frac{(x-y)}{\delta_{mxy}}$$ Equation 3.5.

$$d_c = \frac{(x-y)}{\delta_{mxy} + \delta_c}$$ Equation 3.6.

where *(x-y)* represents the mean difference in measurements between patients and controls, δ_{mxy} is the standard deviation of measures pooled across both groups; δ_c is the standard deviation of the measures across centres. Note, this model assumes that the components of variability, inter-subject component and inter-centre component, are independent.

A different assessment of a method's reliability is the intra-class correlation coefficient (ICC) which relates the within-subject variation to the between-subject variation from the following equation (Vollmar et al., 2010)

$$ICC = \frac{\sigma_{bs}^2}{\sigma_{bs}^2 + \sigma_{ws}^2}$$ Equation 3.7.

where σ_{bs} corresponds to the standard deviation of the population between-subjects and σ_{ws} refers to within-subject standard deviation for repeated measurements. Note that the ICC expresses the fraction of the total variance in the data that is caused by true biological variation between subjects rather than by measurement error within subjects. With this in mind, the higher values of ICC correspond to higher values of data reproducibility and, consequently, to larger degrees of reliability. In MRI data processing, an ICC larger than 0.7 is considered a reasonable level of data reproducibility; for ICC values larger than 0.8, the data shows a good reproducibility, and for values higher than 0.9 the level of reproducibility is considered excellent (Schnack et al., 2004).

In the literature, for multicentre studies, the definition of a baseline is also suggested, relative to which the performance of the scanners used in the study can be compared. The baseline is established by a scanner defined as the target scanner. This choice is based in estimation of coefficients of variation across subjects, whose scans were acquired in the same scanner. In this way, the scanner that shows the lowest COV, regarding the comparison of subjects' scans, is chosen as the target scanner. The difference between the scanner defined as the target and the other scanners corresponds to a measure of variability caused by the hardware (Moorhead et al., 2009).

(Vollmar et al., 2010) used the coefficients of variation (CV) and intra-class correlation coefficients (ICC) of FA measures of the whole brain, in three regions of interest (ROI) and in three tracts derived from these ROI by probabilistic tracking. In this study, Vollmar et al. assessed the impact of affine, nonlinear and template based methods for spatially aligning FA maps on their reproducibility. The intra-site COV for FA ranged from 0.8% to 3.0%, with ICC from 0.90 to 0.99. On the other hand, the inter-site COV ranged from1.0% to 4.1% with ICC of 0.82 to 0.99. The results of this study showed an improvement in the reproducibility when nonlinear image co-registration was performed instead of affine co-registration. However, the normalization to template space reduced the between-subject variation, resulting in lower ICC values and indicating a possibly reduced sensitivity.

Performing an ROI analysis to analyse the inter-subject variability, (Veenith et al., 2013), obtained a COV value for FA images of 7.9% (3.3–31.7%) and for MD images of 7.3 (2.4–33.7%). In addition, the same study suggested that the COV for

repeated DTI measurements obtained during the same session was lower than that obtained from measurements obtained in a different imaging session separated by up to six months, which is consistent with previous studies.

3.3.3. Quantification methods for MPRAGE and DTI images

Different kind of approaches can be used to quantify the variability among scanners. Examples of that are General Linear Models (GLM) and classifiers, such as Support Vector Machine (SVM). In the first case, each scanner used is considered as a variable of the model (Equation 3.8). The beta coefficients estimated in the model are a measure of variability brought by the variable with which it is associated. It is important to note that other variables can be introduced in the model, in order to absorb the variability introduced by other sources, such as the variability across subjects resulting from pathological manifestations.

$$y = x_1 \times \beta_1 + \cdots + x_i \times \beta_i + \varepsilon$$ Equation 3.8.

Considering the concepts behind SVM, this approach does not produce a direct quantification of variability. However, using a classifier it is possible to determine how significant the level of bias introduced by the scanners is. In other words, if the classifier distinguishes the scans acquired with different scanners - considering that the subjects analysed do not exhibit significant anatomic and physiological differences (the group must be composed by healthy controls, with a homogenous distributions of age and gender) - this means, with a high level of accuracy, that the variability introduced by the scanner is too significant and, consequently, that the scans are not comparable (Kostro et al., 2014). In the absence of any inter-scanner effects, it can be assumed that the scanners in analysis produce similar and comparable images, and the classifier would not distinguish the scanners above chance level (50%).

3.4. Approaches to Variability Correction

As detailed above, medical images, such as MRI, vary among institutions and/or operators. These differences could be critical for the effectiveness of the diagnosis.

Most of the methodologies proposed to correct for this variability are based on empirical protocols, which depend on the radiographer and neurologist that execute and analyse the exams. In addition, the standard approaches, most of the time, sacrifice the potential gains in signal-to-noise ratio, in order to improve the gradient performance, which might result in a poor image resolution. Thus, this kind of procedure is not a solution to correct the data after acquisition and, consequently, cannot solve the issues of multicentre studies.

3.4.1. Structural Magnetic Resonance Imaging – MPRAGE

(Moorhead et al., 2009) in a multicentre study using MRI images, demonstrated with the implementation of scanner specific priors to correct the data that it is possible to reduce the differences between scanners. The model developed (Figure 3.1) was applied to grey matter and white matter maps. To develop the prior, a group of 6 subjects, chosen randomly among 13 subjects, was used. The procedure was iterated until the convergence of segmentations was achieved.

The evaluation of the methodology proposed was performed using 7 subjects as a test group, for whom scans were acquired with three different devices.

As mentioned previously, the error introduced by the scanner might be especially problematic for the performance of machine learning algorithms. (Kostro et al., 2014), explored the use of regression models to correct for the scanner variability, demonstrating the usefulness of regression in the reduction of data variability, without removing the disease effects. For that, a Gaussian Process Regression (a nonlinear and regularized regression method – Equation 3.9, the $y_{GPR}^{(*,v)}$ correspond to the values at voxel v for each centre after correction, $y^{(*,v)}$ is the vector of the values of voxel v for each centre in analysis before correction, $\left(k_{\theta,\alpha}^{*}\right)$ is the kernel achieved using the relationship between the training and the voxel that will be corrected from the test group – α and Θ are parameters present at the functions used to calculate the kernel - $K_{\theta,\alpha}$ is the Kernel matrix calculated with the training set) and an unregularized simple linear regression with ordinary least squares solution (OLS – Equation 3.10, $y_{OLS}^{(*,v)}$ is the vector of the corrected values for each centre in voxel v, $y^{(v)}$ is the vector of the values of voxel v from each centre before correction, x^{*} corresponds to the design matrix used to explain the variation of several covariates, such as age, gender and centre, and finally, β^{v} is the vector of the weight of each covariate in the regression to explain the variability of the voxel v) were applied to correct for the bias of the scanner. In both cases, the model developed by Kostro et al. applies a correction to the input data, including also confounding effects such as age, sex, scanner, total intracranial volume. In this case, as a machine learning algorithm, the model executes an evaluation of the correction effect during its implementation, allowing the iteration of the model until the best correction has been achieved. In other words, when applying this algorithm to MRI data correction, a model estimates the grey matter density value for every voxel based on examples of subject-specific covariates and corresponding grey matter density maps; after that, the model is applied to new data to obtain a subject specific template; finally, the template is subtracted from the observed grey matter map. This procedure is implemented iteratively until the model parameters which reduce the variability as low as possible are found.

$$y_{GPR}^{(*,v)} = y^{(v)} - \left(k_{\theta,\alpha}^{*}\right)^{T} K_{\theta,\alpha}^{-1} y^{(v)} \qquad \text{Equation 3.9.}$$

$$y_{OLS}^{(*,v)} = y^{(v)} - x^{*}\beta^{v} \qquad \text{Equation 3.10.}$$

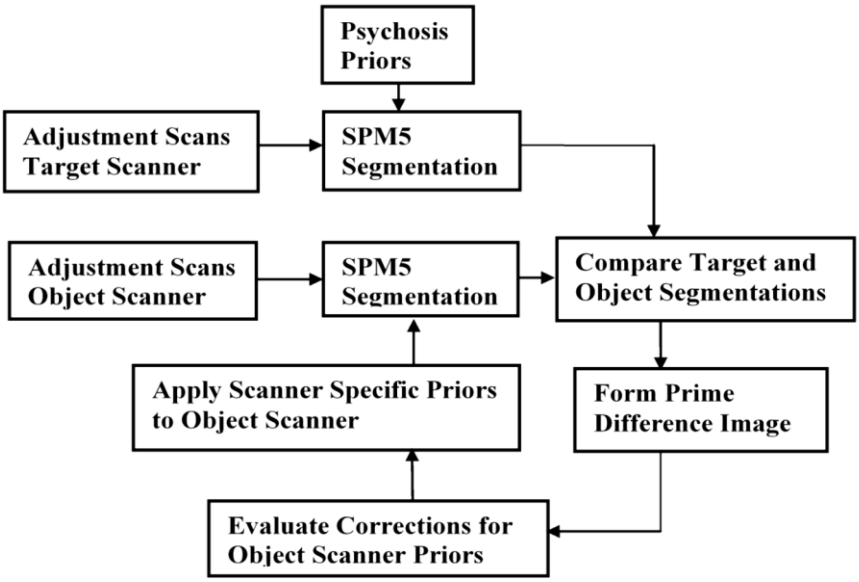

Figure 3. 1 – Schematization of the model applied by Moorhead et al., to correct the variability between scanners. Image adapted from (Moorhead et al., 2009).

3.4.2. Diffusion Tensor Imaging

The easiest way to correct for the variability across scanners is to apply a smoothness equalization. This approach is particularly useful in fMRI, and based on the premise that with the improvement of spatial smoothness an increase of temporal SNR in MRI can be achieved. Besides, it is important to note that the scanner smoothness is strongly dependent on k-space filtering. However, this approach reveals to be equally useful to correct DTI and MPRAGE images.

Considering these facts, Friedman et al. applied a methodology to reduce the differences between scanners, which is based on smoothness equalization. This approach consists in removing smoothness differences inherent to the images acquired in scanners with different level of magnetic fields[7], using a Gaussian kernel. To implement this methodology, a custom computer program was devised to assess iteratively the smoothness of a data set and then apply a Gaussian kernel to the data until the target smoothness, previously defined, was achieved. The result achieved demonstrated a 20% reduction of the activation effect size ratio between high (3 Tesla) and low-field scanners (1.5 Tesla). However, this approach has some limitations, namely the fact that data smoothness decreases the spatial resolution, and can remove important features of the signal, such as signs of pathology.

[7] In this study, the scans were acquired in scanners with two different field strengths: high-field scanner – 3 Tesla – and low-field scanner – 1.5 Tesla.

3.5. SUMMARY

In summary, multicentre studies provide several advantages over single-centre studies, such as the capability to study rare diseases and the capability to accumulate large samples for situations in which the effect size under study is small (V. a Magnotta et al., 2012; Teipel et al., 2011; Zhu et al., 2011).

However, this kind of project has some limitations, such as the variability present in the scans and introduced by the scanner. The reliability of these studies depends of the data reproducibility, which might not be achieved due to the bias introduced by the scanner. Nowadays, the presence and quantification of the variability between scanners is well documented; however, most of the approaches used to correct for this issue are applied during scan acquisition and are strongly dependent of the technician responsible for acquiring the scan. Just a few studies (Friedman et al., 2006a; Friedman, Glover, & The FBIRN Consortium, 2006b; Kostro et al., 2014; Moorhead et al., 2009), present a solution for processing the data and correct for the variability across scanners after the acquisition phase. For that reason, the aim of the present study was to develop a method to quantify and correct for the intra- and inter-scanner variability.

Chapter 4

Quantification of variability Intra- and Inter-scanner

4.1. Introduction

The usefulness of multicentre studies depends of data reproducibility (Friedman, Glover, & The FBIRN Consortium, 2006). As described in Chapter 3, the most significant limitation of multicentre studies is the variability across scans resulting from the hardware characteristics or differences in the sequence implemented, which compromises the reliability of the study (Teipel et al., 2011).

In this way, to ensure the reliability of the study, it is crucial guaranteeing the reproducibility of the data. Comparing controls' scans, a significant variability would not be expected among them, since there aren't any morphological or functional differences between the subjects' brains – absence of disease, similar age and similar educational level – and the protocol of image acquisition used is the same for all scans acquired. Even so, as is suggested by recent studies (Huppertz et al., 2010; Teipel et al., 2011; Zhu et al., 2011) and mentioned earlier, significant differences can be caused by the scanner.

In the literature, as detailed in Chapter 3, the Coefficient of Variation (COV) is suggested as a measure of the variability intra and inter-scanner to determine the data reproducibility. In a previous multicentre study using DTI data, Christian Vollmar et al (2010) obtained an intra-scanner COV ranged from 0.8 % to 3%, comparing FA maps from subjects acquired with the same scanner. For inter-scanner variability, the same study, presented a COV value ranging from 1 to 4.1 %. In more recent studies, (Fox et al, 2012) and (Venkatraman, Gonzalez, Landman, Goh, & Reiter, 2015) pointed to a higher variability for FA compared to MD maps in intra- and inter-centres analyses. In the first case, Fox and his collaborators found a COV value for voxel based analysis within scanner of 9.1% for FA maps and of 4.8% for MD maps. In the same study, in an analysis performed in by regions of interest previously defined, the COV value obtained was 6.8% for FA maps and 2.2% for MD maps. These values tend to increase when the same analyses is performed inter-centres. In that case, for voxel analysis the values obtained were 9.6% and 2.2% for FA and MD respectively, and 8.7% in FA images and 2.2% in MD images for ROI analysis. On the other hand, Venkatraman obtained a 4.57% and 3.56% for FA and MD images, when an intra-centre analysis based on ROIs were performed, and 5.04% and 6.03% when the analysis was done across centres.

Finally, in case of MR images acquired with the MPRAGE sequence, the variation within scanner pointed to 0.70% for grey matter (GM) maps and 0.72% for white matter (WM) maps (Huppertz et al., 2010). In the same study, the COV value across scanners obtained for GM maps was 3.79% and 4.14% for WM maps.

In this chapter, an analysis of the variability within and across scanners is presented. For that purpose, the COV was determined for each voxel of the FA, MD, GM and WM maps, and then the results were compared from the three centres analysed. Two types of analysis were performed. The first was an intra-centre analysis which was done independently, using a sample of images created for each centre. The second was an inter-centre analysis, using a different sample of images from that of the first study, created to determine the COV variability among centres.

The determination of the COV was performed to characterize the variability of the data and to prove the existence of the problem proposed by this dissertation and the need for improved solutions, which are also suggested in this document. Although this type of analysis was already performed in previous studies, it is crucial to perform a study for this specific data because the variability introduced by the hardware has specific sources of error which are influenced by factors such as the sequence applied, the environment in which the scanner is, and the specific monitoring protocols of the scanner performance, i.e, factors that are closely related to the scanner in analysis. With this in mind, the data will show a specific variability, which will require an accurate measure in order to characterize the data and perform an accurate evaluation of the models applied to correct it.

4.1. METHODOLOGY

4.1.1. Group definition and Data pre-processing

In this section three sample groups were defined, which were composed by seven healthy subjects from each centre, denominated as Controls from this point:

Centre 1: Cognition and Brain Sciences Unit (CBU), age (mean ± standard deviation) – 35 ± 11 gender (m/f [8]± standard deviation) – 0.62 ± 0.03;

Centre 2: Finland Centre, age (mean ± standard deviation) – 35 ± 12, gender (m/f ± standard deviation) – 0.64 ± 0.04;

Centre 3: Wolfson Brain Imaging Centre (WBIC), age (mean ± standard deviation) – 35 ± 13 gender (m/f ± standard deviation) – 0.50 ± 0.05.

The MRI images (DWI and T1 weighted images) were acquired in a 3T Siemens Trio Scanner (Centre 1) and in 3T Siemens Verio Scanners (Centres 2 and 3), and the parameters and sequences adopted were specified below.

DWI were acquired using an EPI sequence. The same sequence was applied at the three centres, with slight differences in the sequence parameters. Centre 1: field-of-view 192 mm, repetition time 2000 ms, echo time 30 ms, voxel size 2×2×2 mm^3, b-value=1000s/mm^2, 64 directions. The sequence parameters adopted at Centre 2 were: field-of-view 192 mm, repetition time 11700 ms, echo time 160 ms, voxel size 2×2×2 mm^3, b-value=1000s/mm^2, 63 directions. The sequence parameters adopted at Centre 3 were: field-of-view 192 mm, repetition time 7800 ms, echo time 90 ms, voxel size 2×2×2 mm^3, b-value=1000s/mm^2, 63 directions. Note that these differences between scanners and sequence parameters are representative of the differences involved in multicentre studies, and therefore these data are useful in quantifying the extent of the issue. In fact, in literature the differences between scanners and sequences parameters are even more significant, for example, scanners with different field strength were used and from different manufactures. The

[8] The proportion of males/females was calculated considering the number of males in the samples, dividing by the total number of subjects in the sample.

DTI invariant measures were determined using a homemade script, developed in order to calculate the non-linear estimation of the tensor. The pre-processing was perform using the FMRIB Software Library v.5.0 (FSL v5.0) software. This software consists of a comprehensive library of analysis tools for FMRI, MRI and DTI brain imaging data, developed by the Analysis Group, FMRIB, Oxford. (Stephen M. Smith, et al, 2004). The FSL routines were implemented for eddy current distortion correction and mask creation. The pre-processing was then completed using a spatial normalization of all images to a common space – MNI template. This final step is crucial for head position correction, multi-modal fusion and mapping to Talairach space (Jenkinson & Smith, 2001). Thus, a study was performed to test the best approach for registration. This study concluded that the linear normalization followed by non-linear normalization to an FA-template provided the best correspondence among images and a reduction of the variability across them (Appendix D). Thereby, the FLIRT[9] function was used followed by the FNIRT[10] function, both FSL tools. The registration was initialised and ran to convergence with sub-sampled images, a field of low resolution and a high regularisation weight. After that, the images from the first step were up-sampled, the regularisation modified and it was again ran to convergence (Andersson, Jenkinson, & Smith, 2007).

The T1 images were acquired using an MPRAGE sequence. At Centre 1 the sequence parameters adopted were: field-of-view 256x256x192 mm^3, repetition time 2530 ms, echo time 1.64 ms, voxel size 1×1×1 mm^3, TI 1200 ms, flip angle 7 degrees, bandwidth 6.51Hz/Px. At Centre 2 and 3 the sequence parameters adopted were: field-of-view 256x256x176 mm^3, repetition time 2300 ms, echo time 2.98 ms, voxel size 1×1×1 mm^3, TI 900 ms, flip angle 9 degrees, bandwidth 240Hz/Px.

The post-processing was performed using the SPM12 (http://www.fil.ion.ucl.ac.uk/spm/software/spm12/) routines for voxel-based analysis (tissue segmentation, DARTEL and normalization to a common space). Further details about the post-processing functions used can be found in (Ashburner, 2010)

For each group, two DTI invariant measures were analysed – Fractional Anisotropy (FA) and Mean Diffusivity (MD) – and Grey Matter (GM) and White Matter (WM) images obtained from T1 segmentation.

4.1.1. Quantification of variability – Coefficient of Variation

The variability across scans was measured using the COV, as suggested by the most recent studies developed on context of multicentre projects. (Huppertz, 2010; Magnotta et al., 2012; Teipel et al., 2011; Venkatraman, 2015; Vollmar, 2010).

Bearing this in mind, the COV was defined as the proportion between the standard deviation and the mean value for the same measure across samples- Equation 3.1. Note the value obtained using the Equation 3.1 is expressed as percentage value.

The COV was used to characterize the data from the three different centres mentioned above, which contributed for a preliminary analysis of the multicentre data. Although one of these centres doesn't belong to the main project (CENTER-TBI project, CENTER-TBI, 2014), the variability present in the scans acquired in this centre was studied to increase the

[9] The FLIRT function performs translations, rotations, zooms and shears.

[10] This function modelled the displacement fields as linear combinations of basis-functions, which may be the discrete Cosine transform (DCT) or cubic B-splines placed on a regular grid.

statistical power of this first part of the study and to prove the viability of the assumptions about the variability originated by the scanners.

The COV value was used to quantify the variability within scanner and across scanners. In both cases the COV was measured in each voxel of the brain and in specific ROI defined in white matter and grey matter, the procedure used is detailed in section 4.2.2 (a) and (b), respectively. The algorithm to determine the COV was developed in MATLAB™. Further analyses, such as statistical analysis, were performed using MATLAB™ as well.

(a) Coefficient of Variation: voxel based analysis

The post-processing step of registration applied to all images before the data characterization was essential, in order to compare all the analysed scans. It is expected that each voxel from any scan corresponds to the same brain region. Thus, towards a voxel based analysis, a similar value for the corresponding voxels from different scanners is expectable. Assuming this, any relevant differences found are due to artefacts, whose sources are mainly due to the hardware. However, it is also expected to find individual differences between subjects since the same subjects were not sent across all sites.

In order to remove the artefacts caused by the interaction between magnetic field and the scalp, as well as the artefacts present in boundaries between the different tissues in the cranium, two binary masks were created for each centre, which were applied depending on the kind of tissue under analysis: GM or WM. For that, the MPRAGE scans from the same Controls were used. For each image a segmentation and a normalization of the raw image were produced. Then the various images were averaged in order to determine a final image from which was a binary image then achieved.

(b) Coefficient of Variation: Regions of Interest

Forty-eight ROI were defined in WM, using the Jülich histological (cyto- and myelo-architectonic) atlas (JHU) (Eickhoff et al., 2005). The description of the ROIs is available in Appendix A.1 .These ROIs were applied to all FA, MD and WM maps for each subject and ROI statistics were determined using MATLAB™. On the other hand, for GM images, 48 ROIs were defined in GM using the Harvard-Oxford cortical structural atlases (Desikan et al., 2006).

A mean value for each the voxels in the same ROI was used as the intensity for the ROI under analysis, for each subject. Mean and standard deviation across scans for each type of image were extracted per ROI. The difference between the same ROI across scans was determined through the COV value, which was calculated using the Equation 4.1.

The comparison among scans, intra and inter-centre, was done using boxplots, which display the variation of a sample taken from a statistical population without making any assumptions of the underlying statistical distribution. The boxplot indicates also the degree of dispersion (spread) and skewness in the data, and shows outliers. On the other hand, the degree of variability among ROI is represented by bar plots.

4.2. RESULTS

4.2.1. Variability quantification within centre

(a) Coefficient of Variation: voxel based analysis

The variability within each centre was compared for the four types of images analysed (Figure 4.1.). The values for each centre are shown as a bar which corresponds to the mean value of all voxels analysed. The results obtained showed a high variability which is shown by the deviation from the mean value. The main reason for this value of deviation, is due to some voxels which have a COV raised (approx. 150%). These higher values of variability are due to the localization of these voxels. As mentioned above, the voxels located in the boundaries between different brain tissues are more affected by partial volume artefacts (Le Bihan, Poupon, Amadon, & Lethimonnier, 2006), which result in a higher value of variability. Despite the attempt to reduce this effect by the use of a mask, the mask used was obtained through a mean image across subjects. Thus, in some cases the mask was not able to cut the voxels by the boundaries and effectively reduce the effect of partial volume which contributes to a higher value of variability at those specific voxels.

In order to represent the distribution of the variability in the brain, coloured maps were created which represent the COV value in each voxel. The coloured maps can be found in Appendix B.

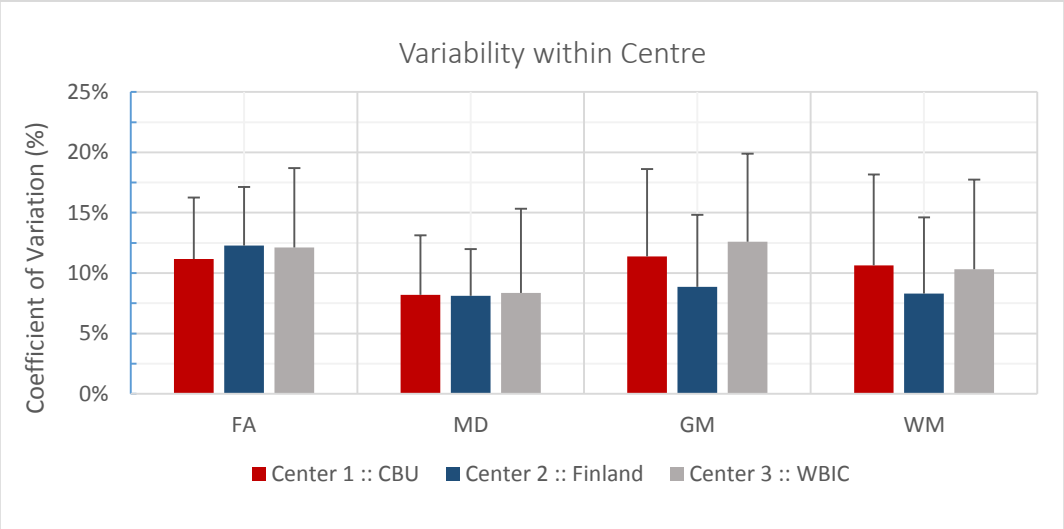

Figure 4.1 – Coefficient of Variation expressed as a percentage value for each type of image analysed. Each bar represents the mean COV value for each centre. The line corresponds to the standard deviation of the COV value.

(b) Coefficient of Variation: Regions of Interest

The ROI analysis allows to verify which ROI have a higher value of variation for the different types of image, namely the variability behaviour on FA, MD, GM and WM within each centre (Figures 4.2, 4.3, 4.4 and 4.5, respectively). In order to compare the results for each ROI and the results from the previous analysis, coloured maps were created to express the variability in each ROI (Appendix B). This type of analysis could also be compared with the previous analysis performed through the results presented in Figure 4.6, which summarizes the behaviour of variability across scans from the same centre. As shown in Figures 4.2, 4.3, 4.4 and 4.5, the ROI analysis seems to suggest a tendency for a lower variability in

scans from Centre 3, when compared to the Centres 1 and 2, namely in FA, MD and GM images. The exceptions are the WM images from Centre 2 which show a lower variability.

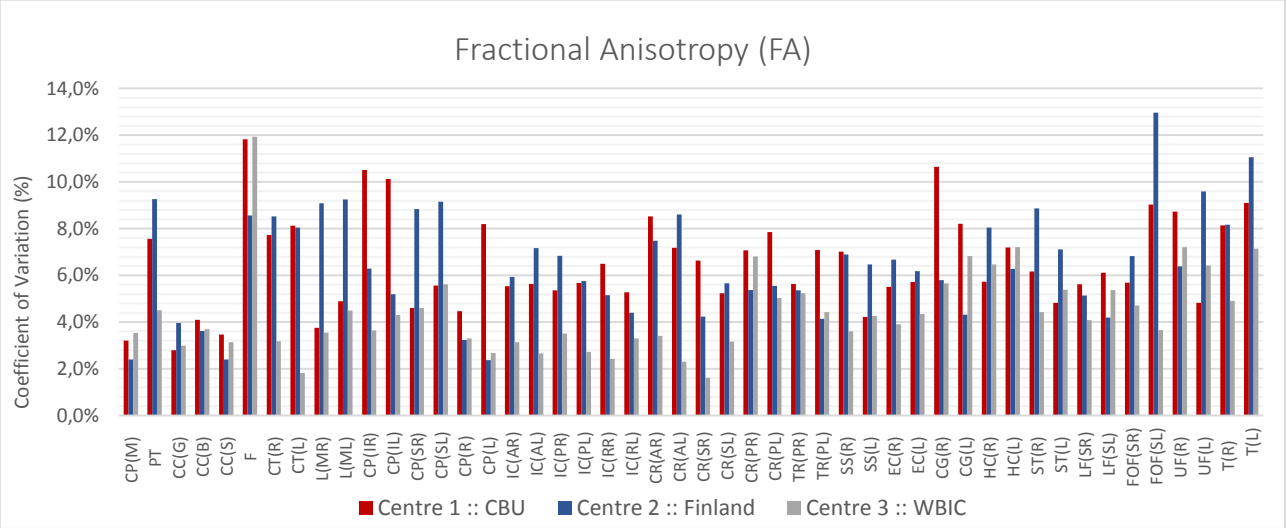

Figure 4.2 – Coefficient of Variation expressed as a percentage value for FA images. Each bar represents the mean COV value for each ROI in the three centres.

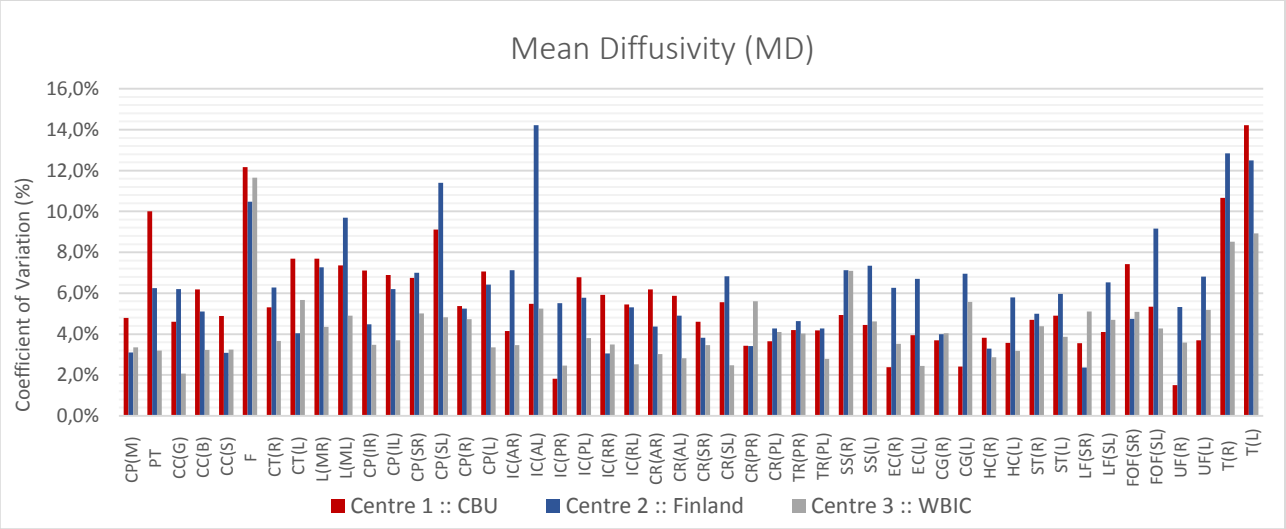

Figure 4.3 – Coefficient of Variation expressed as a percentage value for MD images.

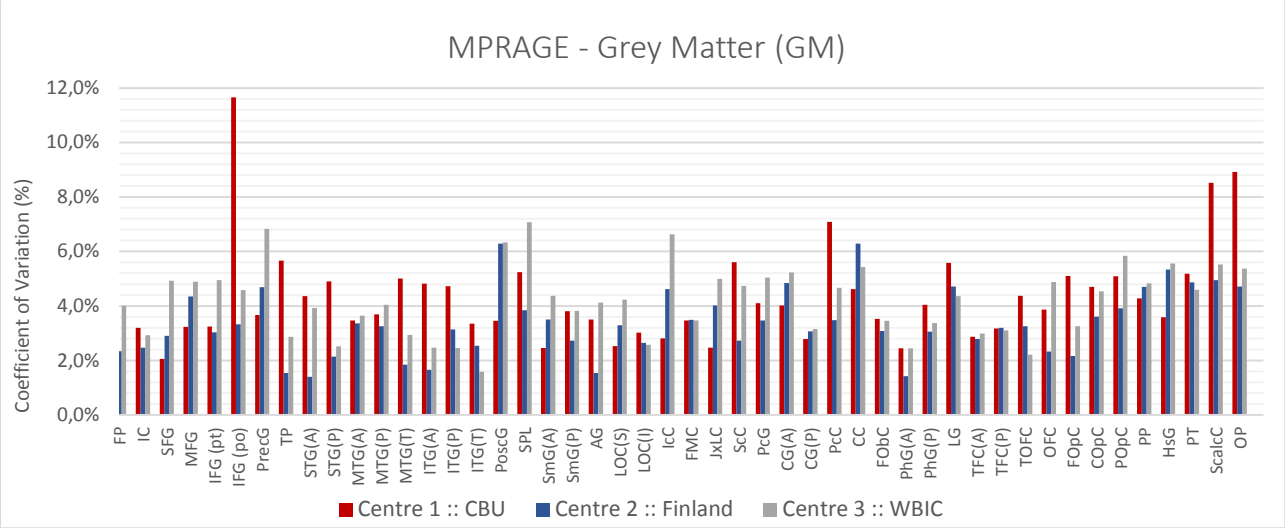

Figure 4.4 – Coefficient of Variation expressed as a percentage value for GM images.

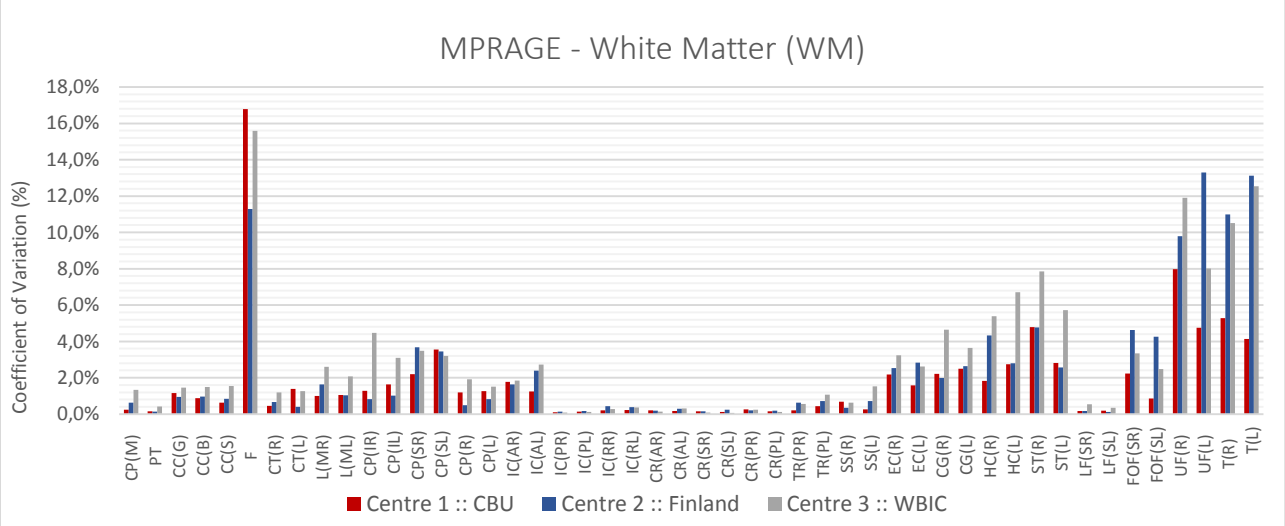

Figure 4.5 – Coefficient of Variation expressed as a percentage value for MD images.

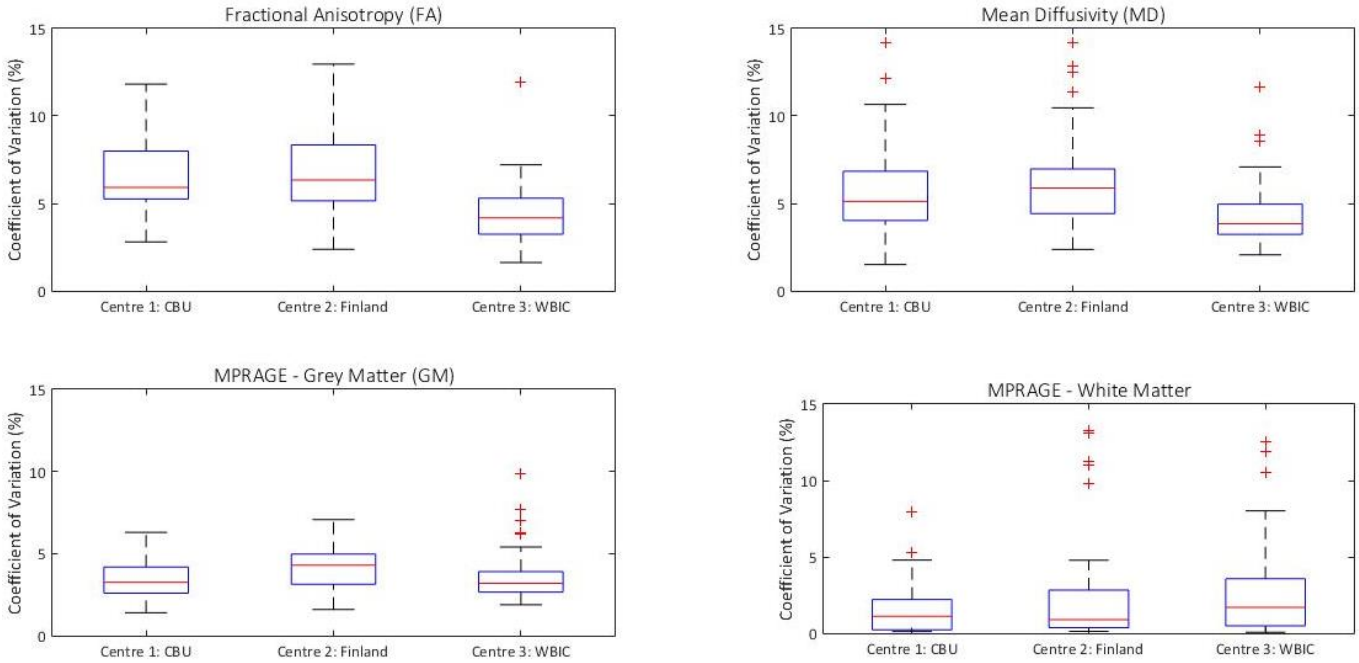

Figure 4.6 – Summary of the analysis performed for the 48 ROI defined. The red stars correspond to the values defined statistically as outliers.

The results from the quantification of variability within centre are summarized in Table 4.1. The data show that values of COV obtained by the ROI analysis (denominated at the table as column (b)) are lower than the COV values obtained by voxel analysis.

On the other hand, regarding the fact the data from ROI analysis came from an average of the voxels in each ROI, the standard deviation of the COV value obtained by ROI analysis are also lower than the standard deviation for the voxel based analysis (Table 4.1 - shaded area).

Table 4.1 – Summary of the Coefficient of Variation obtained within centre. The labelling (a) and (b) refers to the voxel based analysis and ROI analysis described.

	Mean COV ± Standard deviation (%)					
	Centre 1		Centre 2		Centre 3	
	(a)	*(b)*	*(a)*	*(b)*	*(a)*	*(b)*
FA	11.2 ± 5.1	6.5 ± 2.0	12.3 ± 4.9	6.5 ± 2.3	12.1 ± 6.6	4.4 ± 1.8
MD	8.2 ± 5.0	5.6 ± 2.5	8.1 ± 3.9	6.2 ± 2.6	8.4 ± 7.0	4.3 ± 1.8
GM	11.4 ± 7.3	3.4 ± 1.2	8.9 ± 6.0	4.2 ± 1.3	12.6 ± 7.3	3.7 ± 1.6
WM	10.6 ± 7.6	1.8 ± 2.8	8.3 ± 6.3	2.5 ± 3.5	10.3 ± 7.4	3.0 ± 3.6

4.2.2. Quantification of inter-scanner variability

In order to determine the inter-scanner variability, an average of the COV was performed considering all the scans from the three centres. The procedure behind the COV calculus was the same used for the quantification of the variability within scanner. The mask used for each type of images was obtained from the MPRAGE scans from the three centres.

(a) Coefficient of Variation: voxel based analysis

The voxel based analysis for the variability across centres suggests the variability is higher in FA and GM images (Figure 4.7). That result is consistent with the previous results for analogue studies performed within centres.

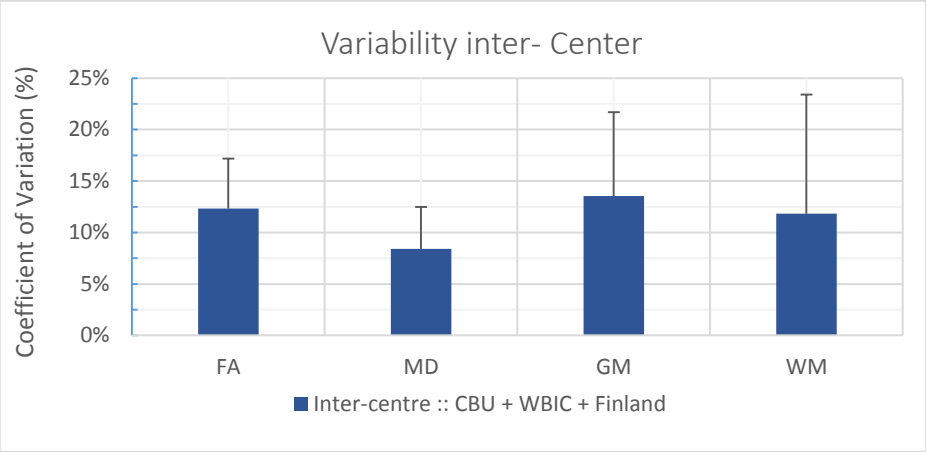

Figure 4.7 – Coefficient of Variation expressed as a percentage value for each type of image analysed. Each bar represents the mean COV value. The line corresponds to the standard error of the COV value.

(b) Coefficient of Variation: Regions of Interest

As described above, the variability introduced by the scanner was quantified in the 48 ROIs previously defined. The following figures (Figures 4.8, 4.9, 4.10, 4.11) present the mean value of COV in each ROI, for the four types of images analysed.

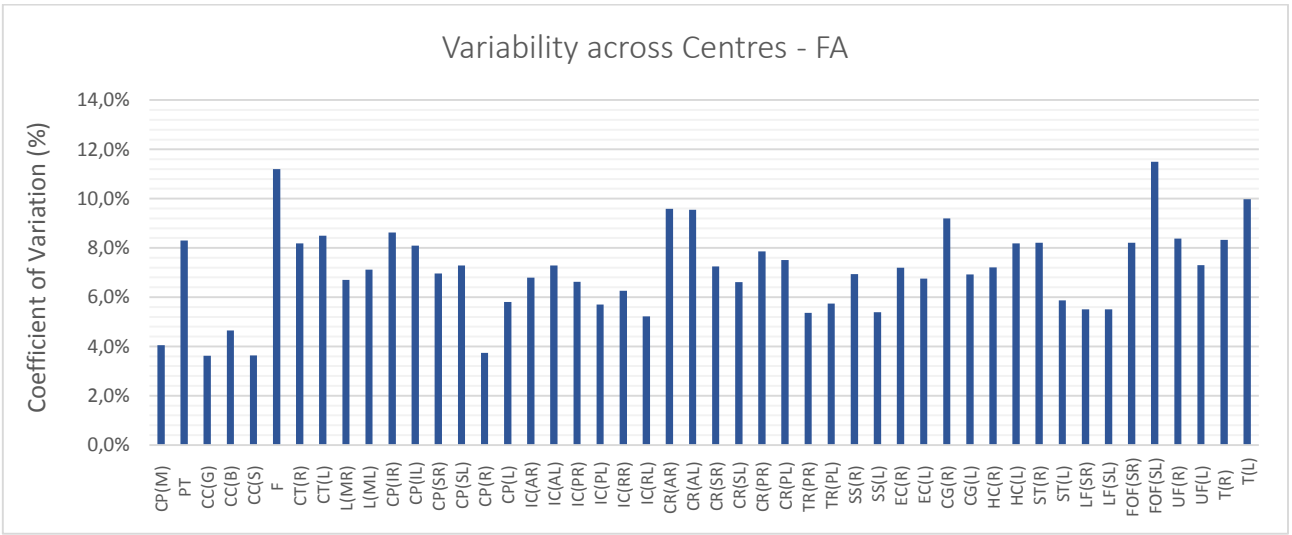

Figure 4.8 – Coefficient of Variation expressed as a percentage value for FA images. Each bar represents the mean COV value for each ROI across centres.

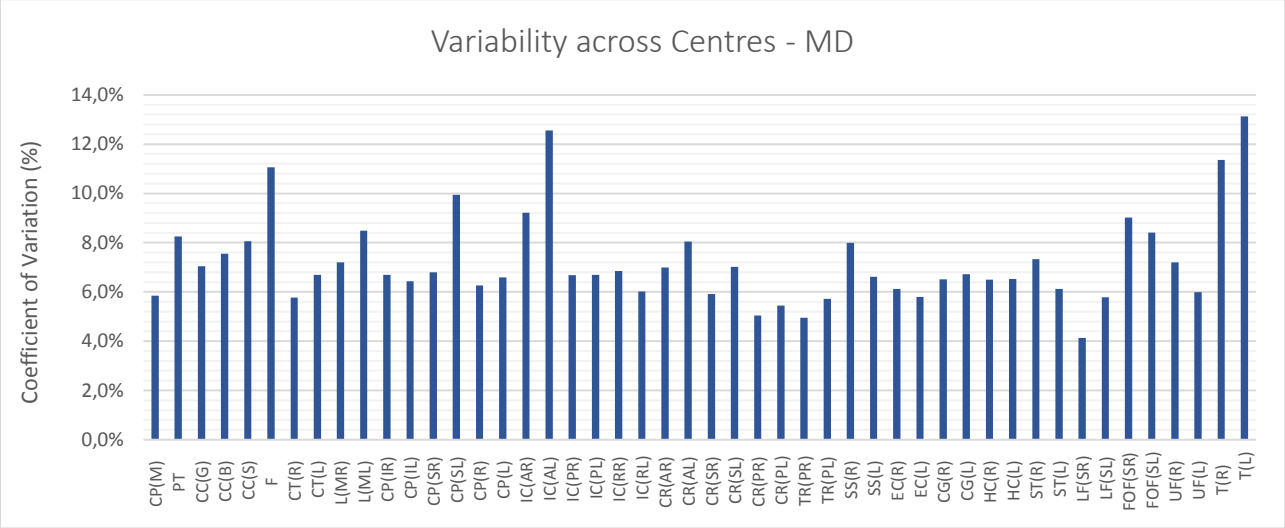

Figure 4.9 – Coefficient of Variation expressed as a percentage value for MD images for each ROI across centres.

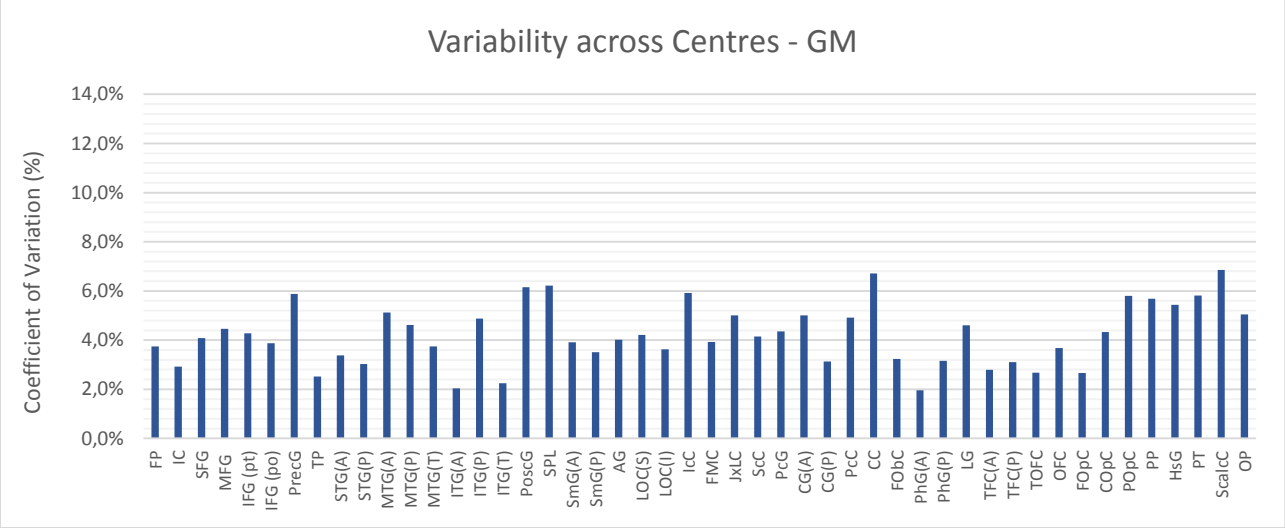

Figure 4.10 – Coefficient of Variation expressed as a percentage value for GM images across centres.

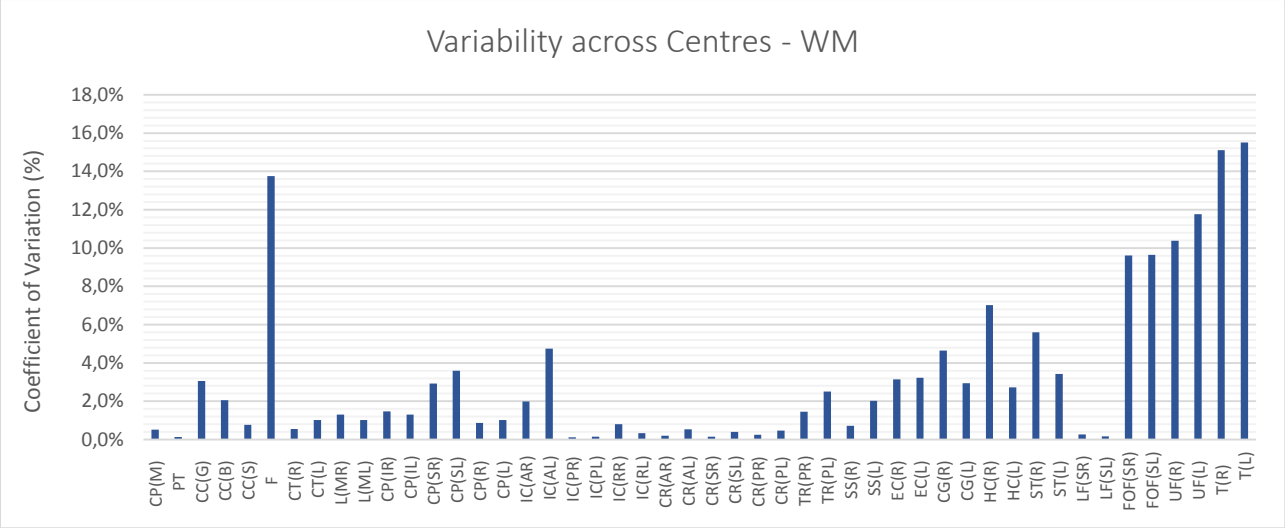

Figure 4.11 – Coefficient of Variation expressed as a percentage value for WM images for each ROI across centres.

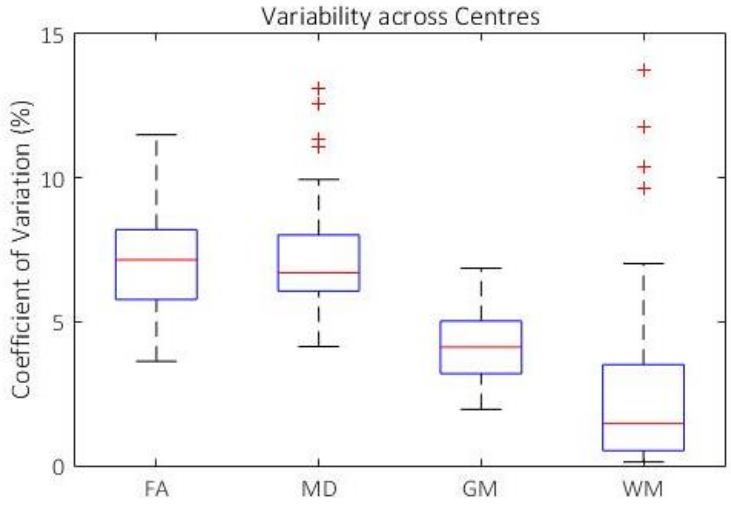

Figure 4.12 – Summary of the analysis performed for the 48 ROI defined. The red starts correspond to the values defined statistically as outliers.

The results from the quantification of variability across centres are summarized in Table 4.2. The difference between the types of analyses, labelling as (a) and (b) in the following table, suggest a reduction of variability when ROI analysis is performed. The exceptions are the WM images that have higher values of standard deviation for the ROI analysis, not consistent with the rest of the images analysed. The explanation behind this could be found in Figure 4.12, in which four COV values defined as outliers are visible.

Table 4.2 - Summary of the Coefficient of Variation obtained across centres. The labelling (a) and (b) refers to the voxel based analysis and ROI analysis described.

	Mean COV ± Standard deviation (%)	
	(a)	*(b)*
FA	12.3 ± 4.9	7.1 ± 1.8
MD	8.4 ± 4.1	7.2 ± 1.9
GM	13.5 ± 8.2	4.2 ± 1.2
WM	11.8 ± 11.5	3.7 ± 5.6

4.2. Discussion

4.2.1. Quantification of intra-scanner variability

In the quantification of the variability, which is useful to infer about the level of bias introduced by the scanner and the protocol, for the three scanners, the value of COV were similar among scanners. However, for both the voxel based analysis or for the ROI analysis, the values obtained are slightly higher than the values referenced in the literature for similar studies. In spite of the difference, these results are not meaningful, because this slight difference can be due to differences in the sequence (not all the studies applied the same sequence parameters, even if the sequences applied were also EPI and MPRAGE sequences), and the reduced number of subjects which might be not be enough for a robust statistical test; besides, this reduced number of subjects was critical when the standard deviation was calculated for the intra-scanner analysis, since with just 7 controls the standard deviation is higher than the expected value for a sample with more subjects. However, it was impossible to increase the number of subjects used in the samples tested, since it is the number available in order to have matched groups. Therefore, the reduced number of subjects compared with the previous studies, in which the sample size was 9 (Vollmar et al., 2010), and the reduced number of centres might be an issue in the comparison of the COV achieved when compared with previous studies.

The COV values achieved can be due to a MRI calibration or the environmental conditions where the scans were acquired, which can be determined by the stability of the magnetic field and the interactions between the magnetic field and the subject scanned.

In case of the voxel based analysis, none of the centres shows a tendency to have a higher variability across subjects. Thus, for the FA images the Centre 2 has a higher COV value (12.3±4.9%), while the Centre 3 shows a higher value for the MD and the GM images (8.4±7.0% and 12.6 ±7.3% respectively), and the WM images have a higher COV in Centre 1 (10.6±7.6%). On the other hand, for the ROI study, the Centre 2 shows consistently more variability when compared with the other centres. The exception are the WM images in which the higher value of COV came from Centre 1 (10.6±7.6%). Also note that the variability is lower in the ROI study. The ROI analysis allows to reduce sources of variability across scans, such as the partial volume effect. In this way, this type of analysis could be useful in multicentre studies as a first approach to reduce the variability across scanners. However, in certain studies where a more accurate spatially analysis is needed, such as Parkinson's and Huntington, this approach is not recommended since small features present in specific tracts in some region can be removed when the mean value of that region is calculated. In fact, the ROI analysis reduces biological heterogeneity within each ROI, and if the biological changes are subtle and ROI analysis might miss it. In this particular

study, as the illness in focus is the TBI this approach could be applied without significant consequences. Considering the TBI characteristics, in some situations, namely in severe TBI in which the patients' brain is seriously damaged and deformed, the ROI analysis is the only approach available, since in these particular cases the segmentation is not efficient considering the level of brains' deformation presented in MRI scans, and automated registration algorithms are not able to normalise these abnormal brains.

4.2.3. Quantification of inter-scanner variability

As expected, the variability across centres is higher than the variability within centre, in spite of the scanners in the three centres being from the same company. However, the sequence used had slightly differences which can explain the increasing of the variability across scanners.

The stability of magnetic field could be affected by external factors, such as the temperature and humidity of the room. Also, the scanners are located in different centres in different locations, and this fact implies that the person responsible for the scanner calibration and maintenance is different for each one, and thus the calibration procedure might have small differences, which may constrain the scanner performance. These factors may also contributed for the raised variability across scanners.

Considering to the ROI analysis, it is possible to conclude that the diffusion measures are more susceptible to scanner artefacts. This fact is consistent with previous studies in which the same result was observed. It is also possible to observe that the Fornix and the Posterior thalamic radiation have a higher value of COV, whilst the GM regions with more variability are the Supracalcarine Cortex and the Cuneal Cortex, with COV value above 6%.

4.3. Conclusion

Due to initial comparisons between the performances of these three scanners, it is possible to conclude that the data reproducibility is not achievable without an initial correction, since an error inferior to 5% would be acceptable. Besides, visual assessment of the variation of COV among ROIs (Appendix B), enables to verify that the pattern of variation is not the same for the data acquired for the various centres, which suggests that the variability measured is random and caused by the scanner, which means that the sources of error are independent from any particular feature and inherent to each subject. However, note that this conclusion could only be supported by an analysis using the same controls in each centre in order to compare the variability pattern for the same subject in different centres.

Chapter 5

SPATIAL FILTERING MODEL

5.1. INTRODUCTION

After confirmation of the presence of variability introduced by the hardware, which could undermine the reliability of the multicentre study, three corrections models were tested in order to ensure the minimization of this error in the MRI images.

The first model was based on a spatial filtering by smoothing method, which is usually applied to reduce artefacts such as the ones caused by subjects' movement (Le Bihan et al., 2006).

Spatial filtering is applied by convolving the images with an isotropic Gaussian kernel. In practice, each voxel in the result image - smoothed image - will contain the weighted average concentration of the voxels around each original voxel. Note that the region around the voxel is determined by the shape of the smoothing kernel.

Even though the smoothing process can be advantageous, since it increases the validity of parametric statistical tests, according to the central limit theorem, as smoothing has the effect of rendering the data more normally distributed. This fact was not the purpose of using this method due to the groups' size and nature of the data available. For this reason, the statistical tests performed were non-parametric tests. However, many studies mention smoothing as the first pre-processing step to treat MRI scans, namely structural images such as MPRAGE scans. Moreover, the smoothing step also helps to compensate for the inexact nature of the spatial normalization.

Whenever possible, the size of the smoothing kernel should be comparable to the size of the expected regional differences between the groups of brains.

The evaluation of this method was performed through the analysis of the variation of the z-score values before and after the correction, which could be understood as a measure of the difference between the patients' values when expressed in terms of the baseline defined by the controls or between the controls from one centre to another. This analysis was performed for each subject, in the 48 ROIs previously defined.

Moreover, statistical tests were also performed in order to understand if the correction has a significant effect on the detected differences in the context of group analysis. For that purpose, the tests applied consisted of non-parametric 2 sample t-tests, used to evaluate how different were the subjects from Centre 2 compared to the subjects from Centre 3, before and after the correction was applied. The Centre 1 was not included in this analysis, since the dataset from this centre had not TBI patients. The CENTER-TBI project did not include the CBU as research centre in which the TBI patients were analysed.

5.2. METHODOLOGY

5.2.1. Group definition and Data pre-processing

In this section two sample groups were defined, composed by subjects from Centre 2 and from Centre 3. The subjects in these samples were part of the groups defined in Chapter 3. In order to evaluate the developed models, these two groups were separated in two subsamples defined as Controls (composed by healthy subjects) and Patients (subjects with mild TBI diagnosis, which were analysed by MRI in their hospital assessment), taking into account the mean age and the proportion of gender that makes the subsamples suitable and comparable.

- Centre 2: Finland
  - Controls: 4 subjects
  - Patients: 58 subjects
- Centre 3: WBIC
  - Controls: 7 subjects
  - Patients: 21 subjects

The MRI images (DWI and T1 weighted images) were acquired using the scanners and sequences described in Chapter 4.

Furthermore, the pre-processing data pipeline adopted in this study is detailed in the previous chapter (section 4.2.1 of this document).

As before, for each group two DTI invariant measures were analysed –FA, MD – and GM and WM for the images obtained from T1 segmentation.

5.2.2. Spatial Filtering Method

To minimise the effects of noise, a spatial filtering method was applied to the raw DWI volumes using a Gaussian kernel with FWHM=2mm. The FWHM value was chosen taking into account the size of the voxel (2 mm), in order to achieve the noise reduction without removing too many relevant features of the scan. The smoothing filtering was applied using the FSL routines, which use an isotropic kernel.

In the same way, the MPRAGE images were corrected using a Gaussian kernel with FWHM=2mm. Considering previous studies (Ashburner, 2010), a more standard value for MPRAGE correction was also tested, in order to evaluate the effectiveness of the smoothing with a higher value for the FWHM (FWHM=8mm) without blurring the scans and remove relevant features. These two types of correction performed on MPRAGE images were denominated as Correction A and Correction B, respectively. The MPRAGE scans were treated using SPM8.

(a) Z Score

The Z score, also denominated as standard score, is a statistical measure, dimensionless, which expresses the deviation of a value to the mean of the population. Thus, this measure provides an assessment of how off-target a value is. The Z score was used to express the values obtained for the Patients from each centre in terms of the corresponding

measure obtained for the Controls from the same centre. In this way, the Controls were used to calculate the population mean and standard deviation, and were defined as baseline for each centre, while the Patients were expressed as a difference to the expected value - baseline.

The Z score was obtained by subtracting the Controls' population mean ($\overline{x_c}$) from an individual raw score, corresponding to each Patient (y), and then dividing the difference by the Controls' population standard deviation (σ_c) (Equation 5.1).

The analysis performed consisted on the assessment of how many Patients were significantly different from the Controls. In addition, it was also tested how many Controls from Centre 2 were significantly different from the Controls from Centre 3. An ideal dataset would show a high number of Patients significantly different from Controls, and a reduced number of Controls significantly different among centres. Note that the subject was categorized as significantly different if the Z score value was above 2 or if it was lower than -2.

This method was applied to the 48 ROIs defined in WM and GM.

$$z\ Score = \frac{y_i - \overline{x_c}}{\sigma_c} \qquad \text{Equation 5.1.}$$

(b) Group Analysis

A group analysis was performed in order to infer on the effectiveness of the spatial filtering correction. Thereby, a nonparametric 2-sample t-test was performed before and after the data correction. The first test used raw images, and the goal was to distinguish Patients from Centre 2 from the ones from Centre 3. This approach was informative in assessing the method analysed, since using this approach it was possible to test if the method was able to reduce the variability of the group of subjects. Actually, with this test it was analysed if the method is able to reduce the variability introduced by the scanner in a group of subjects by reducing the variability in each patient individually. If the method achieved this goal, it was possible to mix scans from different centres and using a spatial filtering in each of them to proceed to a group analysis without any further concern for the study reliability.

The second test was performed with the same goal, using however the smoothed images. A design matrix was used to define the main characteristics of the samples, such as the centre, age and gender. Note that, the same characteristics, namely the age and gender, were considered as nuisance variables and were ignored by the contrast established for the t-test. Finally, the number of permutations N was calculated using the Equation 5.2, where n1 is the size of group 1 and n2 is the size of the second group, in which the value obtained was $3.81 \times 10^{+38}$. Considering the high number of permutations needed, the number of permutations used was 10000, taken into consideration that this is the appropriate value for biological samples and that this number of permutations are required to reduce the margin-of-error to below 10% of the nominal alpha (Winkler, Ridgway, Webster, Smith, & Nichols, 2014).

$$N = \frac{(n_1 + n_2)!}{n_1! \times n_2!} \qquad \text{Equation 5.2.}$$

5.3. RESULTS

5.3.1. Spatial Filtering Model

(a) Z score

An intra- and inter-centre analysis was performed in order to evaluate the effectiveness of the correction in both cases. The results for intra-centre correction can be found in Appendix C.

In the Figures 5.1 to 5.8, it is possible to see the Z scores for inter-centre scans for Controls (Figures 5.1., 5.3., 5.5., 5.7) and for the Patients' group (Figures 5.2., 5.4., 5.5., 5.6.). The second set of figures (Figures 5.9 to 5.20) corresponds to the Z score obtained after correction. Note that in the case of MPRAGE images, as mentioned above, two different levels of correction were applied, and consequently an additional image for both Controls and Patients can be found as a result of this additional correction (Figures 5.19 and 5.20 respectively).

In all figures each subject is represented by a blue circle and a shadow represents the values between -2 and 2, which were considered as non-significant statistically.

(a.1) Before Correction

The data analysed in this section corresponds to the raw data – without any correction. Thereby, the Controls denomination refers to the Controls from the two centres, WBIC and Finland, and the Patients also refers to the dataset composed by the patients from both centres.

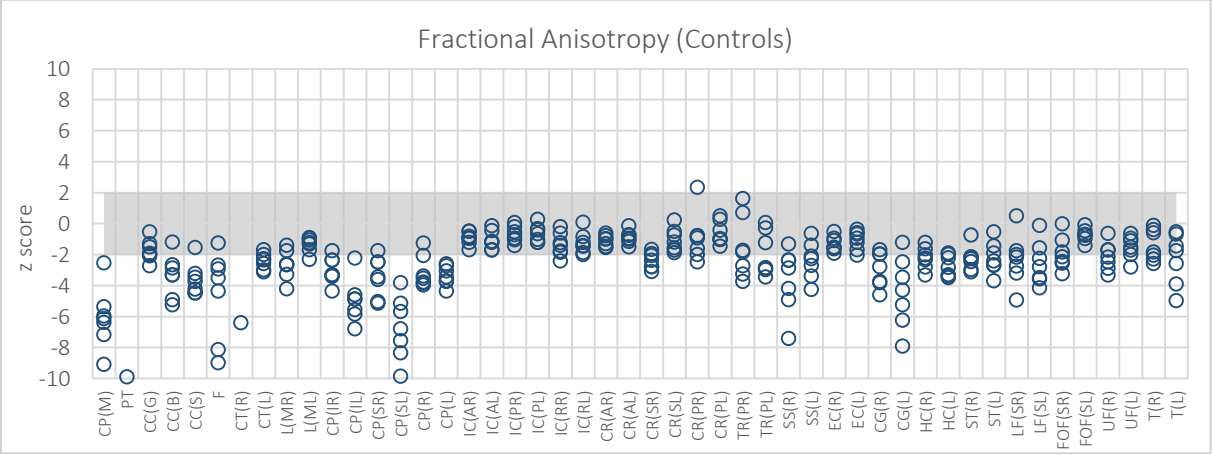

Figure 5.1 –FA maps. Z score of the Controls from Centre 3 – WBIC – when expressed in terms of controls from Centre 2 - Finland. The shadow region corresponds to the region of non-significant differences among subjects. Each circle represents a different subject.

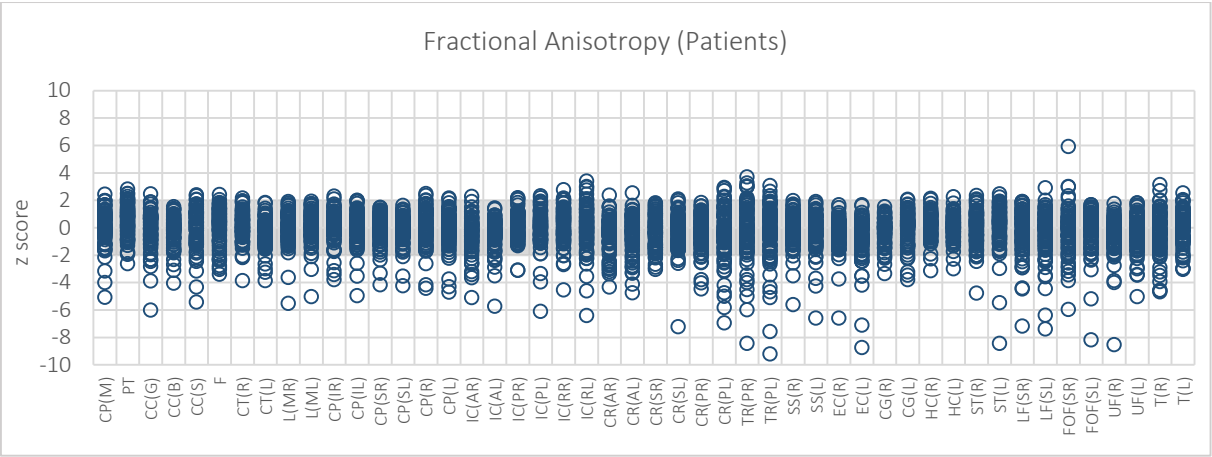

Figure 5.2 – FA maps. Z score of the Patients when expressed in terms of a baseline defined by the controls from the two centres under study.

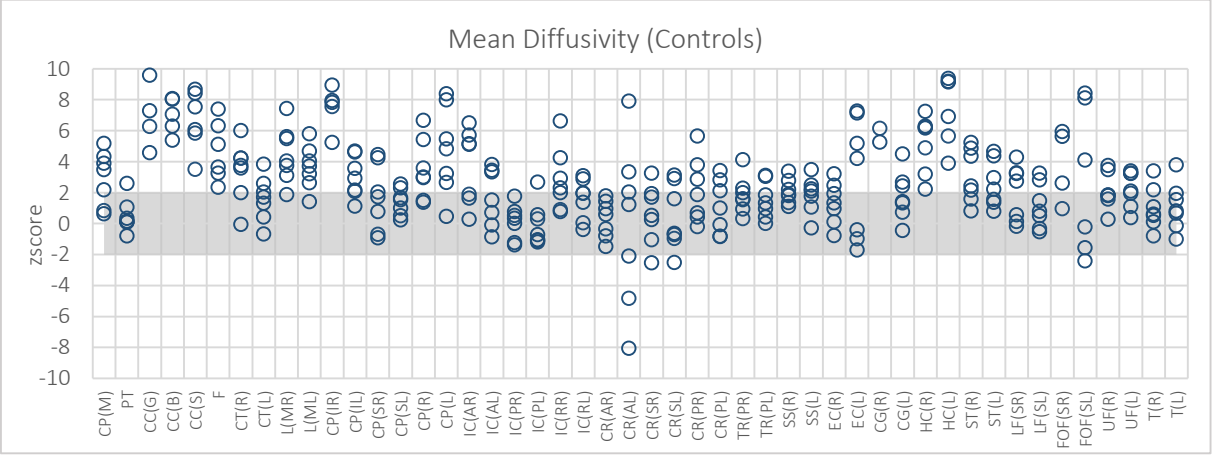

Figure 5.3 – MD maps. Z score of the Controls from Centre 3 – WBIC – when expressed in terms of controls from Centre 2 - Finland.

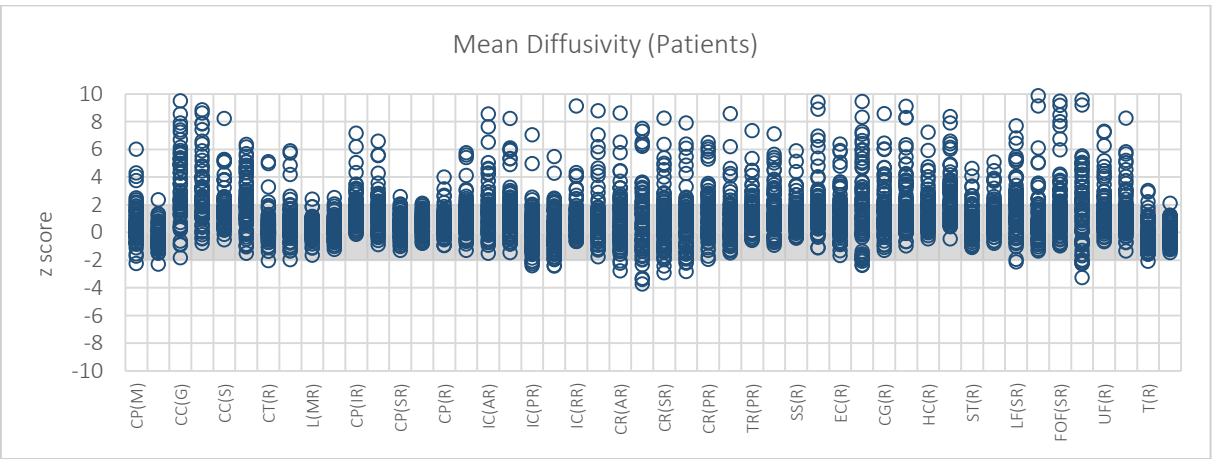

Figure 5.4 – MD maps. Z score of the Patients for the Mean Diffusivity data.

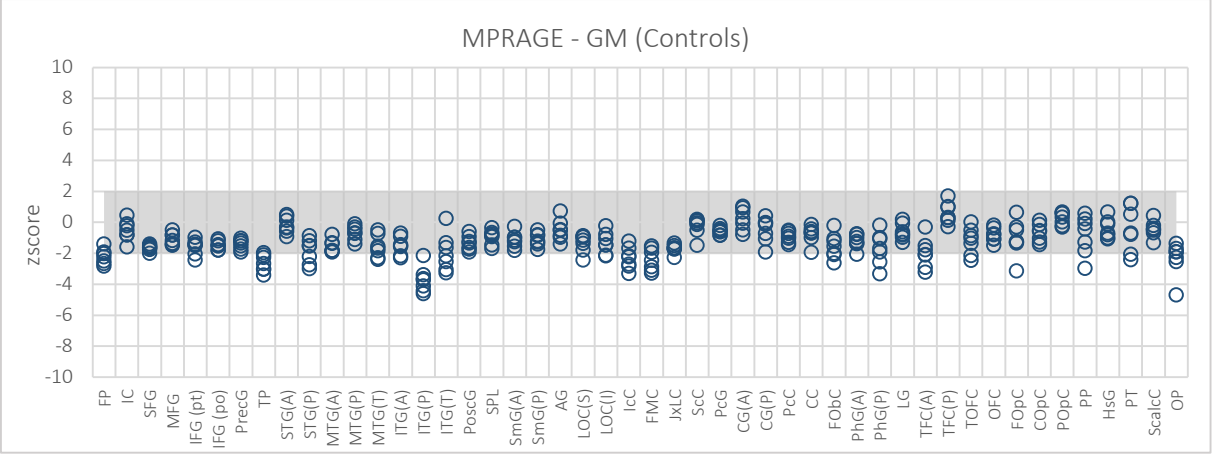

Figure 5.5 – GM maps. Z score of the Controls from Centre 3 – WBIC – when expressed in terms of controls from Centre 2 – Finland.

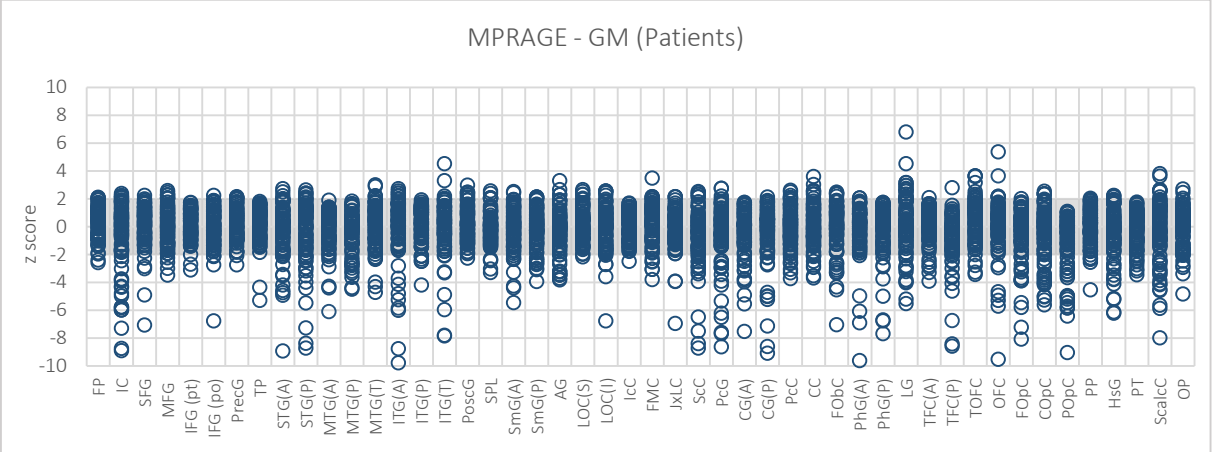

Figure 5.6 – GM maps. Z score of the Patients when expressed in terms of a baseline defined by the controls from the two centres under study.

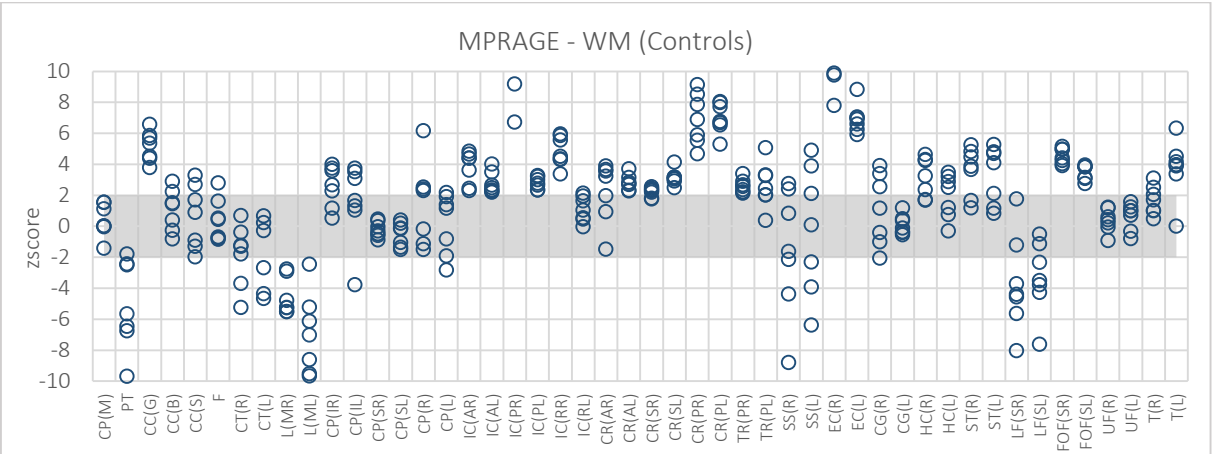

Figure 5.7 – WM maps. Z score of the Controls from Centre 3 – WBIC – when expressed in terms of controls from Centre 2 - Finland.

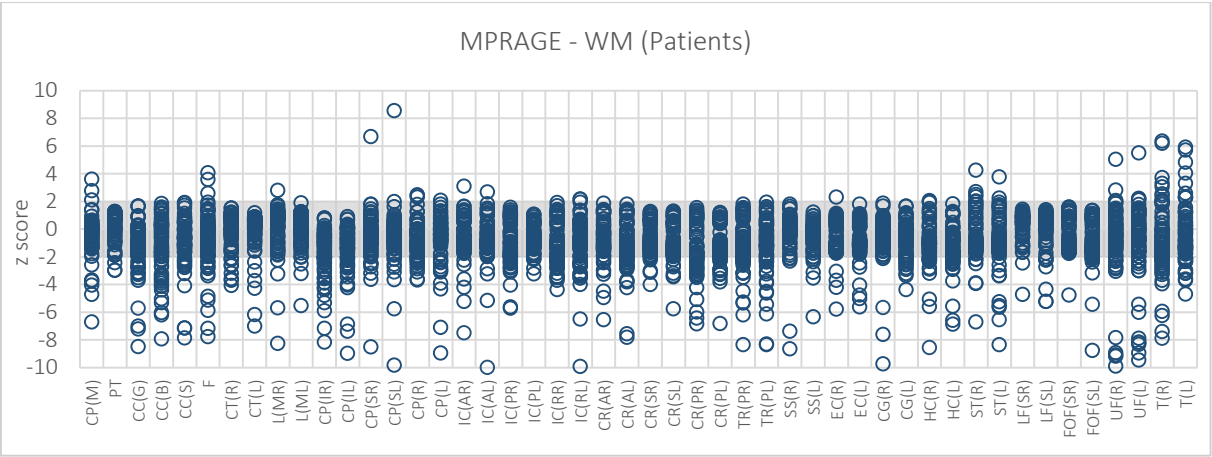

Figure 5.8 – WM maps. Z score of the Patients when expressed in terms of a baseline defined by the controls from the two centres under study.

(a.2) After Correction

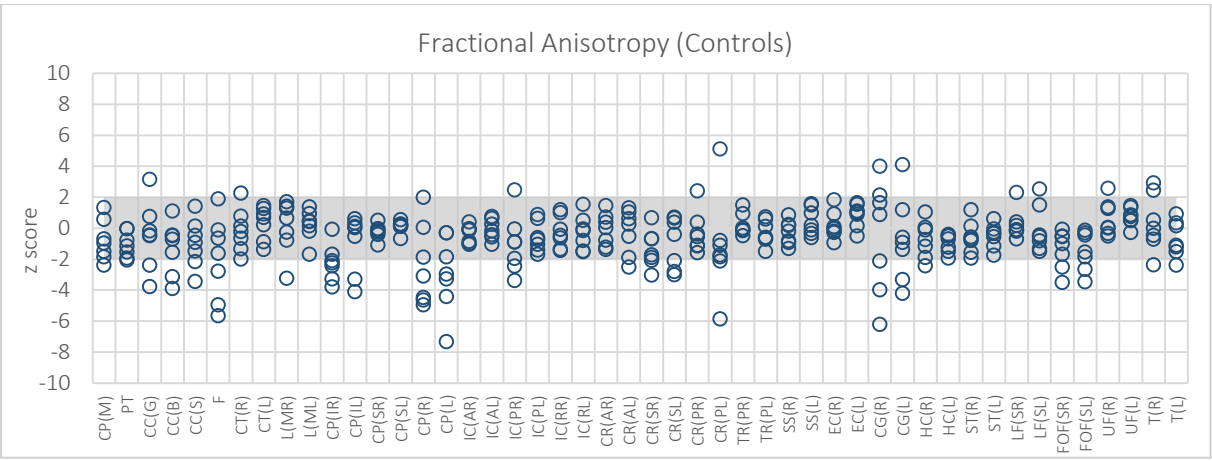

Figure 5.9 – FA maps. Z score of the Controls from Centre 3 – WBIC – when expressed in terms of controls from Centre 2 - Finland.

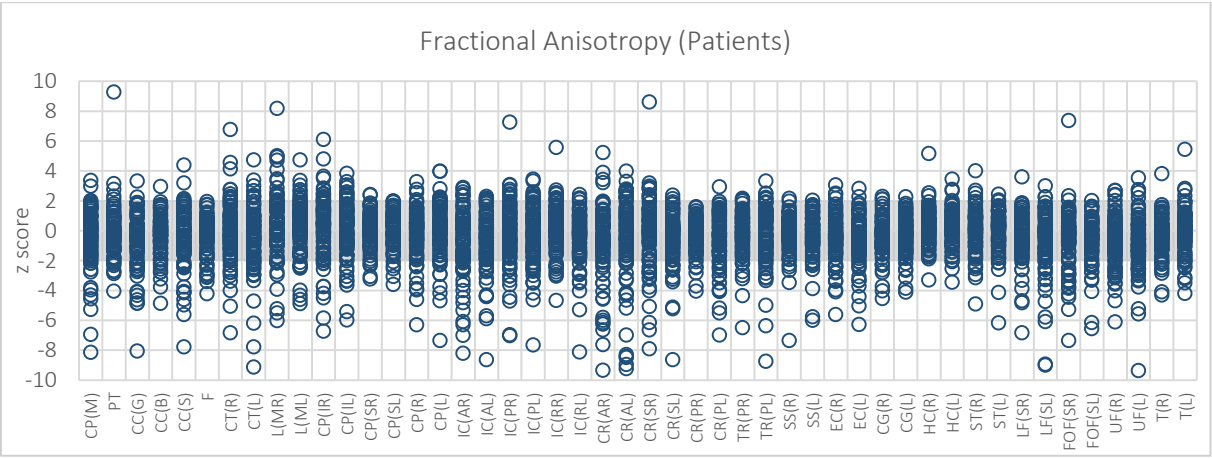

Figure 5.10 – FA maps. Z score of the Patients when expressed in terms of a baseline defined by the controls from the two centres under study, after correction.

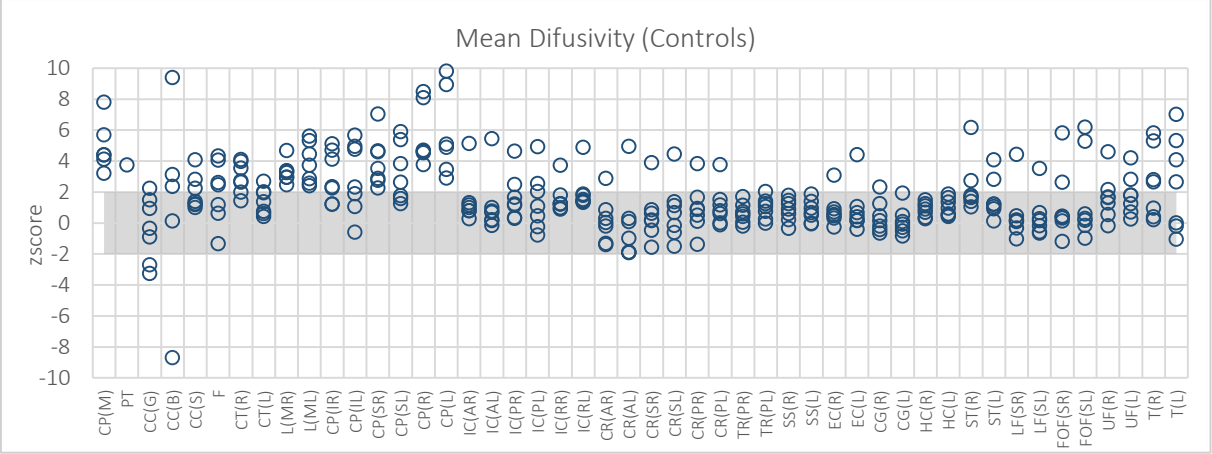

Figure 5.11 – MD maps. Z score of the Controls from Centre 3 – WBIC – when expressed in terms of controls from Centre 2 - Finland.

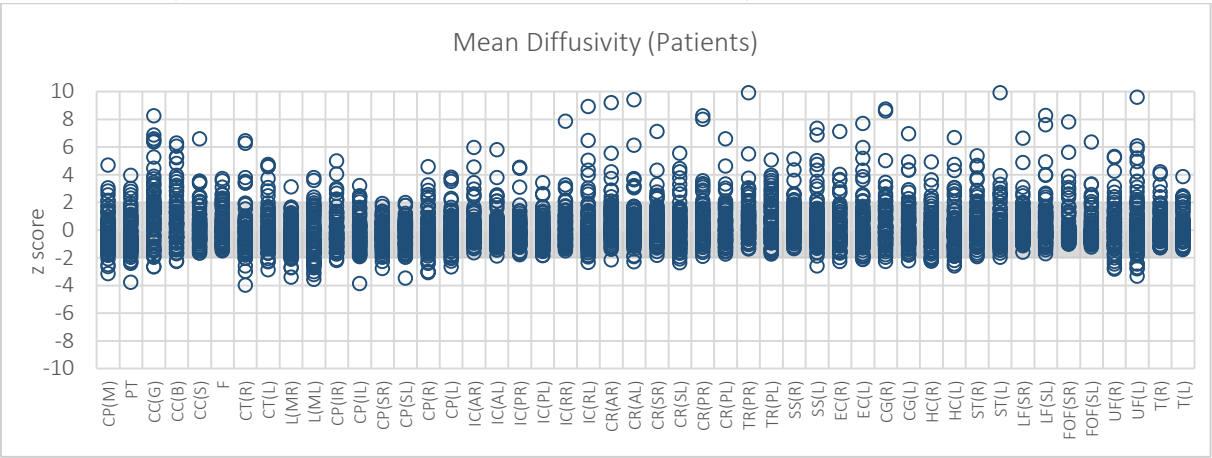

Figure 5.12 – MD maps. Z score of the Patients when expressed in terms of a baseline defined by the controls from the two centres under study.

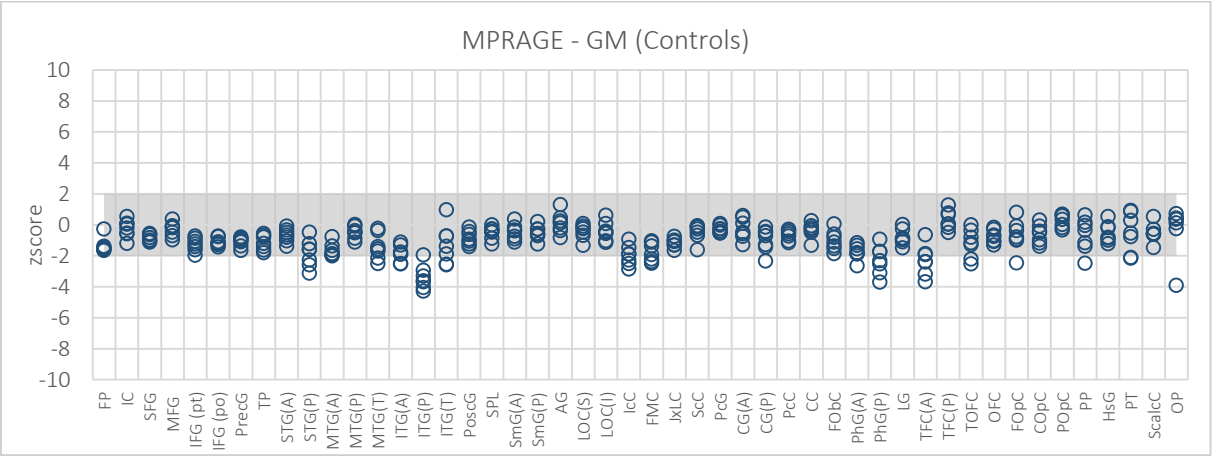

Figure 5.13 – GM maps. Z score of the Controls from Centre 3 – WBIC – when expressed in terms of controls from Centre 2 - Finland. The correction was performed using a FWHM=2mm.

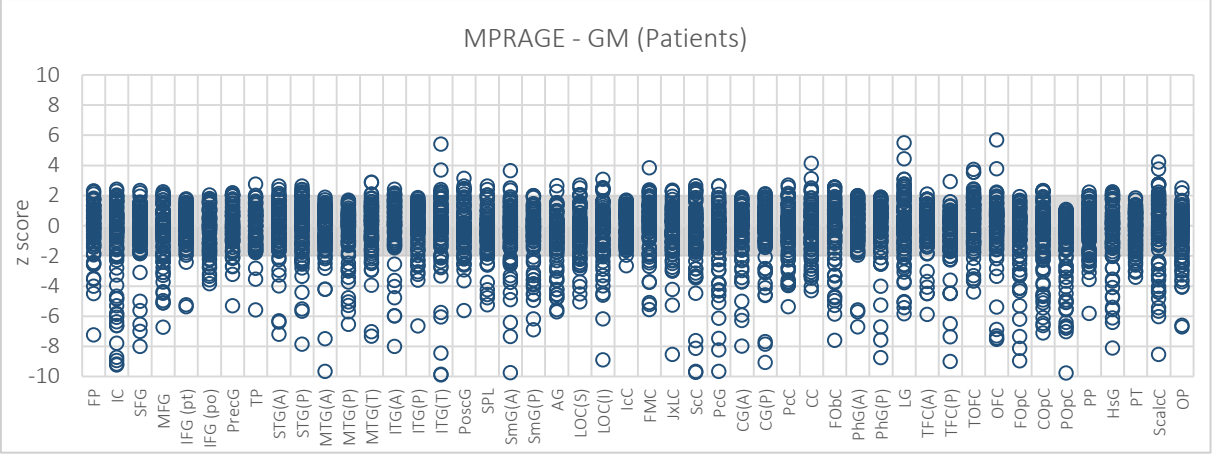

Figure 5.14 – GM maps. Z score of the Patients when expressed in terms of a baseline defined by the controls from the two centres under study. The FWHM used was 2mm.

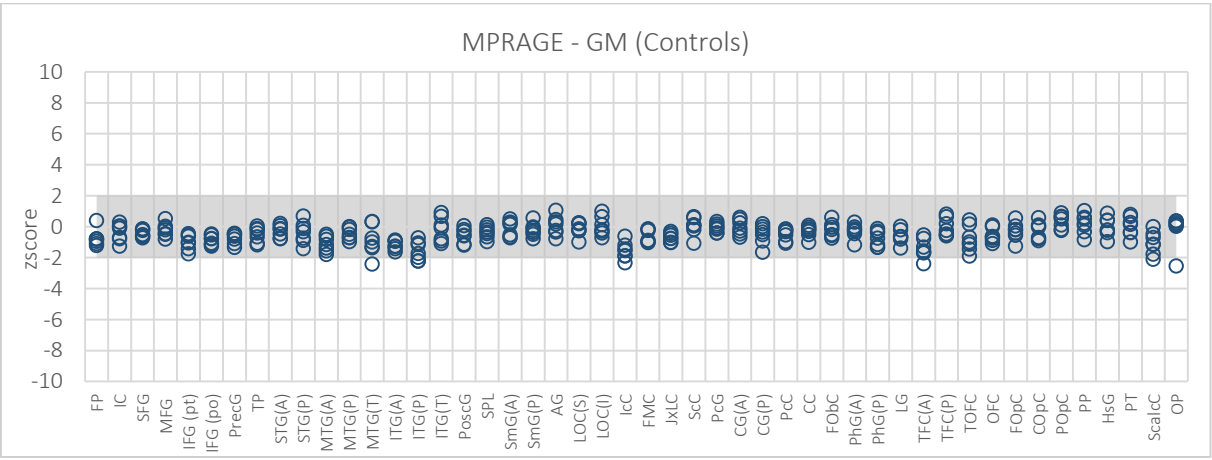

Figure 5.15 – GM maps. Z score of the Controls from Centre 3 – WBIC – when expressed in terms of controls from Centre 2 - Finland. The correction applied a FWHM=8mm.

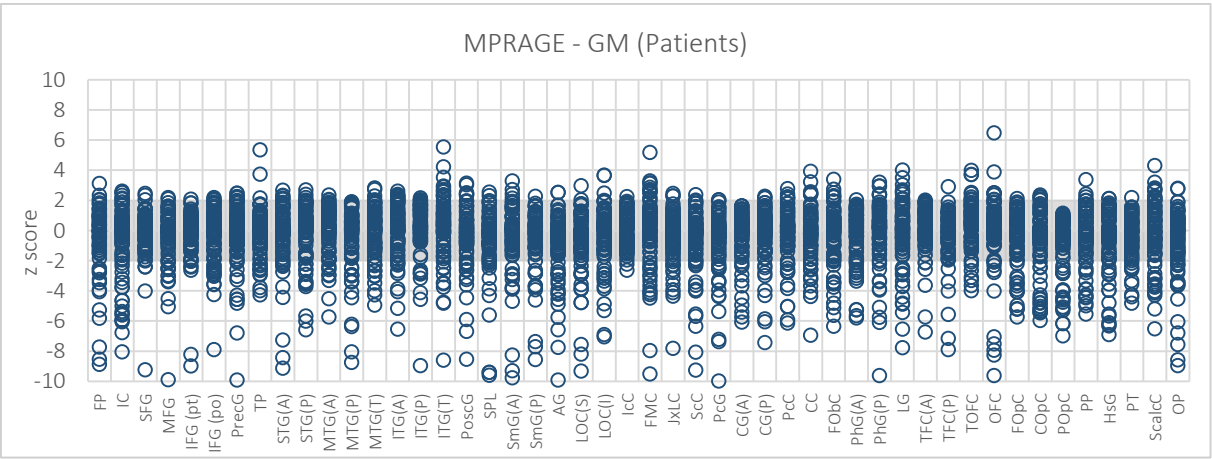

Figure 5.16 – GM maps. Z score of the Patients when expressed in terms of a baseline defined by the controls from the two centres under study. The factor used in spatial filtering was FWHM= 8mm.

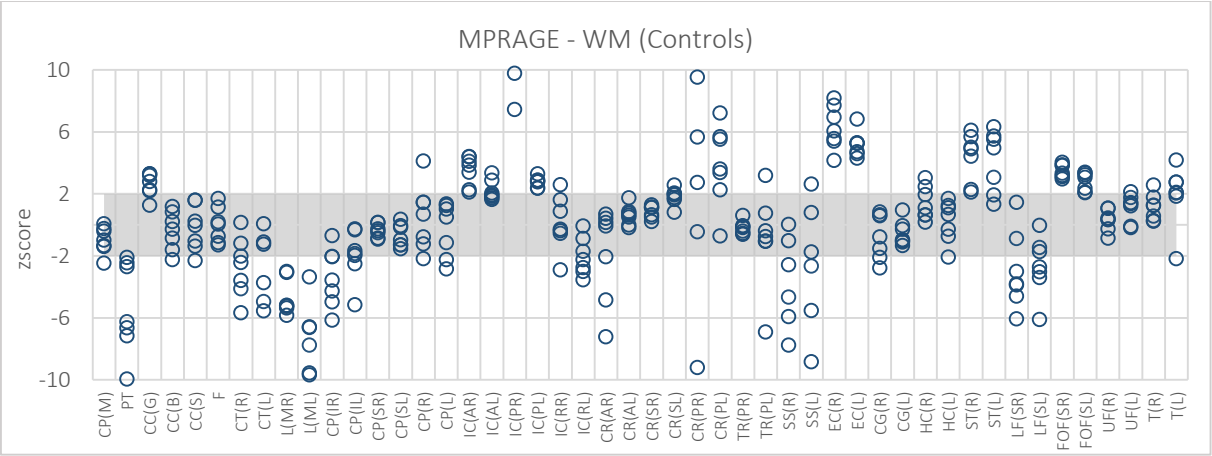

Figure 5.17 – WM maps. Z score of the Controls from Centre 3 – WBIC – when expressed in terms of controls from Centre 2 - Finland. The FWHM used was equal to 2 mm.

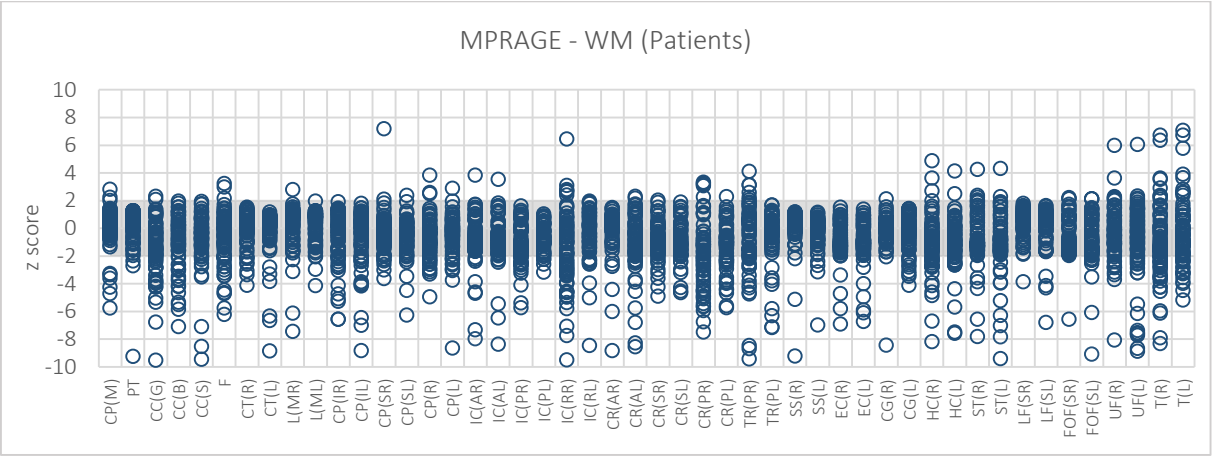

Figure 5.18 – WM maps. Z score of the Patients when expressed in terms of a baseline defined by the controls from the two centres under study. The correction was performed using a FWHM=2mm.

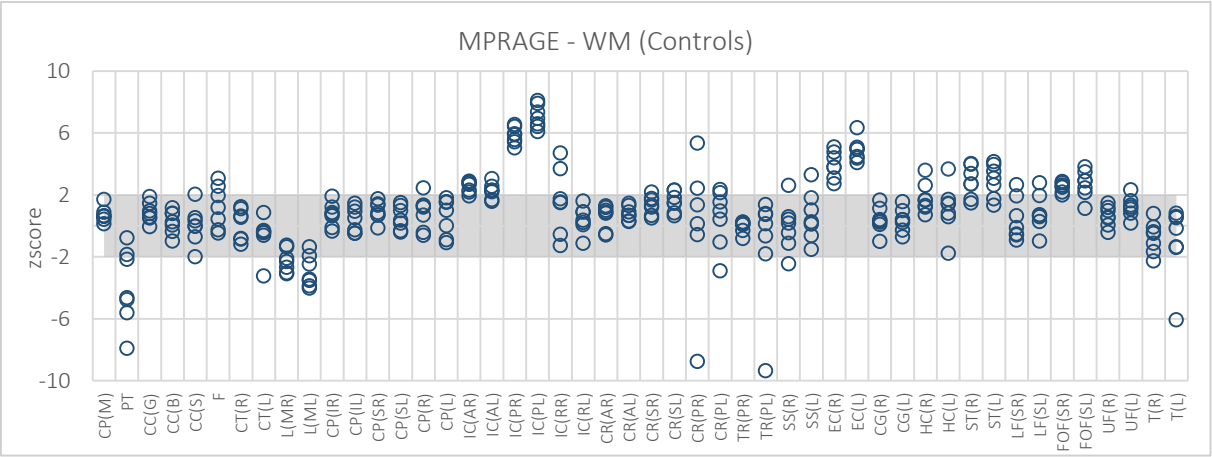

Figure 5.19 – WM maps. Z score of the Controls from Centre 3 – WBIC – when expressed in terms of controls from Centre 2 - Finland. The correction was performed using a FWHM=8mm.

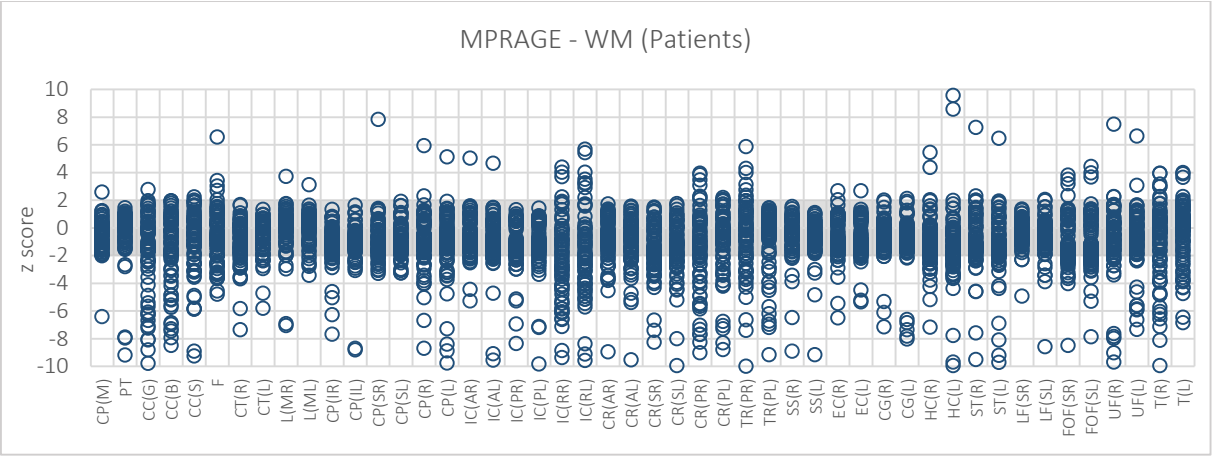

Figure 5.20 – WM maps. Z score of the Patients when expressed in terms of a baseline defined by the controls from the two centres under study. The FWHM value used was 8mm.

(a.3) Correction Evaluation

In order to obtain a quantitative measure of the effectiveness of the correction model applied, the number of Patients for which differences among all group were significant for each ROI was counted before and after the correction. In this way, the Figures 5.21 to 5.28 express the number of patients significantly different from the all patients in the group (number of patients localized outside of the confidence interval [-2 2]) before and after the correction for the four types of images present in this study. Note that the number of subjects present in the Controls group (11 subjects) differs from the number of subjects present in Patients group where the number of subjects is 79.

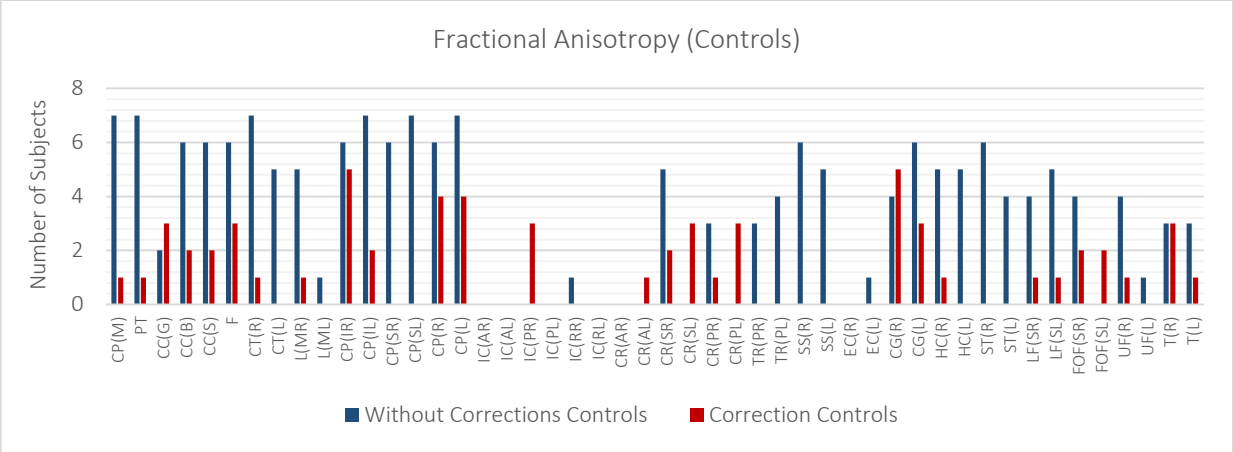

Figure 5.21 – Number of controls significantly different for each ROI before and after correction for FA maps.

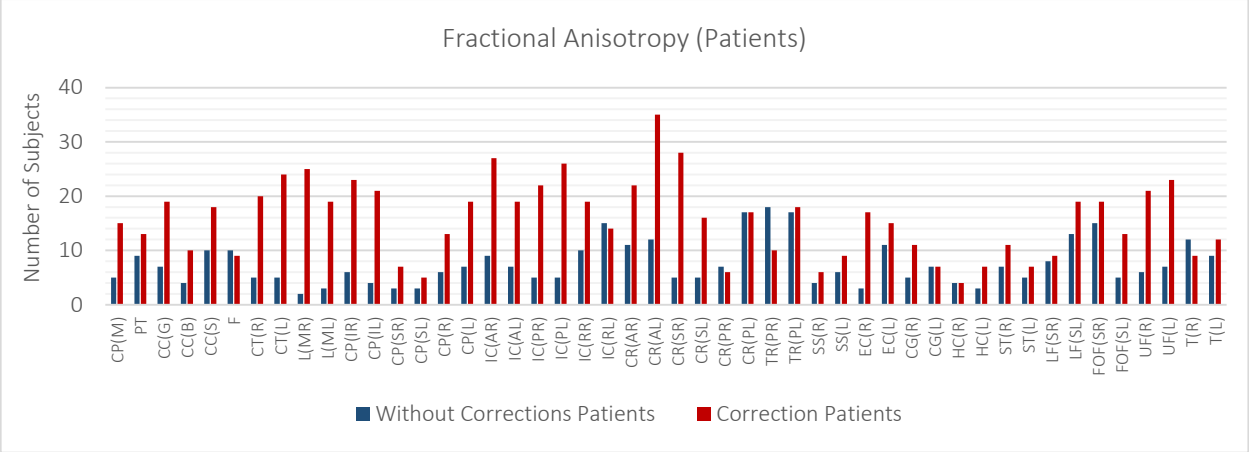

Figure 5.22 – Number of patients significantly different for each ROI before and after correction for FA maps.

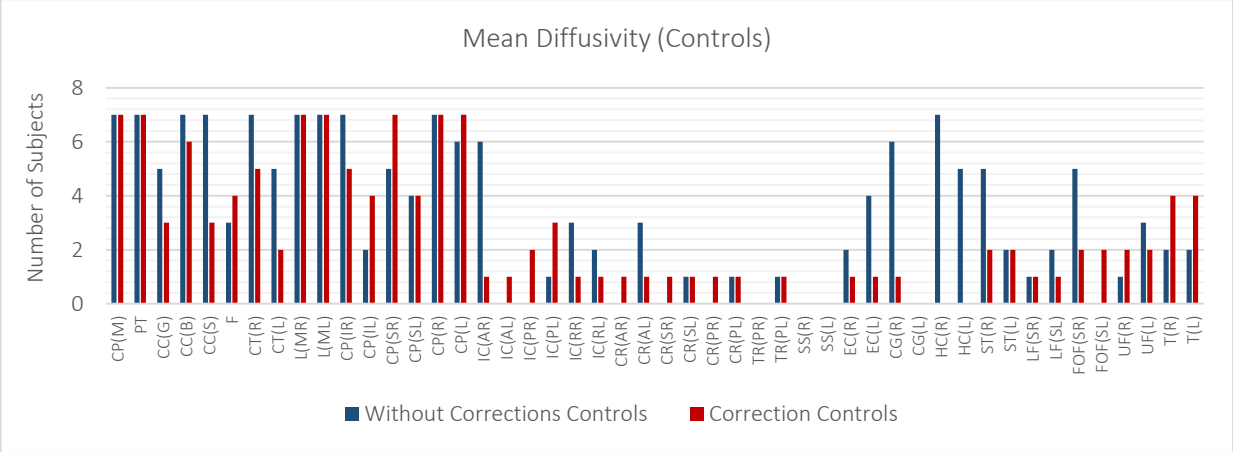

Figure 5.23 – Number of subjects significantly different for each ROI before and after correction for MD maps.

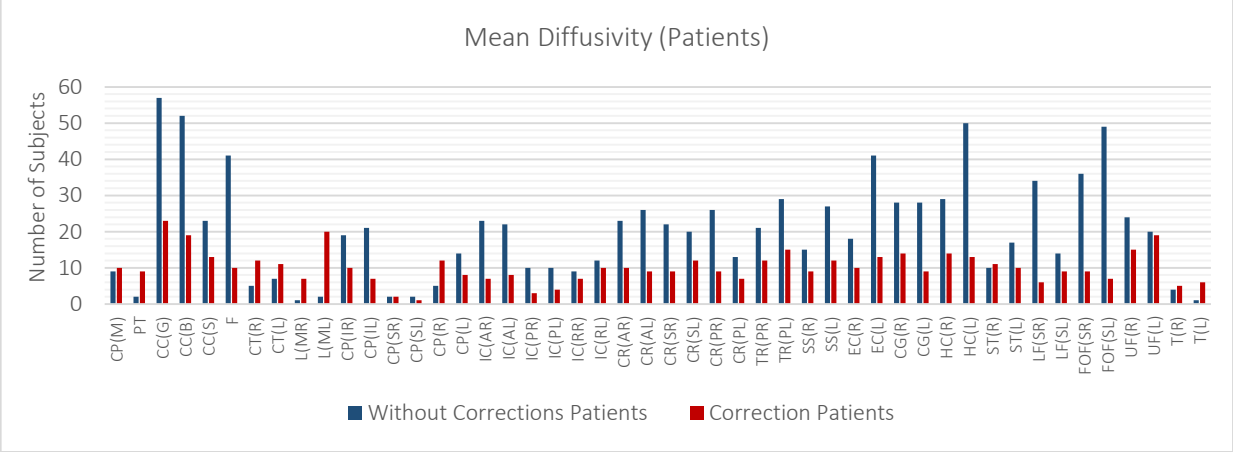

Figure 5.24 – Number of patients significantly different for each ROI before and after correction for MD maps.

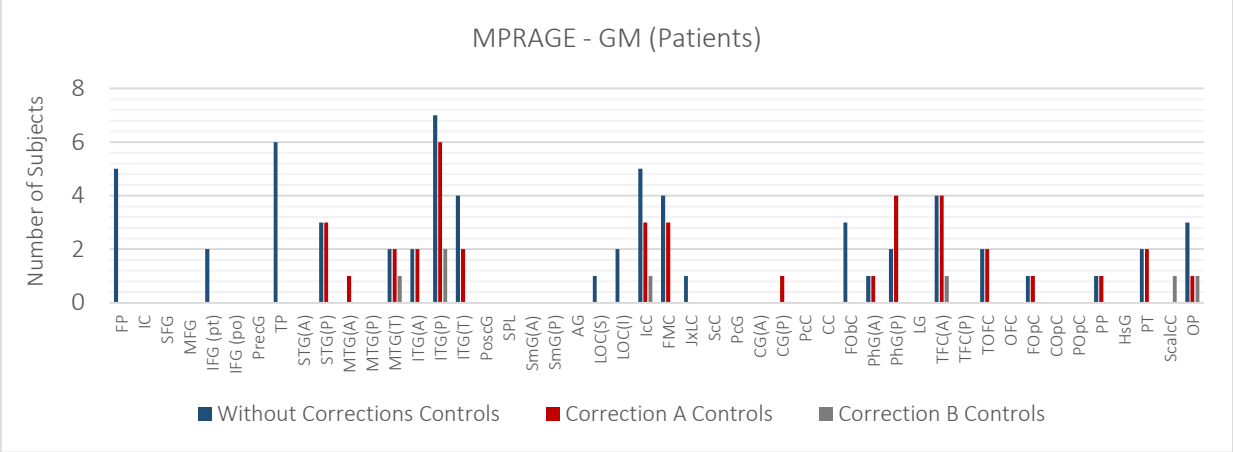

Figure 5.25 – Number of controls significantly different for each ROI before and after correction, Considering to the two types of correction applied, for GM maps.

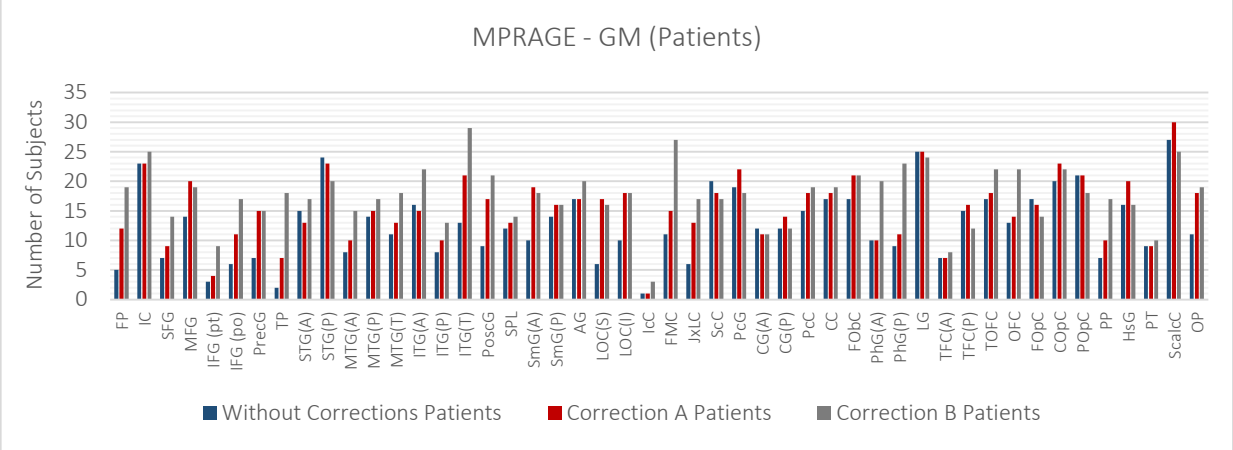

Figure 5.26 – Number of patients significantly different for each ROI before and after correction, considering the correction A and B applied, for GM maps.

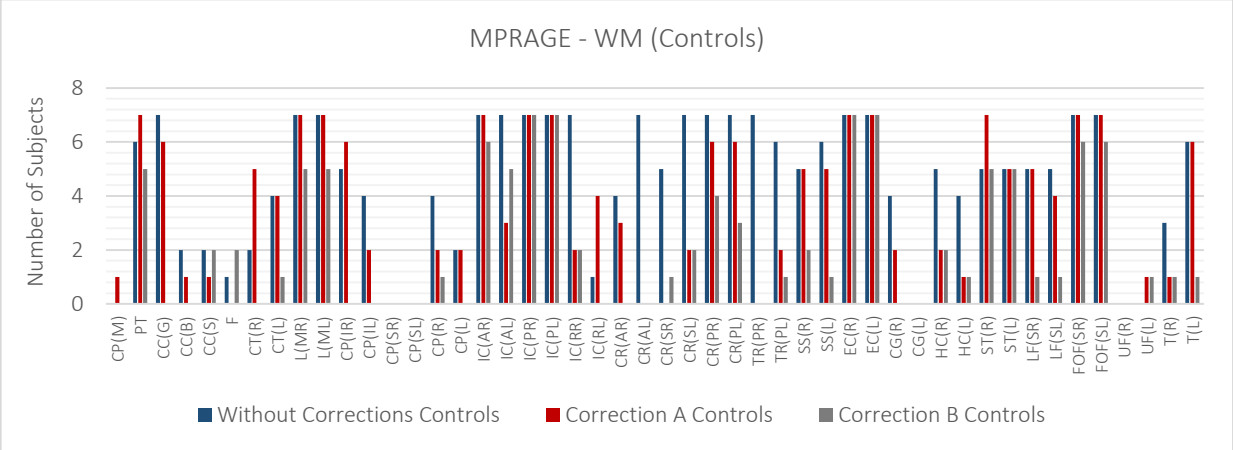

Figure 5.27 – Number of controls significantly different for each ROI before and after correction, considering the correction A and B applied, for WM maps.

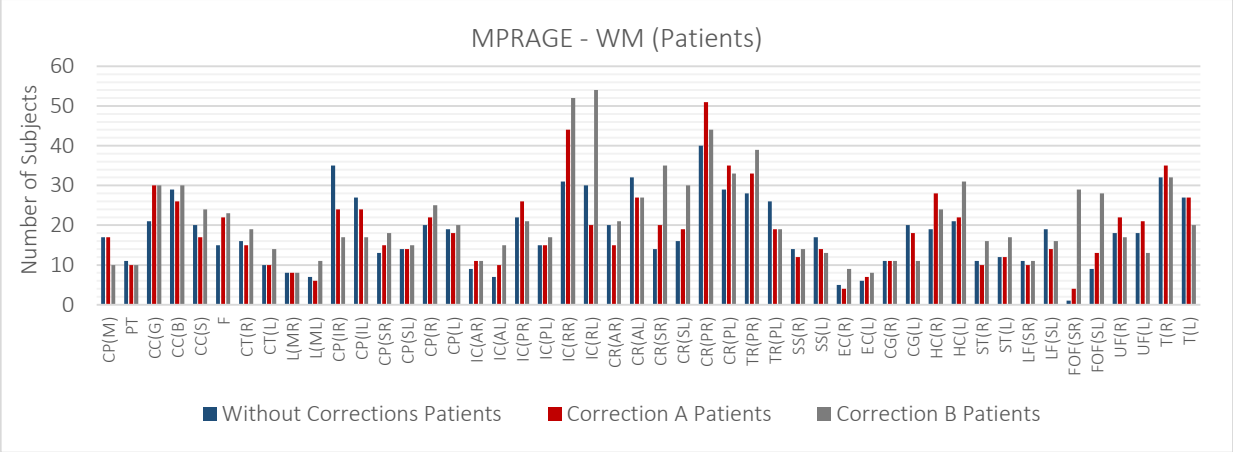

Figure 5.28 – Number of patients significantly different for each ROI before and after correction, considering the correction A and B applied, for WM maps.

(b) Group Analysis

The results of the group analysis are presented using colour-maps, in which the significant voxels (p-value <0.05) are represented with a colour variation between red and yellow. However, for an easier assessment of the results, the number of voxels was calculated with a significant weight for the statistical analysis (Table 5.1).

No significant differences were found for WM images between groups. Thus, these images were not included in this section. In the same way, the GM images obtained after applying the Correction B did not show significant differences between groups, and so the images were not included either.

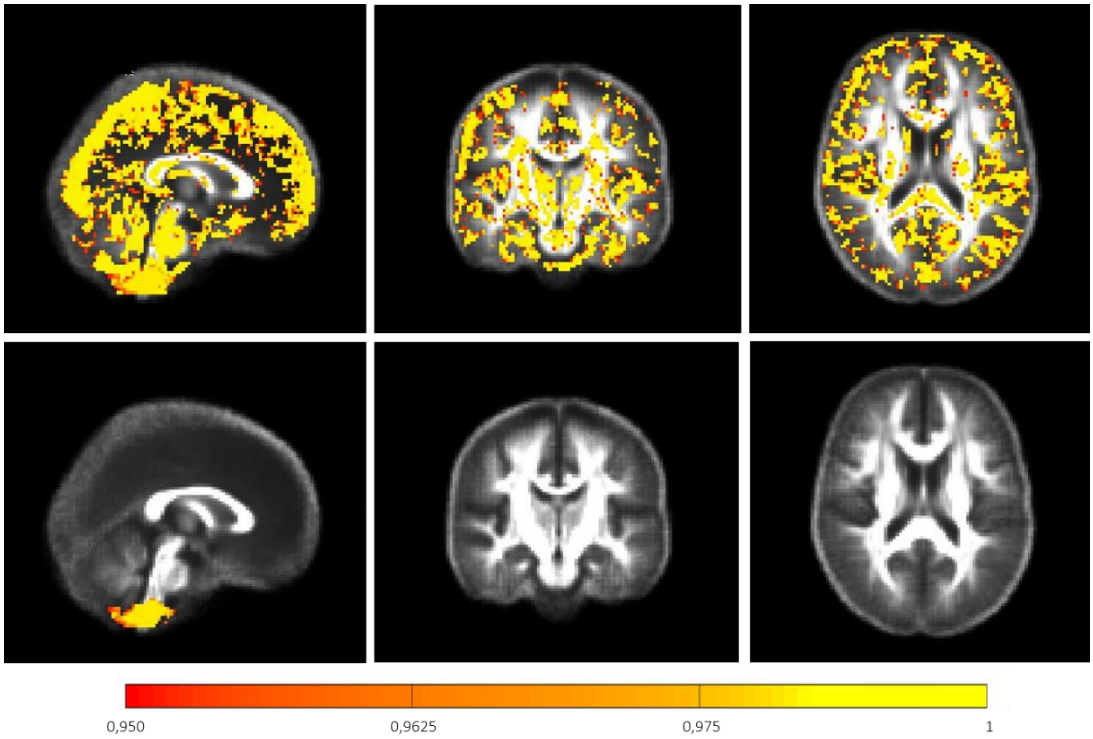

Figure 5.29 – Evaluation of differences for FA maps displaying voxels with significant differences between centres. The top row corresponds to the FA slices before any correction, and the bottom row corresponds to the FA slices after correction.

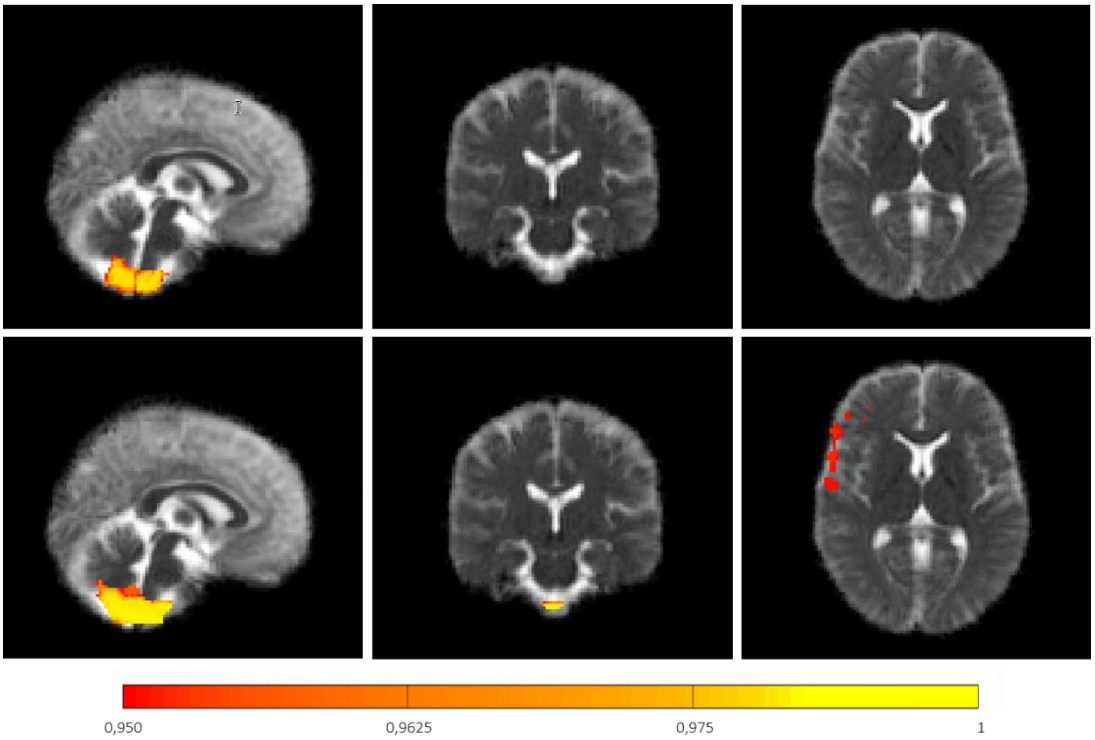

Figure 5.30 – Evaluation of differences for MD maps. Voxels with significant differences between centres. The top row corresponds to the MD slices before any correction, and the bottom row corresponds to the MD slices after correction.

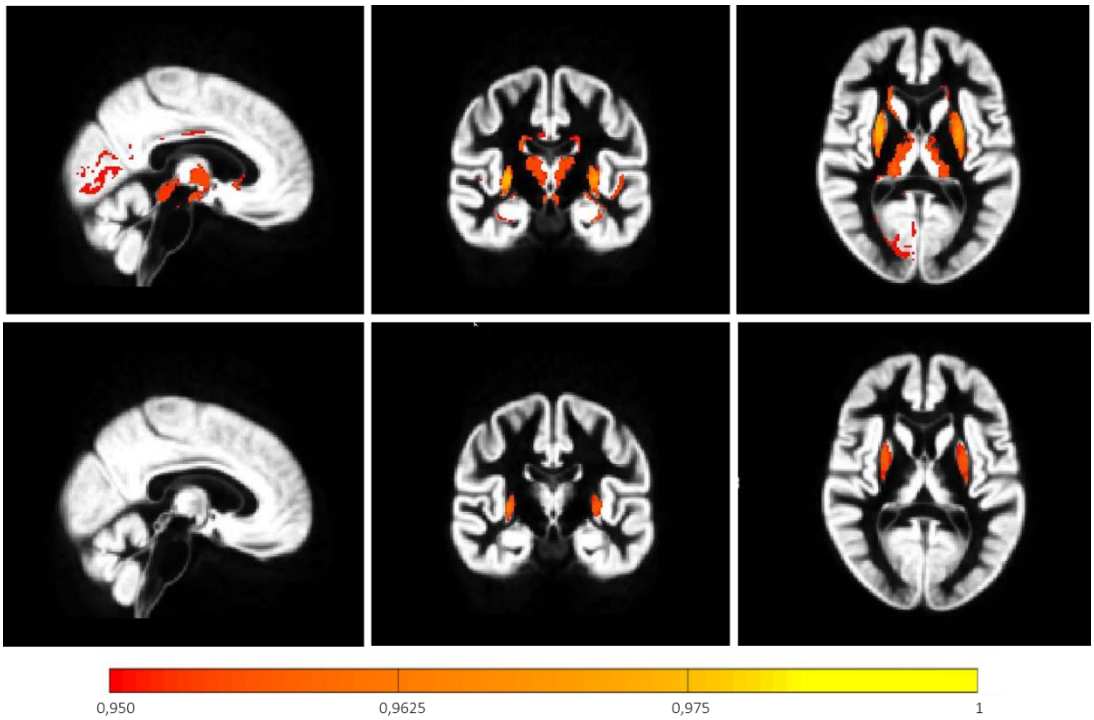

Figure 5.31 – Evaluation of differences for GM maps. Voxels with significant differences between centres. The top row corresponds to the GM slices before any correction, and the bottom row correspond to the GM slices after correction A.

Table 5.3 – Summary of statistical analysis. The results presented in this table correspond to the number of voxels where significant differences were detected. The significance was defined using the p-value, namely p-value<0.05. Given that the results of the test implemented are corrected, the intervals presented correspond to 1-(p-value).

		1-(p-value)		
		[0.85; 0.9[	[0.9; 0.95[	[0.95; 1]
FA	Before Correction	11556	11412	89797
	Correction	644	1067	2762
MD	Before Correction	579	1642	1982
	Correction	30632	22864	4594
GM	Before Correction	34386	42798	1508
	Correction A	28046	7793	0
	Correction B	2766	0	0
WM	Before Correction	0	0	0
	Correction A	0	0	0
	Correction B	0	0	0

5.4. Discussion

5.4.1. Spatial Filtering Method

Although most studies produced on the topic of multicentre analysis using MRI suggest a set of preventive measures to avoid as much as possible the variability introduced by variants such as the scanner, one of the most common approach to correct artefacts and sources of error which could compromise the reliability of the study is spatial filtering by smoothing. With this in mind, one of the approaches tested to correct the CENTER-TBI data used an isotropic smoothing kernel to correct artefacts in the scans as well as the variability introduced by hardware issues.

This approach, denominated as approach 1 in this dissertation, was applied to DTI invariant measures and MPRAGE scans, which were segmented in GM and WM.

Furthermore, this method should be able to increase the differences across patients, because the patients' scans should present a high level of variability since the diseases manifestations are heterogenic among patients, and not lead to full homogeneity in the Controls' scans, since the variability introduced by the scanner is not the only source of differences among scans. Besides, if this approach eliminates the variability among patients, it is likely that the main features which allow patients' diagnosis and TBIs to be studied are being removed as well. This is the main issue of the correction models. In fact, it is very important to test the model in order to verify its effectiveness in theory, but it is also important to guarantee its applicability to real studies, without compromising further analysis. In other words, a good correction method should be the one that removes differences between controls, but preserves differences between patients.

Considering the results obtained for FA images (Figures 5.1, 5.2, 5.21, 5.22), it is possible to infer that the model successfully removed the variability among Controls, without homogenising the subsample_Patients. In fact, with a small

number of exceptions, the number of patients significantly different in each ROI was lower after the correction approach was applied (Figure 5.21).

Similarly, for MPRAGE scans, for either GM or WM, the models seemed to be successful as well, with better results when compared to FA. However, it is important to notice that the Correction B – using a smoothing kernel with FWHM=8mm seemed to be more effective to correct the data (Figures to 5.25 to 5.28). In fact, in the particular case of GM images, this approach was able to correct almost all the subjects (this fact is possible to see analysing Figure 5.15). In contrast, the results obtained using Correction A suggest that the value used for FWHM is not enough to remove the artefacts in the scan. Note that, with this analysis it was proved that the FWHM=8 mm is not a value too high, which could cause a blurring effect in the image and render it useless for patient diagnosis or other studies performed using the MRI scans.

The only exception to the positive results obtained using this method were the MD images, in which the differences among Patients seemed to be reduced. However, since these images have a smaller level of variation when compared with the other types of images analysed (FA, GM and WM), the correction model could be less successful when compared to other images with higher values of COV.

The performance of the model was not equally efficient for the intra-scanner correction. The results of this correction can be found in Appendix C. In this case the correction is not efficient for all ROIs, where in some particular cases, the number of patients was significantly different, and lower than before correction. This fact seems to suggest that the correction could be prejudicial for the intra-centre correction since the patients are homogenised, losing particular features inherent to the disease, which are crucial for the diagnosis and to distinguish them from the healthy subjects. This effect is significant for MD scans from Finland, in which the correction was inefficient. These results can be caused by the reduced number of Controls in each centre, which might be insufficient to establish a baseline, i.e., 7 subjects in each centre is not enough to establish a baseline sufficiently robust in statistical terms. Nevertheless, reaching an accurate conclusion about the performance of the model in the context of intra-scanner correction would only be possible by studying the effect of the model on the scans from controls within the centre, before and after correction.

(b) Group Analysis

Although the group analysis has been performed using patients, considering that the t-test performed took into account all the features which could contribute to an increase in variability, the only factor under focus in this test was the variability introduced by the centre. Therefore, in theory, the number of voxels significantly different across centres should be lower after correction.

The results for the group analysis suggested a high effectiveness of the model in the correction of FA and GM maps (Figures 5.29 and 5.30). However, the results for MD images showed a slight increase of the number of voxels with significant differences among centres, after correction. These results were consistent with the previous study for individual subject analysis, in which the number of Patients with significant differences relative to the baseline increased after correction.

Thereby, as concluded before, the correction applied seems to be a good choice for FA, GM and WM maps, however the results for MD maps revealed issues.

5.4.2. Limitations of the Method and Future Improvements

Despite of the good results of the method, this correction approach has limitations.

The first limitation of the model is the number of controls used to establish the baseline. Considering the number of Patients present in this study (79 subjects), the number of controls is not sufficient to ensure a robust statistical test. Besides, regarding the features of the sample (age, gender, centre), a controls' sample with a bigger size would be beneficial, since the present group shows a low level of heterogeneity in the subjects features. On the other hand, to guarantee well matched groups (Patients and Controls) the subjects present in Control group have a high level of deviation in age, in order to achieve the medium age of 35 years old. In spite of the issues behind the number of controls used, it was not possible include more controls in this study because the number of controls available in CENTER-TBI is only 9 per centre. Besides, to ensure the groups were well matched, two controls were excluded, and only 7 controls were used per centre.

Another issue behind this method is the fact of the spatial filtering using a Gaussian kernel increases the partial volume and can remove biological features of the images, jeopardize the viability of biological interpretations after correction.

Finally, using the z-scores to express the Patients in terms of Controls seems to be a good way to evaluate and ensure the study reliability, since the variability introduced by sources of error could be detect and removed. However, using this approach the biological meaning of the voxels values were lost, and it is impossible to analyse the images using qualitative and visual methods. After this approach was applied ,the Patients were always analysed in terms of the baseline established by the controls; thus it is impossible to compare the values of the Patients from the two centres analysed with a third centre, without established a new baseline, in which the controls from a third centre are included. In this way, when a new comparison and/or analysis between centres are required, the method described in this section will need to be applied from the beginning, which requires a loss of time and efforts not justified.

To improve the method described in this section and overcome the limitations mentioned above, it would be good to increase the number of subjects present in the Controls group. This fact would be useful to ensure the robustness of the statistical tests performed. On the other hand, it would be interesting to develop an algorithm, which applied a personalized Gaussian kernel with a specific FWHM to the scanners from a particular centre. This algorithm would be able to test which specific FWHM for each centre could reduce the variability across centres without improve unnecessarily the partial volume of the images. This improvement was not tested regarding the number of combinations required to test different levels of smoothing from each centre and their effect in the variability across scanners.

Finally, in order to ensure the effectiveness of the method suggested, it would be useful apply it to a third centre (if it possible with a different scanner, e.g. Phillips) to prove the results achieved.

5.5. CONCLUSION

The results suggest a good performance of the model. Besides its simplicity, it provides an easy and quick way to correct the data and achieve the reliability needed in multicentre studies. However, using isotropic smoothing kernels can significantly increase partial volume or voxel averaging artefacts, and this model could therefore adversely affect the true diffusion properties of the underlying fibre tissue. In this way, it is important to understand if the study could be compromised by this issue.

The performance of the model is not so good for an intra-scanner correction, as there seems to be a tendency of the model to homogenise the patients and the controls, and a tendency to remove important features of patients which could be useful to distinguish them and to reach an accurate diagnosis.

Regarding the results of group analysis, the model was efficient in removing the variability introduced by the noise from the scanner, did not allowed to distinguish the groups from the different centres.

With this fact in mind, it is possible to conclude that the model is highly efficient in the context of group analysis, and seems to perform well when used to correct the scans from different centres, allowing to accurately proceed to an analysis of each subject independently. However, in the context of intra-scanner correction, the model seems to remove the particular features of the patients and to homogenise the dataset.

Chapter 6

Regression Models

6.1. Introduction

Kostro et al., (Kostro et al., 2014), explored the use of regressions to correct the scanner variability, demonstrating the usefulness of regression in the reduction of data variability, without removing the disease effects. For that, a Gaussian Process Regression (a nonlinear and regularized regression method – Equation 3.9) and an unregularized simple linear regression with ordinary least squares solution (OLS – Equation 3.10) were applied to correct the variability of the *i* scanner. Kostro et al. used all the covariates defined with equal relevance for the model. However, Kostro et al., only applied the model on context of voxel based analysis.

Considering the problem behind this project and the different types of MRI images acquired, the model mentioned above was modified in order to be possible apply it to the CENTER-TBI data. Therefore, regarding the disease effects at the TBI patients, namely in cases of severe TBI, in which the brain is deformed and the segmentation is not efficient, the voxel based analysis is difficult to apply. Thus, to overcome the limitations of voxel based analysis in severe TBI the model was also adapted to perform an ROI analysis. In the present chapter, an optimization of the model was performed iteratively and independently for each ROI, in order to achieve the best model for each of them. Note that the model was constructed with the possibility of the application of this optimization to a voxel-based analysis. This variation of the model was not applied since the computation time required not justify the improvements achieved.

In this chapter a different test was applied to evaluate the model performance. Despite of this new approach used to evaluate the model did not contribute for the consistency and comparability of the methods used across chapters, it was not possible to use the approaches adopted in Chapter 5. The main reason for that was the insufficient number of controls available in the dataset. In fact to ensure the robustness of the model developed to correct the data, all the controls available were used for the model calibration. Therefore, these controls cannot be used for model evaluation, in order to avoid the introduction of bias in the method used to evaluate the performance of the model.

Different kind of approaches can be used to quantify the variability among scanners. Examples of that are General Linear Models (GLM) and classifiers, such as Support Vector Machine (SVM). In the first case, each scanner used is considered as a variable of the model. The beta coefficients estimated in the model are a measure of the variability brought by the covariate with which it is associated. It is important to note that other variables can be introduced in the model, in order to absorb the variability introduced by other sources, such as the variability across subjects resulting of age and gender.

The SVM does not produce a direct quantification of variability. However, using a classifier it is possible to determine how significant the level of variability introduced by the scanners is. An accuracy of classification higher than 80% means that the variability introduced by the scanner is too significant and, consequently, the scans are not comparable (Kostro et al., 2014). In the absence of any inter-scanner effects, it can be assumed that the scanners under analysis produce similar and comparable images, and the classifier would not distinguish the scanners above chance level (50%).

6.2. METHODOLOGY

6.2.1. Group definition and Data pre-processing

In this section, two groups were created using the subjects from the Centre 2 and Centre 3, described in Chapter 4. The first group corresponds to the Controls from Centre 2 and Centre 3 joined (total of 25 subjects). The second group was formed by the Patients from both centres, which corresponds to a total of 79 subjects. The subjects' characteristics could be found in the Chapter 4.

The parameters and sequences adopted to acquire the MRI scans were described at Chapter 4.

Furthermore, the pre-processing data pipeline adopted in this study is detailed in the previous chapter (section 4.2.1 of this document).

As before, for each group two DTI invariant measures were analysed –FA, MD – and GM and WM for the images obtained from T1 segmentation.

6.2.2. Regression Model: Ordinary Least Squares (OLS)

The first step of this model is the construction of the design matrix, which will define the covariates taken into account in the regression model, and their values for each subject. The design matrix used in this study, denominated as X matrix (Equation 6.1), included the age, gender, Centre 2 and Centre 3 as covariates. The last three covariates were code {0 1}, while the age was scaled to the interval [0 1].

The Ordinary Least Squares model (OLS) for each voxel was then defined in matrix form (Equation 6.3), y is the vector of the observations for each scan at the voxel v, the ε is the residual error with zero mean and variance σ^2 and β (Equation 6.4) corresponds to the matrix of the parameters estimated by the model, which were used to correct the voxel value using the pseudo-inverse transformation (Equation 6.5).

Note that the Training set is only composed by Controls in order to avoid confounding disease with non-disease effects during the estimation of the regression parameters. On the order hand, given the reduced number of controls in this dataset, the Test group is only composed by patients. In this way, only the patients' data were corrected and the information from de controls was lost for subsequent statistical analyses. This fact is the principal issue of this model, since it will not be possible to use the controls for further studies or to do an extensive evaluation of this model. Given the reduced number of controls available in each centre[11], it was not possible to use the same metrics, applied in Chapter 4,

[11] The CENTER-TBI protocol has established that only nine subjects are analysed in each centre, in order to be used as reference/controls.

to evaluate the model performance. In other words, splitting the controls in two groups to be used as model estimation and then as model evaluation was not considered, since in that situation the model would not be able to correct efficiently the data and it would lose the statistical power and the robustness essential to apply this methodology. Another option would be to use part of the data used to estimate the model to evaluate it. However, using the same data to estimate and evaluate the model would introduce bias in the results, making them unviable to be discussed and presented as a solution to the problem under analysis in this dissertation.

$$X = \left[X_0,\ X_{age}, X_{gender}, X_{Centre\ 2}, X_{Centre\ 3}\right] \in R^{N\times 5} \quad \text{Equation 5.3.}$$

$$x^i = \left[1,\ x_{age}{}^i, x_{gender}{}^i, x_{Centre\ 2}{}^i, x_{Centre\ 3}{}^i\right] \in R^5 \quad \text{Equation 5.4.}$$

$$y^v = X\beta^v + \varepsilon\,, y^i = \left[y^{(1,v)}, \dots, y^{(N,v)}\right] \in R^N \quad \text{Equation 5.5.}$$

$$\beta^v = \left[\beta_0{}^v,\ \beta_{age}{}^v, \beta_{gender}{}^v, \beta_{Centre\ 2}{}^v, \beta_{Centre\ 3}{}^v\right] \in R^5 \quad \text{Equation 5.6.}$$

$$y_{corrected}{}^v = y^v - x\widehat{\beta^v} \Leftrightarrow y_{corrected}{}^v = y^v - x(X^TX)^{-1}X^TY \quad \text{Equation 5.7.}$$

To apply the model detailed above the following assumptions were taken into account:

a. The effects in analysis were linear;
b. The covariates were independent;
c. The residuals had a Gaussian distribution;
d. The values of β were equally likely a priori.

This model was applied also to ROI analysis. In that case, the model could also be described by the Equations 6.1 to 6.5, however, the v index should be understood as definition of one ROI.

In the particular case of ROI analysis, the model was optimized by the determination of which estimated βs, in each ROI, had a higher significance in the regression. The significance of the βs were assessed using a stepwise algorithm, in which the R squared adjusted was used to evaluate the significance of the βs token into account in the specific step for the correction of the particular ROI in analysis. Thus, the covariates taken into to account to correct each ROI, were the ones with a relevant meaning (R squared adjusted > 0.7) to correct that specific ROI. This optimization was not applied to the voxel analysis, since the computation time required for the optimization did not justified the improvement obtained in the results[12].

[12] This optimization was only tested for FA maps, in which the time required to calculate the βs was approximately 3 days, in contrast with the 1.5 hours required for the calculation of the betas when a voxel based analysis without optimization. Note that, the AUC achieved for a voxel based analysis optimized, for FA maps, was 0.613. However, the algorithm optimized for voxel based analysis should be tested more carefully for the rest of the images analysed (MD, GM, WM), in order to prove the conclusion mentioned above.

The effectiveness of these three variations of this model were then compared through the AUC obtained by classification of the centre using a linear classifier, calculated for the three situations detailed above.

6.2.3. Support Vector Machine (SVM)

A machine learning method was adopted to evaluate the results achieved. Considering the concepts behind SVM, this classifier was applied to categorize the patients' data in terms of the centre where they came from. The Centre 2 was defined as category A, instead the Centre 3 was defined as category B. The patients' data (the corrected data) was split in two groups – Training (20 subjects) and Test set (39 subjects)[13] – and a design matrix to define the variable in analysis – the centre – was constructed. A dataset with a reduced inter-centre variability would not be efficiently categorized in terms of the centre using a classifier, while using a dataset with a high level of inter-centre variability it would be possible to categorize the data in terms of the centre. With this in mind, the SVM was used as an indirect measure of variability. Therefore, in the classification of the uncorrected data a high level of accuracy is expected (accuracy >80%), whilst it was expected that the better the performance of model correction, the percentage of correctly classified Patients in terms of the centre would be nearest to 50 percent (this value corresponds to the situation in which the classifier is attributing the category randomly).

To categorize the results of the SVM, the following parameters were used:

a. Area under the curve (AUC) – quantified the overall ability of the test to discriminate between two classes, in this particular situation the classes correspond to the two centres. A truly useless test (one no better at identifying true positives than flipping a coin) has an area of 0.5. A perfect test (one that has zero false positives and zero false negatives) has an area of 1.00.
b. Accuracy – is the proportion of true results (both true positives and true negatives) among the total number of cases examined.
c. Sensitivity – fraction of subjects from Centre 2 that the test correctly identified as positive.
d. Specificity – fraction of subjects from Centre 3 that the test correctly identified as negative.
e. Precision – proportion of the true positives against all the positive results
f. F-score – (F1 score) is the harmonic mean of precision and recall, thus was used as a measure of accuracy test.

Although all the measures mentioned above had been taken into to account to characterize the SVM results, the AUC was enough to infer about the efficiency of the correction model applied.

6.3. RESULTS

6.3.1. Regression Model: Ordinary Least Squares (OLS)

As mentioned above, the results of the application of OLS model were not brain images with a direct biological interpretation. In fact, the raw data – voxel values which allow to obtain a brain image with direct biological meaning – is

[13] Not all the patients corrected were used in SVM analysis. The main reason for that was the need of the groups were well matched in terms of the age and gender, which made the groups comparable and without features that could contribute to increase the variability.

transformed by the model in corrected data, which voxel values just can be analysed through comparison among subjects, since their biological meaning was loss during the correction steps.

(c) Support Vector Machine

The following table, Table 6.1, summarises the performance of the SVM for the different type of data and the different type of correction applied.

Table 6.4 – SVM results of the data classification of the results of model correction, in terms of the Centre.

		AUC	Accuracy	Sensitivity	Specificity	Precision	F-score
FA	Before Correction	0.833	0.897	0.667	1.000	1.000	0.800
	OLS: voxel based analysis	0.545	0.744	1.000	0.091	0.737	0.849
	OLS ROI analysis (non-optimized)	0.625	0.513	0.917	0.333	0.379	0.537
	OLS: ROI analysis (optimized)	0.552	0.718	0.929	0.182	0.743	0.825
MD	Before Correction	0.801	0.872	0.964	0.636	0.871	0.915
	OLS: voxel based analysis	0.818	0.897	1.000	0.636	0.875	0.933
	OLS: ROI analysis (non-optimized)	0.591	0.769	1.000	0.182	1.000	0.308
	OLS: ROI analysis (optimized)	0.591	0.769	1.000	0.182	0.757	0.862
GM	Before Correction	0.914	0.925	1.000	0.107	0.901	0.947
	OLS: voxel based analysis	0.839	0.769	1.000	0.679	0.550	0.710
	OLS: ROI analysis (non-optimized)	0.681	0.821	1.000	0.348	1.000	0.533
	OLS: ROI analysis (optimized)	0.625	0.462	1.000	0.250	0.344	0.512
WM	Before Correction	0.902	0.974	1.000	0.964	0.917	0.957
	OLS: voxel based analysis	0.821	0.744	1.000	0.643	0.524	0.688
	OLS: ROI analysis (non-optimized)	0.714	0.590	1.000	0.428	0.410	0.579
	OLS: ROI analysis (optimized)	0.647	0.641	1.000	0.500	0.440	0.611

6.4. Discussion

6.4.1. Regression Model: Ordinary Least Squares (OLS)

As mentioned before, the performance of the model cannot be evaluated directly. However, considering the SVM results the model seems to be effective in data correction.

In the raw images, i.e. before any correction, the SVM achieve a high value of AUC, which suggest a high level of variability in the majority of the scans – FA, GM, and WM. These results are the expected considering the COV values averaged before for data characterization – Chapter 4.

Considering the Table 6.1, it is possible to conclude the model adjusted, which was applied on context of ROI analysis, achieved best results when compared with the rest of the OLS variations tested. On the other hand, the several variations of the model tested had a better performance in terms of variability correction for FA maps.

The SVM results also suggest that the model is more effective in the correction of FA and MD maps instead of the GM and WM segmentation. Nevertheless, the model was also effective in GM and WM correction, in which the capability of the classifier was reduced about 29% and 26% respectively.

However, it is important to note that the results mentioned above are not a direct evaluation of the OLS model. In this way, the conclusions about the performance of the model may not be the detailed and accurate ones, since those conclusions depend of the performance of SVM algorithm and the inherent error behind any kind of estimation of the classification function. Even so, the SVM gave a way to evaluate the performance of the model as accurately as possible, since the approaches to evaluate the model applied at Chapter 5 cannot be applied to these results. Besides, applying the optimized model was ensure a R-squared adjusted > 0.7, which means that the covariates estimated are significant to explain the data, and the model is well adjusted to the data.

6.4.2. Limitations and Future Improvements

As detailed above, the model tested assumes that the effects are linear, the covariates are independent, the residuals have a Gaussian distribution, and the parameter estimates are equally likely a priori. Therefore, the OLS model is an unbiased estimator that minimizes the sum-of-squared errors of the training set. Besides, considering the simplicity of the model, the OLS algorithm is computationally very fast, which makes this model a good option for the correction of large dataset, as the CENTER-TBI dataset. Although, this correction model tends to over fit to the corrected data, leading to suboptimal error when tested on the scans, which were not included on the training group.

In order to improve the efficiency of the model tested, it was also developed an algorithm in which a GPR was applied instead to a linear regression. In the same way, this algorithm was applied in the context of voxel-based-analysis and ROI analysis. The GPR applied non-linear regression and took into account interaction between the covariates. On the other hand, this type of model allowed to reduce the overfitting of the results. However, while in OLS the coefficients were the covariates themselves (age, gender and centre), in GPR the coefficients are kernel function values of the patients present at test group with all the controls of the training group. In this way, the GPR model is computationally very slow (approximately 2.5/3 weeks to run a dataset of 80 subjects), and the results of this model were not achieved in time to be included in this dissertation. Even so, the GPR algorithm was completely developed and it is ready to test, which might bring improvements to this correction model. Note further that the GPR regressors express a similarity of the covariates of the test group with the covariates of the training group and therefore do not provide an intuitive interpretation, which could require another type of analysis beyond SVM.

This model has also a limitation in terms of its performance evaluation. This issue is once again related to the reduced number of Controls available. For an accurate evaluation of the model performance would be needed at least twice of the

number of controls, in order to split the controls' group and using half of the controls for the model calibration and the estimation of the covariates. The other half of controls would be then used for the evaluation of model.

Considering the methods applied in this correction model, it is impossible to recover the brain images after the correction model has been applied and the values of the voxels (or the ROIs in case of ROI analysis) lost the biological meaning. In this way, the only interpretation of the Patients' data possible after correction would be by the comparison of the data among Patients. Besides, further analysis of the Patients' data after correction should be always based on the comparison among patients or take into account the comparison between the Patients' data and the Controls' data (in specific situations in which the number of controls available allows the researcher to split the controls group).

6.5. CONCLUSION

The results suggest a good performance of the three variations of the correction model tested, with better results for the optimized version of the model in which a stepwise selection was applied.

Besides, considering the simplicity of the model, this is an easy and quick way to correct the data and achieve the reliability needed in multicentre studies.

However, further analysis should be performed in which a bigger number of controls would be used, in order to evaluate the model using the same approaches which are used in Chapter 5.

Chapter 7

MODELS COMPARISON

7.1. INTRODUCTION

One of the biggest issues of the project developed on context of this dissertation was the number of controls available. This fact makes difficult to test properly the effectiveness of the correction performed by the models. With this fact in mind, in order to compare the performance of the models tested to correct the variability across centres, the SVM algorithm applied to evaluate the effectiveness of the model at Chapter 6 was applied as well to the results obtained at the Chapter 5. Therefore, a comparative measure of the models performance was obtained by SVM analysis.

The concepts behind the methods applied in this Chapter, namely the Spatial Filtering, OLS algorithm and SVM, can be found in Chapters 4 and 5.

7.2. METHODOLOGY

7.2.1. Group definition and Data pre-processing

In this section, two groups were created using the subjects from the Centre 2 and Centre 3, described in Chapter 4. The first group correspond to the Controls from Centre 2 and Centre 3 joined (total of 25 subjects). The second group was formed by the Patients from both centres, which corresponds to total of 79 subjects. The subjects' characteristics can be found in Chapter 4. The parameters and sequences adopted to acquire the MRI scans were described in Chapter 4. Furthermore, the pre-processing data pipeline adopted in this study is detailed in the previous chapter (section 4.2.1 of this document). As before, for each group two DTI invariant measures were analysed –FA, MD – and GM and WM for the images obtained from T1 segmentation.

7.2.2. Support Vector Machine (SVM)

As mentioned in the Introduction of this Chapter, a machine learning method was adopted to compare the performance of the models. Considering the concepts behind SVM, this classifier was applied to categorize the patients' data in terms of the centre where they came from. The Centre 2 was defined as category A, instead the Centre 3 was defined as category B. The patients' data (the corrected data using Spatial Filtering and Regression Models) was split in two groups – Training (20 subjects) and Test set (39 subjects)[14] – and a design matrix to define the variable in analysis –

[14] Not all the patients corrected were used in SVM analysis. The main reason for that was the need of the groups were well matched in terms of the age and gender, which made the groups comparable and without features that could contribute to increase the variability.

the centre – was constructed. The comparison between models was done using the AUC and the Accuracy of the classifier to distinguish the centre were the scans came from. Another statistical measures, such as F-score, were also used to characterize the effectiveness of the classifier to distinguish the centre and, indirectly, to characterize the performance of the several correction methods applied.

7.3. Results

7.3.1. Comparison of the effectiveness of correction methods

The following table, Table 7.1, summarise the performance of the SVM for the different type of data and the different type of correction applied. This table has same repeated results (Table 6.1), however the reader should give special attention to the AUC obtained for Regression models and the Spatial Filtering method.

Table 7.5 – SVM results of the data classification of the results of models of correction, in terms of the Centre.

		AUC	Accuracy	Sensitivity	Specificity	Precision	F-score
FA	Before Correction	0.833	0.897	0.667	1.000	1.000	0.800
	Spatial Filtering methods	0.685	0.814	1.000	0.458	0.863	0.801
	OLS: voxel based analysis	0.545	0.744	1.000	0.091	0.737	0.849
	OLS ROI analysis (non-optimized)	0.625	0.513	0.917	0.333	0.379	0.537
	OLS: ROI analysis (optimized)	0.552	0.718	0.929	0.182	0.743	0.825
MD	Before Correction	0.801	0.872	0.964	0.636	0.871	0.915
	Spatial Filtering methods	0.712	0.827	1.000	0.258	0.754	0.950
	OLS: voxel based analysis	0.818	0.897	1.000	0.636	0.875	0.933
	OLS: ROI analysis (non-optimized)	0.591	0.769	1.000	0.182	1.000	0.308
	OLS: ROI analysis (optimized)	0.591	0.769	1.000	0.182	0.757	0.862
GM	Before Correction	0.914	0.925	1.000	0.107	0.901	0.947
	Spatial Filtering methods	0.841	0.916	1.000	0.780	0.906	0.875
	OLS: voxel based analysis	0.839	0.769	1.000	0.679	0.550	0.710
	OLS: ROI analysis (non-optimized)	0.681	0.821	1.000	0.348	1.000	0.533
	OLS: ROI analysis (optimized)	0.625	0.462	1.000	0.250	0.344	0.512
WM	Before Correction	0.902	0.974	1.000	0.964	0.917	0.957
	Spatial Filtering methods	0.735	0.724	1.000	0.574	0.693	0.651
	OLS: voxel based analysis	0.821	0.744	1.000	0.643	0.524	0.688
	OLS: ROI analysis (non-optimized)	0.714	0.590	1.000	0.428	0.410	0.579
	OLS: ROI analysis (optimized)	0.647	0.641	1.000	0.500	0.440	0.611

7.4. Discussion

Attending to the Table 7.1 it is possible conclude that the Regression models achieved better results in to remove the variability introduced by the hardware. In most of the cases, the Spatial Filtering methods were not able to remove the variability of the scanner, which is the main feature of MRI images that could be used by the classifier to distinguish the images in terms of the centre where were acquired.

In this way, using the SVM to evaluate the correction methods studied during this project, it is possible to conclude that the best option to remove the variability across centres and ensure the reliability of multicentre studies is the Regression models, namely the adjusted OLS applied in the context of ROI analysis.

However, attending to the time consumed by the Regression models, and considering the results obtained for Spatial Filtering methods (Chapter 4), these kind of approaches could be used. In fact, the Spatial Filtering methods are the most intuitive and simpler to apply in the context of CENTRE-TBI project. The results achieved using these methods are enough to ensure the reliability of the study, since there was achieved an elimination of almost every voxels of a significant weight to the variability across scanners (Chapter 5). Besides, the maps corrected will preserve the biological meaning of the values for each voxel, which is lost using the Regression models. Besides, using the Spatial Filtering methods, the number of Controls acquired in CENTRE-TBI would be enough, and the protocol would not need to be change in order to acquire a higher number of controls to split in two groups in order to use one of them to calibrate the Regression Models.

Although, in future studies in which an accurate correction will be required, the Regression Models proved to be a smart choice, since a higher number of controls will be acquired in order to ensure the robustness of the model estimated and the existence of controls to use in posteriors analysis after correction.

7.5. Conclusion

The Regression models achieved in general better results in the elimination of the variability introduced by the scanner. In this way, attending only the performance of the correction method the Regression Models are the obvious choice. Although, considering all the aspects behind the methodologies analysed, namely the number of controls required and the loss of biological meaning of the values of the maps acquired – FA, MD, GM and WM – the Spatial Filtering models are the most effective and simpler solution to reduce the error due to the scanner performance and to ensure the reliability of multicentre studies.

Chapter 8

Conclusion and Future Work

TBI is described as "*the most complex disease in our most complex organ*" (CENTER-TBI, 2014). In fact, TBI cannot be classified as one disease, with symptoms and therapeutic response already well known. This illness has an inherent heterogeneity, which is determined by multiple factors such as injury location, physiology, extracranial injuries, and constitutional effects of the patient (CENTER-TBI, 2014; Hyder et al., 2007; Wilson et al., 2014). That fact makes the study of this pathology a huge challenge, which requires a diversified dataset. A large dataset insures the presence of a large spectrum of different symptoms and of therapeutic responses, allowing a more complete understanding of this illness. Thus, the multicentre studies have proven themselves very useful to collect data from subjects with interesting characteristics for TBI research, contributing to increase the statistical power of the TBI studies and to improve the reliability of the assumptions held about this illness.

However, the multicentre studies just can be useful, if the data collected on context of these studies will be reproducible, in order to be possible pooled all the data from different places in the same database, without constraints about the place where the data came from. Only in this way, the study will benefit of size of sample and heterogeneity among subjects, increasing in this way the robustness of the tests performed and increasing as well the statistical power of the analysis done.

This dissertation is the result of a study developed in the context of CENTRE-TBI project, in which methodologies to reduce the variability among subjects and across centres were developed, due to artefacts and sources of error caused by hardware specifications.

The work developed began with an exhaustive study of the level of variability present in three different scanners, which had differences despite all being produced by the same manufacturer and in which the sequence acquisition parameters had slightly differences. The main goal of this first phase of this study was to simulate a real clinical situation in which it was not possible to match all the sequence parameters and using the same scanner for all scan acquisitiosn. Besides, even in the perfect situation, in which all the sequence parameters and scanners are well matched, the environment – namely temperature and humidity - influence the scanner performance and the variability across scanners must be quantify to determine its influence in the study. In this way, it is possible to conclude, considering the results presented in Chapter 4, that there was a significant level of variability which should be removed to increase the reliability of the studies performed.

After the assessment of the level of variability across scanners, a set of methodologies were developed in order to correct the raw data from two centres – Finland and WBIC. Considering the analysis of the results of both studies (Chapters 5 and 6) it is possible to conclude that both methodologies were successful in removing the variability previously quantified.

In first case, the Spatial Filtering methods had shown an excellent performance in the FA maps correction, and a good performance in the correction of GM and WM segmentations. This model, had not shown such good results for MD maps; however, in this specific situation the variability was also removed, and further analysis should be applied in order to understand if the method was inefficient for these maps in particular or if in this case the number of controls used express the patients in terms of controls was not enough to achieved a statistical robustness of the tests performed.

On the other hand, the Regression models had shown an excellent performance in the correction of MD and FA maps, and a good performance in GM and WM correction. These models seem to be especially useful for ROI analysis in particular for adjusted variations of the models applied on context of ROI analysis. In case of GM and WM, would be useful to include more covariates namely the volume of GM, WM or even the total intracranial volume. Specific priors used for tissue segmentation would be useful as well to include in the model estimation, since these priors have the information relatively to each subject in particular and could be useful to explain variability that in this study was attributed to the scanners' performance but could be due to the subjects' inherent variability.

Considering the results of Chapter 7, the Regression Models seems to be more effective in reducing the variability across centres, since the classifier had a worst performance in distinguishing the data in terms of centre after these models were applied, when compared to the same situation but applied to the results of Spatial Filtering methods. However, as mentioned in the discussion of Chapter 7, the regression methods apply several transformations to the raw data, which remove any biological meaning to the correct data. In this way, these methods just can be used in a quantitative analysis between subjects.

Therefore, with this study it is possible conclude that the variability across centres must be reduced in order to improve the reliability of the studies performed. The best option to reduce the variability is using regression models in which the covariates are able to explain each source of variability that are not a result of specific features of the subjects, such as features of the stage of the illness in the subject. However, if the data is to be used in a clinical context in which a qualitative analysis is usually the most common, these methods should be avoided, since the corrected data have no biological meaning and cannot be analysed by the visualization of images. In that scenario, the Spatial Filtering method must the one to use in order to reduce the variability.

In the future, it would be extremely useful to have a Control group composed by a higher number of subjects in order to be possible to perform an exhaustive analysis of the methods studied. On the other hand, it would be also really useful to send the same controls to each centre, in order to compare the values of variability for the different scanners. In that scenario, the variability measured would be due to the scanner only, since the variability inherent of each subject would be removed, since the subjects in each centre would be the same. This group could be used to calibrate the data, i.e, this group would be used to establish a gold standard, to which all the centres would be compared and corrected. The gold standard group would be an improvement of Regression models performance, but would be also an opportunity to test different approaches of correction models.

On the other hand, the GPR results analysis is needed in order to prove the expectations detailed in Chapter 6, and include also this method as a viable approach to correct the variability.

Finally, a mixed model should be useful to perform a more accurate correction. In this case, the model should start with the Spatial Filtering method, in which a set of FWHM values would be tested until the one required to reduce the variability until a certain value would be found. After that, in case of an accurate quantitative analysis of the images is

required, the Regression models could be applied. This algorithm (Figure 8.1) would be able to increase the reliability of the Multicentre studies.

In summary, this dissertation achieved the expected results, and the main goals established at the beginning of the project have been concluded.

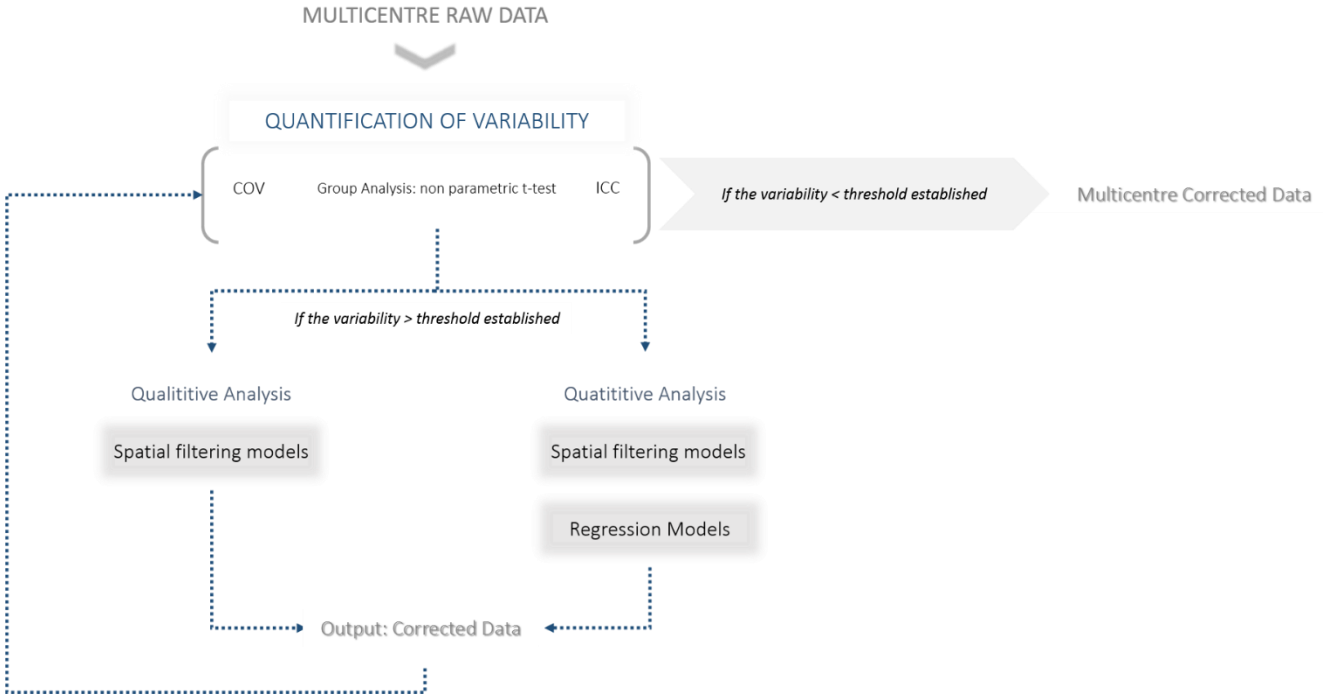

Figure 8.1 – Schematization of the mixed model to correct the variability.

REFERENCES

Abdulkadir, A., Mortamet, B., Vemuri, P., Jack, C. R., Krueger, G., & Klöppel, S. (2011). Effects of hardware heterogeneity on the performance of SVM Alzheimer's disease classifier. *NeuroImage*, *58*(3), 785–92.

Alexander, A. L., Lee, J. E., Lazar, M., & Field, A. S. (2008). Dissusion Tensor Imaging of the Brain. *Neurotherapeutics*, *4*(3), 316–329.

Andersson, J. L. R., Jenkinson, M., & Smith, S. (2007). *Non-linear registration aka Spatial normalisation FMRIB Technial Report TR07JA2*. Oxford OX3 9DU.

Andrew L. Alexander, Samuel A. Hurley, Alexey A. Samsonov, Nagesh Adluru, Ameer Pasha Hosseinbor, Pouria Mossahebi, Do P.M. Tromp, Elizabeth Zakszewski, and A. S. F. (2011). Characterization of Cerebral White Matter Properties Using Quantitative Magnetic Resonance Imaging Stains. *Brain Connectivity*, *1*(6), 423–446.

Ashburner, J. (2010). *VBM Tutorial*.

Basser, PJ; Matitello, J; LeBihan, D. (1994). Estimation of the effective self-diffusion tensor from the NMR spin echo. *Journal of Magnetic Resonance Imaging : JMRI*, *103*, 247–254.

Boven, R. W. Van, Harrington, G. S., Hackney, D. B., Ebel, A., Gauger, G., Bremner, J. D., … Jagust, W. J. (2011). Advances in neuroimaging of traumatic brain injury and posttraumatic stress disorder. *Journal Rehabil Res Dev.*, *46*(6), 717–757.

Cannon, T. D., Sun, F., Mcewen, S. J., Papademetris, X., He, G., Erp, T. G. M. Van, … Toga, A. W. (2014). Reliability of Neuroanatomical Measurements in a Multisite Longitudinal Study of Youth at Risk for Psychosis, *2434*(August 2013), 2424–2434.

CENTER-TBI. (2014). CENTER-TBI. Retrieved from https://www.CENTER-TBI.eu/

Ceric, H. (2005). The Physics of Diffusion. Retrieved from http://www.iue.tuwien.ac.at/phd/ceric/node22.html

Deoni, S. C. L., Williams, S. C. R., Jezzard, P., Suckling, J., Murphy, D. G. M., & Jones, D. K. (2008). Standardized structural magnetic resonance imaging in multicentre studies using quantitative T1 and T2 imaging at 1.5 T. *NeuroImage*, *40*(2), 662–671.

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., … Killiany, R. J. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*, *31*(3), 968–980.

Eickhoff, S. B., Stephan, K. E., Mohlberg, H., Grefkes, C., Fink, G. R., Amunts, K., & Zilles, K. (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *NeuroImage*, *25*(4), 1325–1335.

Einstein, A. (1956). *Investigations on the Theory of the Brownian Movement*. (D. Publications, Ed.).

Friedman, L., Glover, G. H., & The FBIRN Consortium. (2006a). Reducing interscanner variability of activation in a multicenter fMRI study: Controlling for signal-to-fluctuation-noise-ratio (SFNR) differences. *NeuroImage*, *33*(2), 471–481.

Friedman, L., Glover, G. H., & The FBIRN Consortium. (2006b). Reducing interscanner variability of activation in a multicenter fMRI study: Controlling for signal-to-fluctuation-noise-ratio (SFNR) differences. *NeuroImage*, *33*(2), 471–481.

Glover, G. H., Mueller, B. a, Turner, J. a, van Erp, T. G. M., Liu, T. T., Greve, D. N., … Potkin, S. G. (2012). Function biomedical informatics research network recommendations for prospective multicenter functional MRI studies. *Journal of Magnetic Resonance Imaging : JMRI*, *36*(1), 39–54.

Haacke, E. M; Brown, R.W; Thompson, M. R; Venkatesan, R. (1999). *Magnetic Resonance Imaging: Physical Principles and Sequence Design*. (Wiley-Blackwell, Ed.) (1st ed.).

Hagmann, P., Jonasson, L., Maeder, P., Thiran, J., Wedeen, V. J., & Meuli, R. (2006). Understanding Diffusion MR Imaging Techniques: From Scalar Imaging to Diffusion. *RadioGraphics*, *26*, 205–224.

Huppertz, H. J., Kröll-Seger, J., Klöppel, S., Ganz, R. E., & Kassubek, J. (2010). Intra- and interscanner variability of automated voxel-based volumetry based on a 3D probabilistic atlas of human cerebral structures. *NeuroImage*, *49*(3), 2216–2224.

Hyder, A. A., Wunderlich, C. A., Puvanachandra, P., & Gururaj, G. (2007). The impact of traumatic brain injuries : A global perspective. *NeuroRehabilitation*, *22*, 341–353.

Jenkinson, M., & Smith, S. (2001). A global optimisation method for robust affine registration of brain images. *Medical Image Analysis*, *5*(2), 143–156.

John C. Mazziotta, Arthur W. Toga, Alan Evans, Peter Fox, J. L. (1995). A Probabilistic Atlas of the Human Brain: Theory and Rationale for Its Development: The International Consortium for Brain Mapping (ICBM). *NeuroImage*, *2*(2), 89–101.

Kostro, D., Abdulkadir, A., Durr, A., Roos, R., Leavitt, B. R., Johnson, H., … Klöppel, S. (2014). Correction of inter-scanner and within-subject variance in structural MRI based automated diagnosing. *NeuroImage*, *98*, 405–415.

Kraus, M. F., Susmaras, T., Caughlin, B. P., Walker, C. J., Sweeney, J. a, & Little, D. M. (2007). White matter integrity and cognition in chronic traumatic brain injury: a diffusion tensor imaging study. *Brain : A Journal of Neurology*, *130*(Pt 10), 2508–19.

Le Bihan, D., Poupon, C., Amadon, A., & Lethimonnier, F. (2006). Artifacts and pitfalls in diffusion MRI. *Journal of Magnetic Resonance Imaging*, *24*(3), 478–488.

Lehmann, E. L. (1986). *Testing Statistical Hypothesis*. (Wiley, Ed.) (2nd ed.). New York.

Magnotta, V. a, Matsui, J. T., Liu, D., Johnson, H. J., Long, J. D., Bolster, B. D., … Paulsen, J. S. (2012). Multicenter reliability of diffusion tensor imaging. *Brain Connectivity*, *2*(6), 345–55. http://doi.org/10.1089/brain.2012.0112

Menon, D. K., Schwab, K., Wright, D. W., & Maas, A. I. (2010). Position statement: definition of traumatic brain injury. *Archives of Physical Medicine and Rehabilitation*, *91*(11), 1637–40.

Moorhead, T. W. J., Gountouna, V.-E., Job, D. E., McIntosh, A. M., Romaniuk, L., Lymer, G. K. S., … Lawrie, S. M. (2009). Prospective multi-centre Voxel Based Morphometry study employing scanner specific segmentations: procedure development using CaliBrain structural MRI data. *BMC Medical Imaging*, *9*, 8.

Ozcan, A. (2011). Noise and Nonlinear Estimation with Optimal Schemes in DTI. *Magn Reson Imaging*, *28*(9), 1335–1343.

Rachel W. Chan. (2014). Characterization and Correction of eddy-current artifacts in unipolar and bipolar diffusion sequences using magnetic field monitoring. *Journal of Magnetic Resonance Imaging : JMRI*, *244*, 74–84.

Review, B. (2005). Minimally Invasive Neurosurgery, 2005–2006.

Ridgway, J. P. (2010). Cardiovascular magnetic resonance physics for clinicians: part I. *Journal of Cardiovascular Magnetic Resonance : Official Journal of the Society for Cardiovascular Magnetic Resonance*, *12*(1), 71.

Risdall, J. E., & Menon, D. K. (2011). Traumatic brain injury. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *366*(1562), 241–50.

Sakaie, K., Lee, J., Debbins, J. P., Arnold, D. L., Melhem, E. R., Philips, M. D., & Lowe, M. (2012). A Validation Study of Multicenter Diffusion Tensor Imaging : Reliability of Fractional Anisotropy and Diffusivity Values, 695–700.

Sasaki, M., Ida, M., Yamada, K., Watanabe, Y., & Matsui, M. (2007). Standardizing display conditions of diffusion-weighted images using concurrent b0 images: a multi-vendor multi-institutional study. *Magnetic Resonance in Medical Sciences : MRMS : An Official Journal of Japan Society of Magnetic Resonance in Medicine*, *6*(3), 133–137.

Schnack, H. G., Van Haren, N. E. M., Hulshoff Pol, H. E., Picchioni, M., Weisbrod, M., Sauer, H., … Kahn, R. S. (2004). Reliability of brain volumes from multicenter MRI acquisition: A calibration study. *Human Brain Mapping*, *22*(4), 312–320.

Shenton, M. E., Hamoda, H. M., Schneiderman, J. S., Bouix, S., Pasternak, O., Rathi, Y., … Zafonte, R. (2012). A review of magnetic resonance imaging and diffusion tensor imaging findings in mild traumatic brain injury. *Brain Imaging and Behavior*, *6*(2), 137–92.

Sidaros, A., Engberg, A. W., Sidaros, K., Liptrot, M. G., Herning, M., Petersen, P., … Rostrup, E. (2008). Diffusion tensor imaging during recovery from severe traumatic brain injury and relation to clinical outcome: a longitudinal study. *Brain : A Journal of Neurology*, *131*(Pt 2), 559–72.

Stephen M. Smith, Mark Jenkinsona, Mark W. Woolricha, b, Christian F. Beckmanna, Timothy E.J. Behrensa, Heidi Johansen-Berga, Peter R. Bannistera, Marilena De Lucaa, Ivana Drobnjaka, David E. Flitneya, Rami K. Niazya, James Saundersa, John Vickersa, Yongy, P. M. M. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage*, *23*(1), 208–229.

Suckling, J., Barnes, A., Job, D., Brennan, D., Lymer, K., Dazzan, P., … Lawrie, S. (2012). The Neuro/PsyGRID calibration experiment: identifying sources of variance and bias in multicenter MRI studies. *Human Brain Mapping*, *33*(2), 373–86.

Teipel, S. J., Reuter, S., Stieltjes, B., Acosta-Cabronero, J., Ernemann, U., Fellgiebel, A., … Hampel, H. (2011). Multicenter stability of diffusion tensor imaging measures: A European clinical and physical phantom study. *Psychiatry Research - Neuroimaging*, *194*(3), 363–371. http://doi.org/10.1016/j.pscychresns.2011.05.012

Twieg, D. B. (1983). The k-trajectory formulation of the NMR imaging process with applications in analysis and synthesis of imaging methods. *Medical Physics*, *10*, 610.

Veenith, T. V, Carter, E., Grossac, J., Newcombe, V. F. J., Outtrim, J. G., Lupson, V., … Coles, J. P. (2013). Inter subject variability and reproducibility of diffusion tensor imaging within and between different imaging sessions. *PloS One*, *8*(6), e65941.

Venkatraman, V. K., Gonzalez, C. E., Landman, B., Goh, J., & Reiter, D. A. (2015). Region of interest correction factors improve reliability of diffusion imaging measures within and across scanners and field strengths. *NeuroImage*.

Vollmar, C., O'Muircheartaigh, J., Barker, G. J., Symms, M. R., Thompson, P., Kumari, V., … Koepp, M. J. (2010). Identical, but not the same: intra-site and inter-site reproducibility of fractional anisotropy measures on two 3.0T scanners. *NeuroImage*, *51*(4), 1384–94.

Weiskopf, N., Suckling, J., Williams, G., Correia, M. M., Inkster, B., Tait, R., … Lutti, A. (2013). Quantitative multi-parameter mapping of R1, PD(*), MT, and R2(*) at 3T: a multi-center validation. *Frontiers in Neuroscience*, *7*(June), 95.

Williams, L. M., Paul, R. H., Clark, C. R., & Gordon, E. (2007). Fractional Anisotropy : A Diffusion Tensor. *AJNR AM J Neuroradiol*, *28*, 226–235.

Wilson, M., Zolfaghari, P., Griffin, C., Lockey, D., Tolias, C., & Verma, V. (2014). The future of traumatic brain injury research. *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine*, *22*(Suppl 1), A7.

Winkler, A. M., Ridgway, G. R., Webster, M. a., Smith, S. M., & Nichols, T. E. (2014). Permutation inference for the general linear model. *NeuroImage*, *92*, 381–397. http://doi.org/10.1016/j.neuroimage.2014.01.060

Winston, G. P. (2012). The physical and biological basis of quantitative parameters derived from diffusion MRI. *Quantitative Imaging in Medicine and Surgery*, *2*(4), 254–65.

Zhu, T., Hu, R., Qiu, X., Taylor, M., Tso, Y., Yiannoutsos, C., … Zhong, J. (2011). Quantification of accuracy and precision of multi-center DTI measurements: A diffusion phantom and human brain study. *NeuroImage*, *56*(3), 1398–1411.

Appendix A

Regions of Interest

A.1. White Matter Regions

The denomination R refers to right, and L denominates left.

CP(M)	Middle cerebellar peduncle
PT	Pontine crossing tract
CC(G)	Genu of corpus callosum
CC(B)	Body of corpus callosum
CC(S)	Splenium of corpus callosum
F	Fornix column and body of fornix
CT(R)	Corticospinal tract R
CT(L)	Corticospinal tract L
L(MR)	Medial lemniscus R
L(ML)	Medial lemniscus L
CP(IR)	Inferior cerebellar peduncle R
CP(IL)	Inferior cerebellar peduncle L
CP(SR)	Superior cerebellar peduncle R
CP(SL)	Superior cerebellar peduncle L
CP(R)	Cerebral peduncle R
CP(L)	Cerebral peduncle L
IC(AR)	Anterior limb of internal capsule R
IC(AL)	Anterior limb of internal capsule L
IC(PR)	Posterior limb of internal capsule R
IC(PL)	Posterior limb of internal capsule L
IC(RR)	Retrolenticular part of internal capsule R
IC(RL)	Retrolenticular part of internal capsule L
CR(AR)	Anterior corona radiata R
CR(AL)	Anterior corona radiata L
CR(SR)	Superior corona radiata R
CR(SL)	Superior corona radiata L
CR(PR)	Posterior corona radiata R
CR(PL)	Posterior corona radiata L
TR(PR)	Posterior thalamic radiation[15] R
TR(PL)	Posterior thalamic radiation[8] L
SS(R)	Sagittal stratum[16] R
SS(L)	Sagittal stratum[9] L
EC(R)	External capsule R
EC(L)	External capsule L
CG(R)	Cingulum (cingulate gyrus) R
CG(L)	Cingulum (cingulate gyrus) L
HC(R)	Cingulum (hippocampus) R
HC(L)	Cingulum (hippocampus) L
ST(R)	Fornix (cres) / Stria terminalis R
ST(L)	Fornix (cres) / Stria terminalis L
LF(SR)	Superior longitudinal fasciculus R
LF(SL)	Superior longitudinal fasciculus L

[15] Include optic radiation

[16]Include inferior longitudinal fasciculus and inferior fronto-occipital fasciculus

A.2. Grey Matter Regions

FP	Frontal Pole
IC	Insular Cortex
SFG	Superior Frontal Gyrus
MFG	Middle Frontal Gyrus,
IFG (pt)	Inferior Frontal Gyrus, pars triangularis
IFG (po)	Inferior Frontal Gyrus, pars opercularis
PrecG	Precentral Gyrus
TP	Temporal pole
STG(A)	Superior Temporal Gyrus, anterior divison
STG(P)	Superior Temporal Gyrus, posterior division
MTG(A)	Middle Temporal Gyrus, anterior division
MTG(P)	Middle Temporal Gyrus, posterior division
MTG(T)	Middle Temporal Gyrus, Temporooccipital part
ITG(A)	Inferior Temporal Gyrus, anterior division
ITG(P)	Inferior Temporal Gyrus, posterior division
ITG(T)	Inferior Temporal Gyrus, temporooccipital part
PoscG	Postcentral Gyrus
SPL	Superior parietal Lobule
SmG(A)	Supramarginal Gyrus, anterior division
SmG(P)	Supramarginal Gyrus, posterior division
AG	Angular Gyrus
FMC	Frontal Medial Cortex
JxLC	Juxtapositional Lobule Cortex
ScC	Subcallosal Cortex
PcG	Paracingulate Gyrus
CG(A)	Cingulate Gyrys, anterior division
CG(P)	Cingulate Gyrus, posterior division
PcC	Precuneous Cortex
CC	Cuneal Cortex
FObC	Frontal Orbital Cortex
PhG(A)	Parahippocampal Gyrus, anterior division
PhG(P)	Parahippocampal Gyrus, posterior division
LG	Lingual Gyrus
TFC(A)	Temporal Fusiform Cortex, anterior division
TFC(P)	Temporal Fusiform Cortex, posterior division
TOFC	Temporal Occipital Fusiform Cortex
OFC	Occipital Fusiform Gyrus
FOpC	Frontal Operculum Cortex
COpC	Central Opercular Cortex
POpC	Parietal Operculum Cortex
PP	Planum Polare
HsG	Heschi's Gyrus

Appendix B

Quantification of Variability

B.1. Voxel Based Analysis

AS mentioned in Chapter 4, colour maps to express the COV value obtained for each voxel were created. These colour maps (Figure B.1-B.4), were obtained by the calculation of the intra-scanner variability, which can be compared across scanners using these figures.

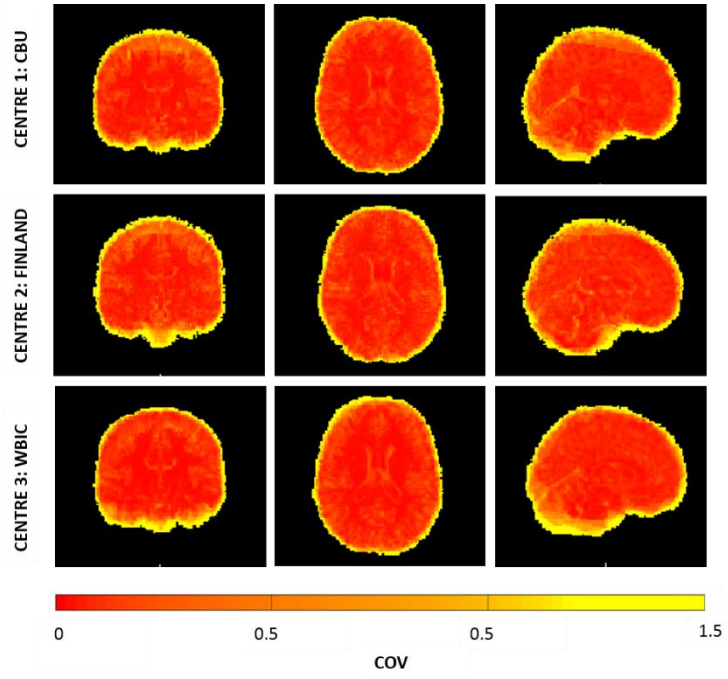

Figure B.1 – Coefficient of Variation for FA maps. It is shown three different views (coronal, axial and sagittal) of the brain, in which a median value of COV for each voxel, obtained for each Centre, is represented using a colour code: red-yellow.

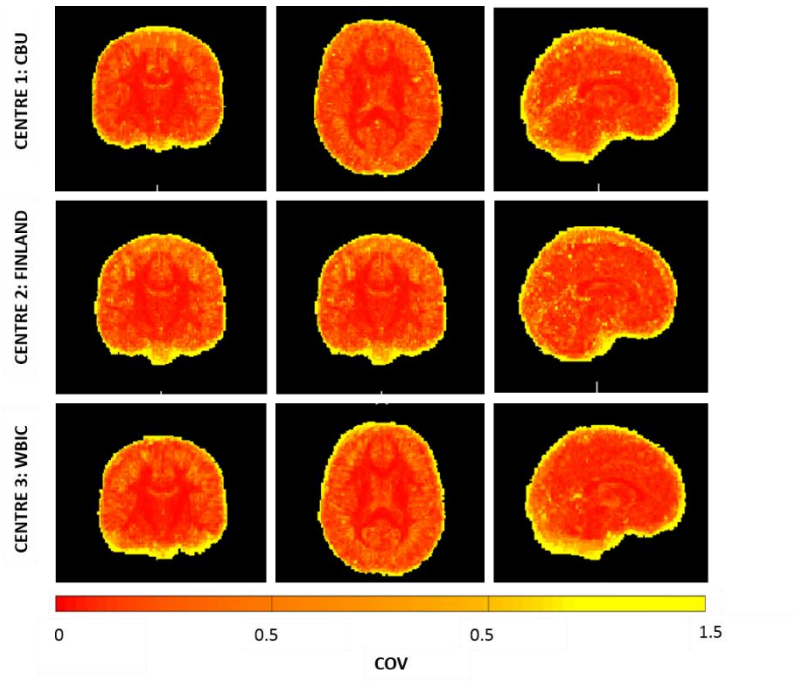

Figure B.2 – Coefficient of Variation for MD maps. It is shown three different views (coronal, axial and sagittal) of the brain, in which a median value of COV for each voxel, obtained for each Centre, is represented using a colour code: red-yellow.

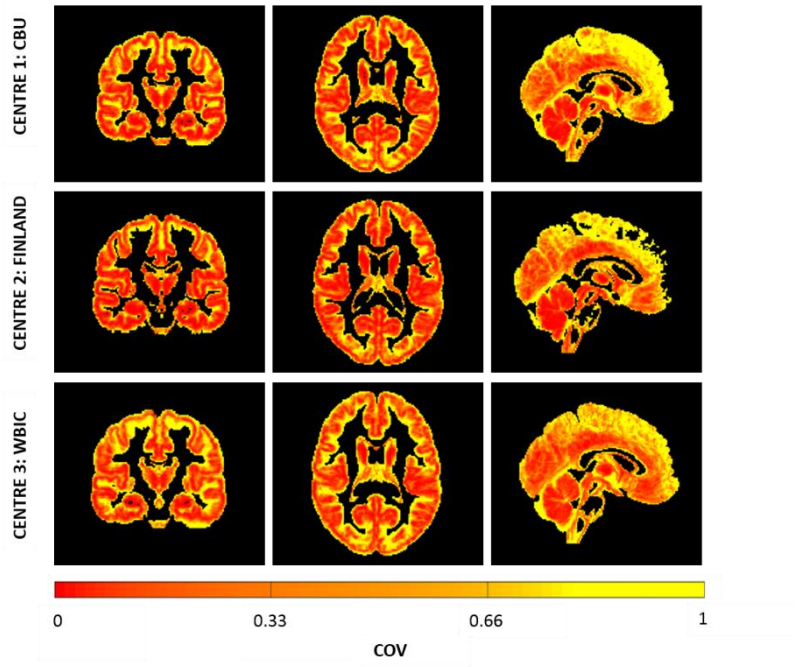

Figure B.3 – Coefficient of Variation for GM maps. It is shown three different views (coronal, axial and sagittal) of the brain, in which a median value of COV for each voxel, obtained for each Centre, is represented using a colour code: red-yellow.

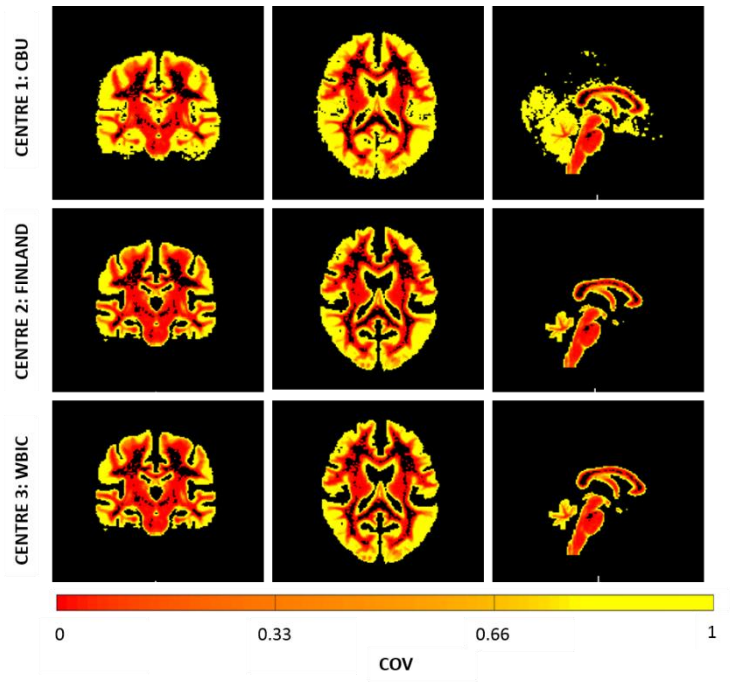

Figure B.4 – Coefficient of Variation for WM maps. It is shown three different views (coronal, axial and sagittal) of the brain, in which a median value of COV for each voxel, obtained for each Centre, is represented using a colour code: red-yellow.

B.2. ROI Analysis

The following figures (Figure B.5 –B.8) were obtained by the COV value for each ROI, which was translate in a colour using a colour code red-yellow.

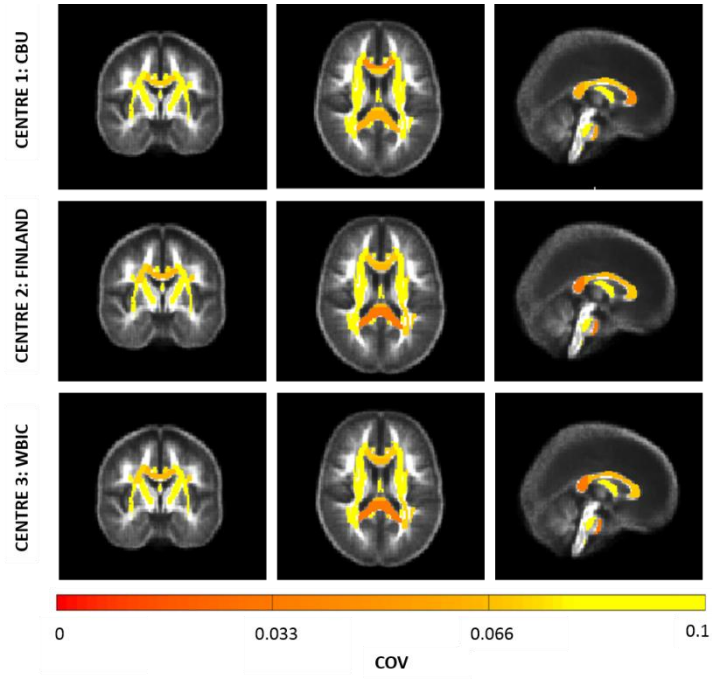

Figure B.5 – Coefficient of Variation for FA maps. It is shown three different views (coronal, axial and sagittal) of the brain, in which a median value of COV for each ROI, obtained for each Centre, is represented using a colour code: red-yellow.

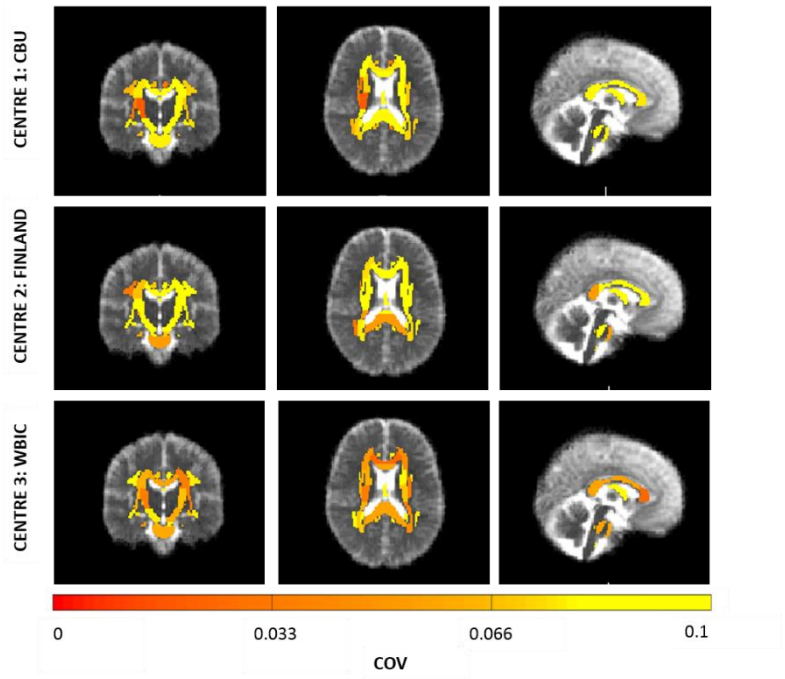

Figure B.6 – Coefficient of Variation for MD maps. It is shown three different views (coronal, axial and sagittal) of the brain, in which a median value of COV for each ROI, obtained for each Centre, is represented using a colour code: red-yellow.

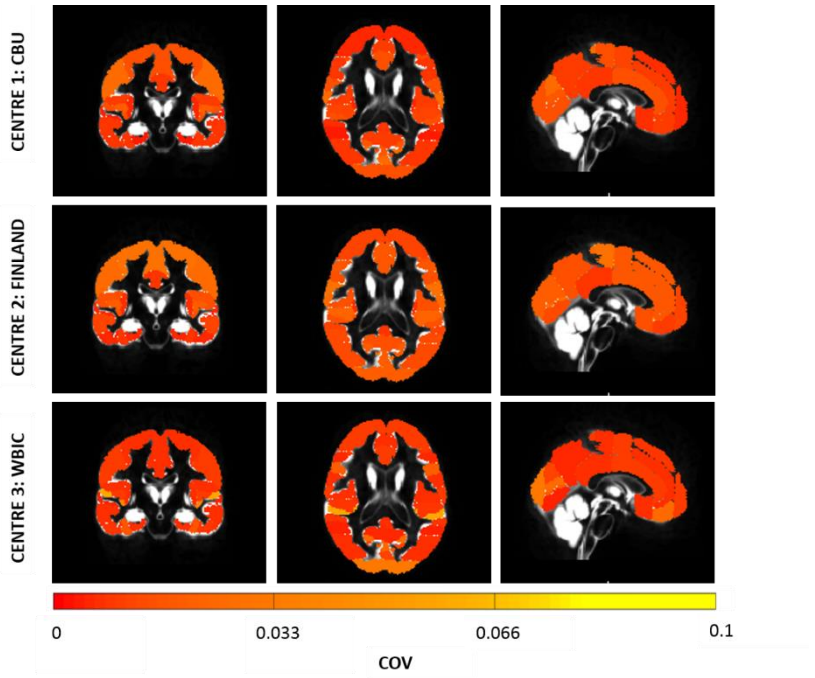

Figure B.7 – Coefficient of Variation for GM maps. It is shown three different views (coronal, axial and sagittal) of the brain, in which a median value of COV for each ROI, obtained for each Centre, is represented using a colour code: red-yellow.

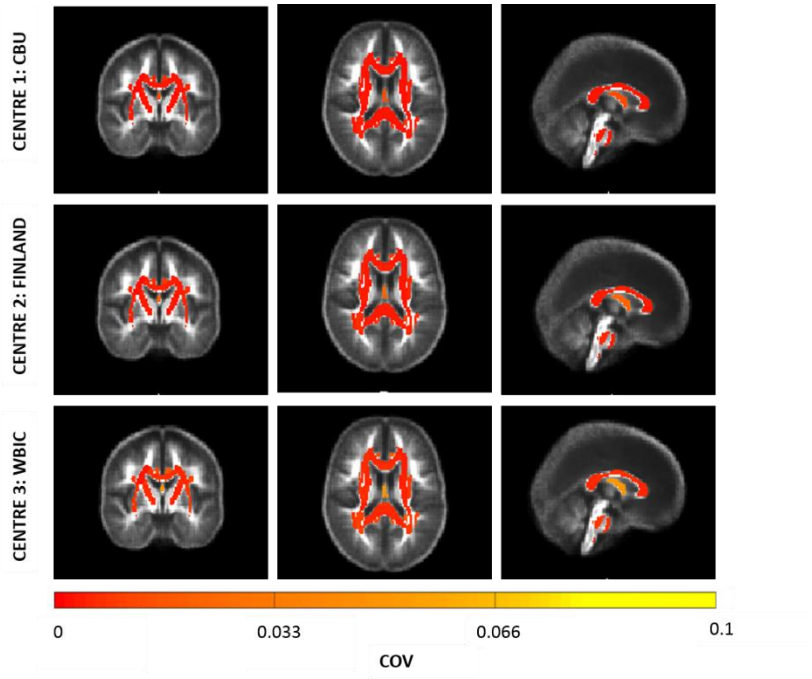

Figure B.8 – Coefficient of Variation for MD maps. It is shown three different views (coronal, axial and sagittal) of the brain, in which a median value of COV for each ROI, obtained for each Centre, is represented using a colour code: red-yellow.

Appendix C

Spatial Filtering Model

The Spatial Filtering Model was also tested for intra-centre correction. In this case, considering the lower number of controls from each centre, it was not possible to proceed to the comparison of the results achieved for the controls before and after correction. Instead, only an evaluation of how different the patients are when compared with baseline, after and before correction, was done. With this in mind, the number of subjects significantly different of the baseline was averaged, and a higher level of subjects significantly different are expected after correction, regarding the design of the correction model applied.

The following images correspond to the results descripted above, for the four types of images tested.

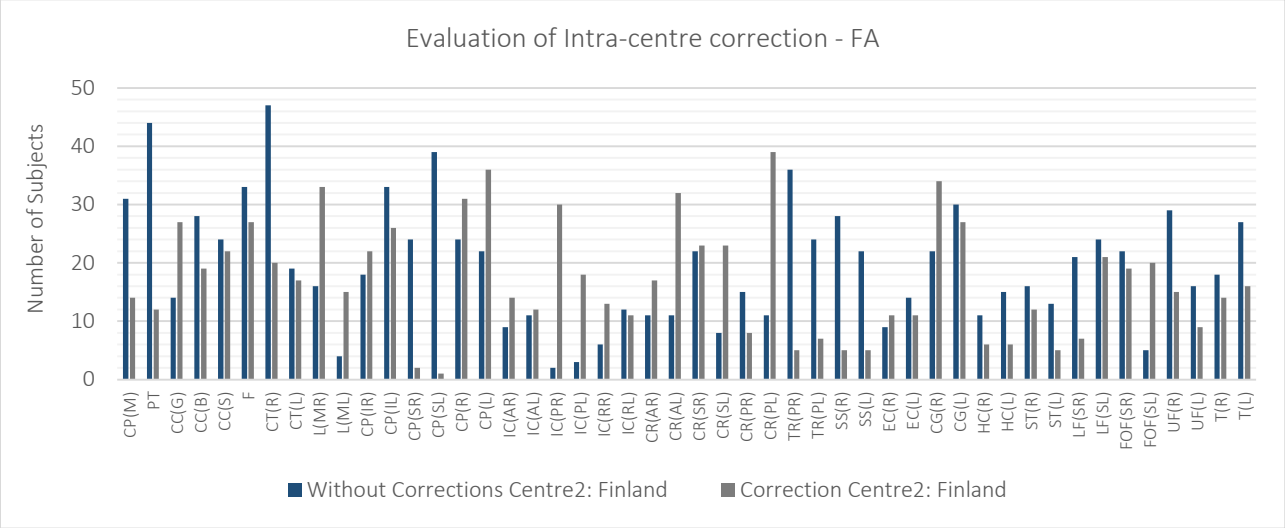

Figure C.1 – Number of Patients significantly different when compared with the baseline, for each ROI, before and after correction.

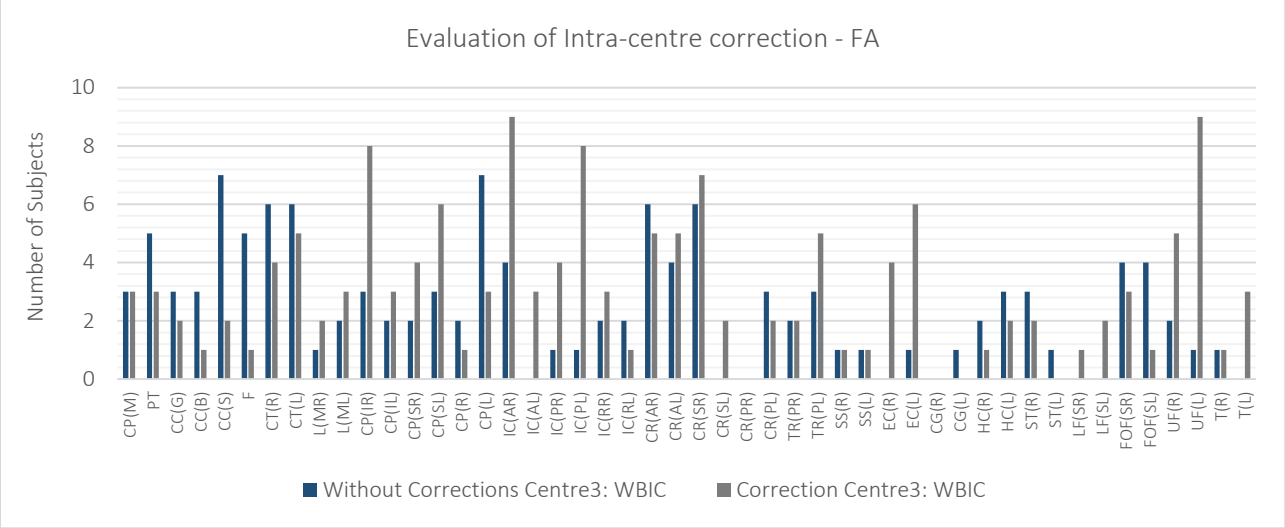

Figure C.2 – Number of Patients significantly different when compared with the baseline from centre 3, before and after correction.

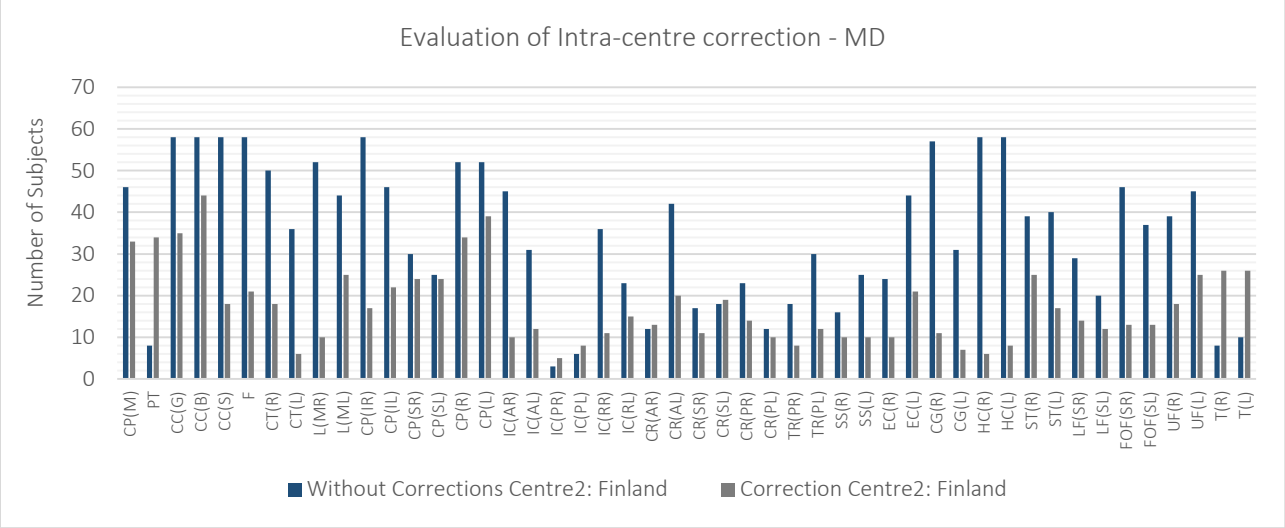

Figure C.3 – Number of Patients significantly different when compared with the baseline from centre 2, before and after correction.

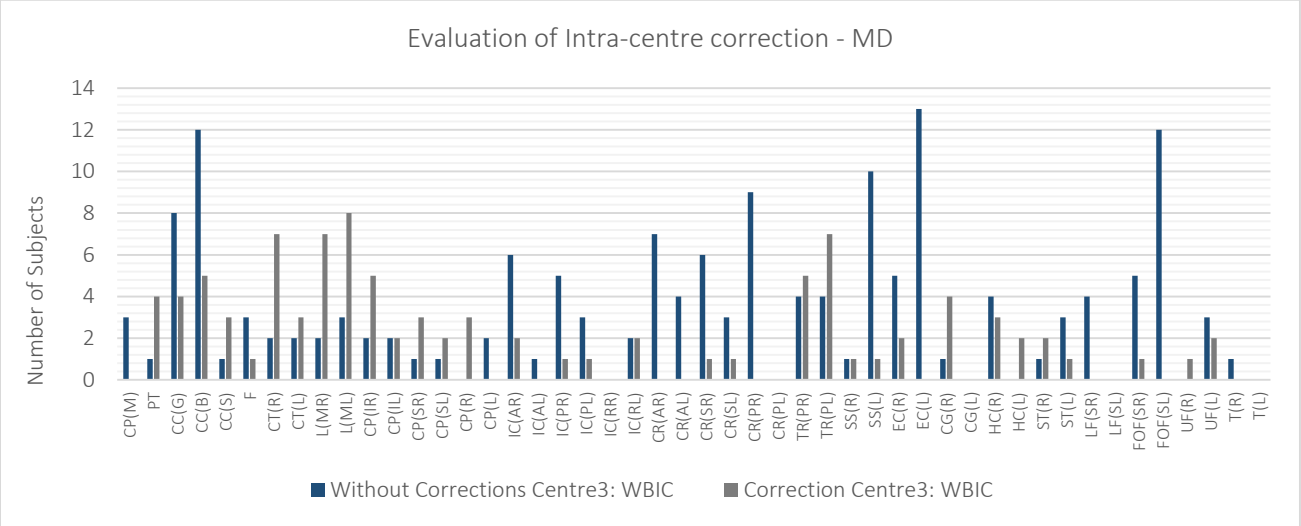

Figure C.4 – Number of Patients significantly different when compared with the baseline from centre 3, before and after correction.

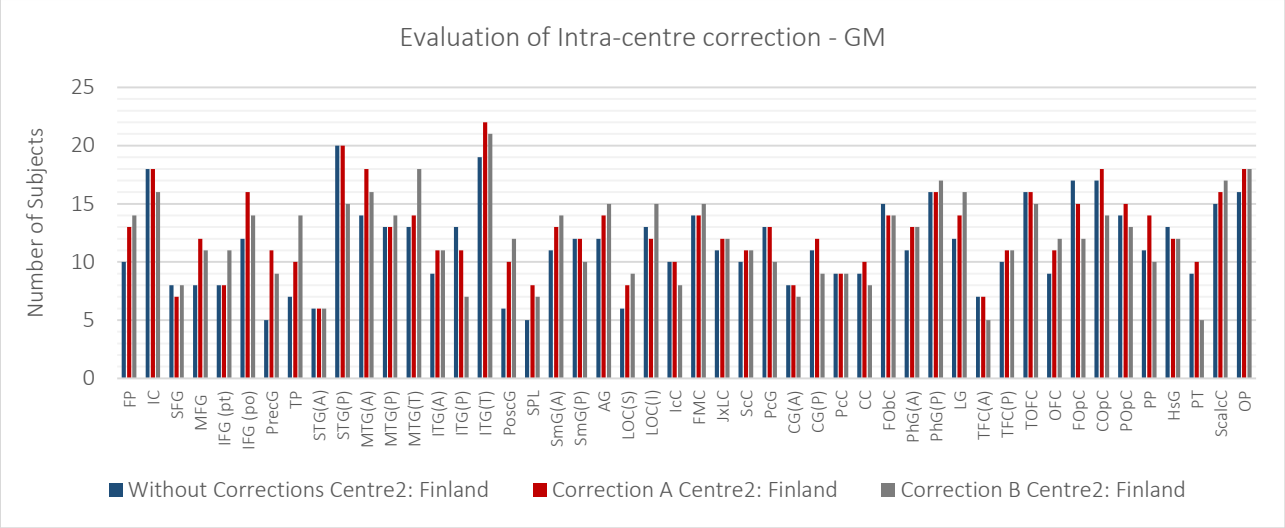

Figure C.5 – Number of Patients significantly different when compared with the baseline from centre 2, before and after the two types of corrections were applied.

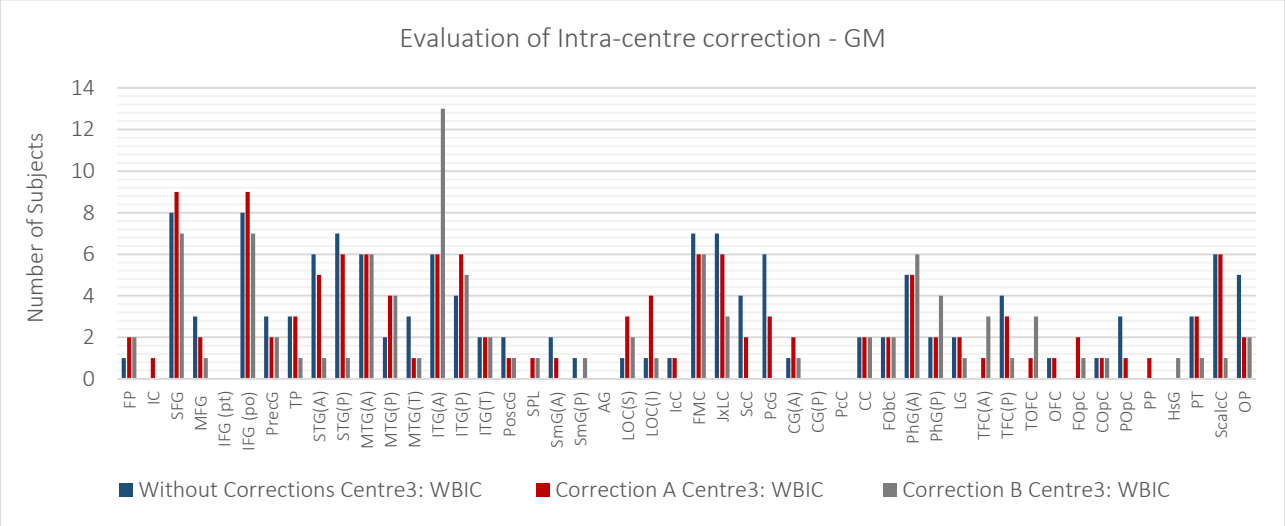

Figure C.6 – Number of Patients significantly different when compared with the baseline from centre 3, before and after the two types of corrections were applied.

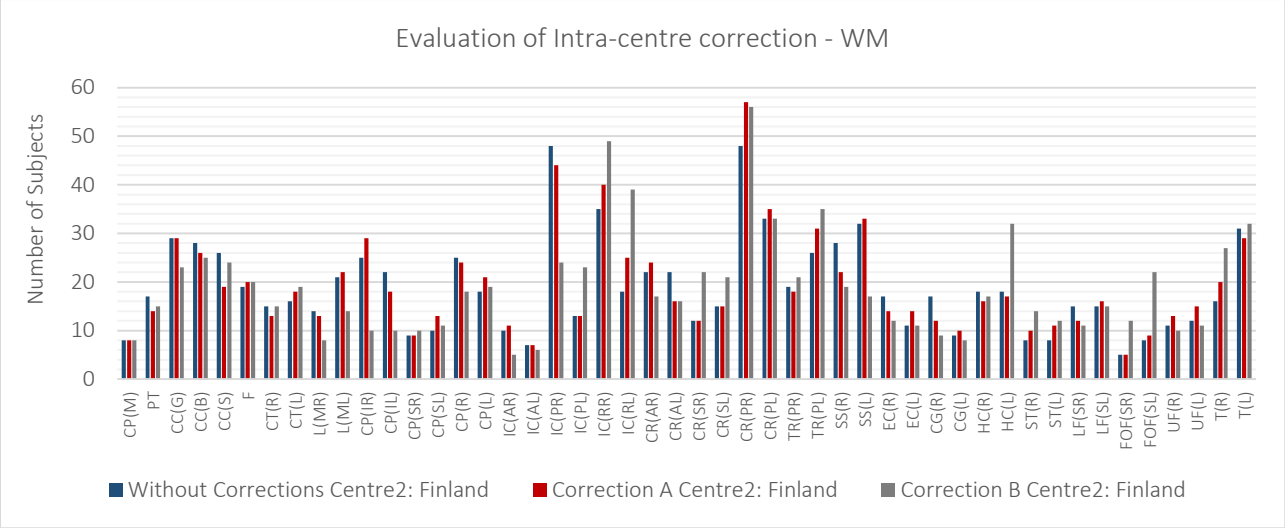

Figure C.7 – Number of Patients significantly different when compared with the baseline from centre 2, before and after the two types of corrections were applied.

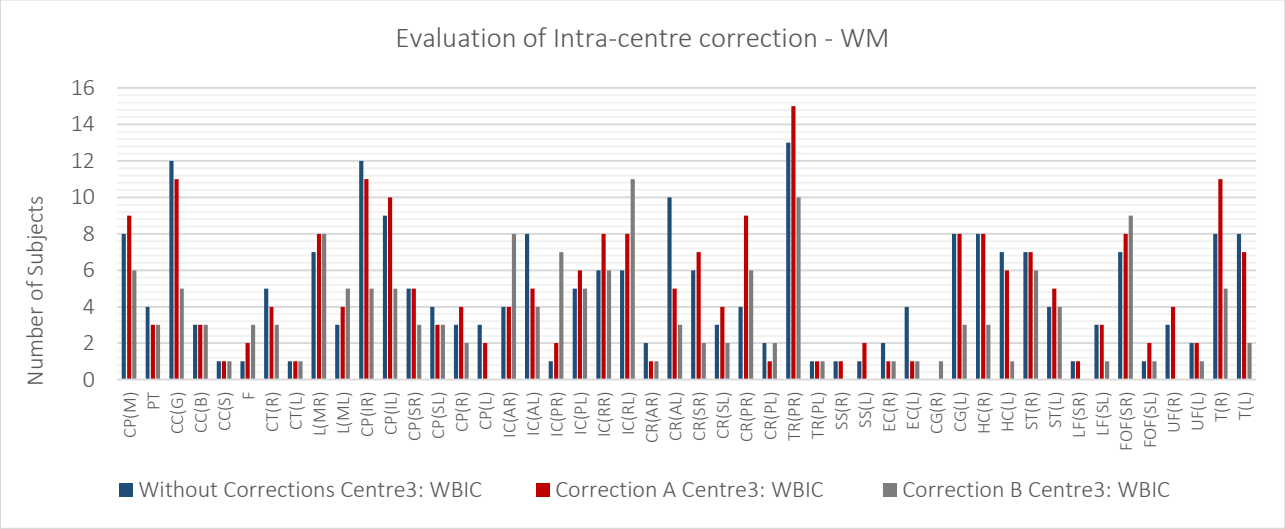

Figure C.8 – Number of Patients significantly different when compared with the baseline from centre 3, before and after the two types of corrections were applied.

Appendix D

DTI DATA PRE-PROCESSING

In this earlier phase of the project, two groups of subjects were considered – Group A and Group B – which correspond to data imaged by two different scanners, in different sites. The Group A is composed by 21 healthy subjects (13 males and 8 females, age 35.4 ± 11.2), and the scans were acquired on a 3T whole body MRI system (Magnetom TIM Trio, Siemens) in Wolfson Brain Imaging Centre (WBIC) Addenbrooke's Hospital, Cambridge. The Group B is composed of 21 healthy subjects (12 males and 9 females), age 34.95 ± 11.05, and the scans were also acquired on a Siemens Trio, but this time at the Cognition and Brain Sciences Unit (CBU), Cambridge. The parameters adopted for MRI acquisition and the sequence details can be found in Chapter 4.

These groups were used to determine the best pre-processing approach for the DTI data and extract the invariant measures from the tensor.

D.1. PRE-PROCESSING AND EXTRACTION OF DTI MEASURES

Maps of DTI measures were estimated from the raw DW images. The measurements of diffusion anisotropy derived from the diffusion tensor are quite sensitive to image noise and artefacts generated by different sources such as hardware and sequence specifications (namely as eddy currents), and movement of subject during the scan (Alexander et al., 2008). A typical artefact present in DTI is caused by eddy currents, due to long EPI readouts combined with strong diffusion-sensitizing gradients. These artefacts are induced in conductive structures of the magnet bore by gradient switching and their intensity variation depends on the magnitude and direction of the diffusion gradients (Rachel W. Chan, 2014). DTI may be distorted and present artefacts, due to eddy currents and due to subjects' motion. In addition, magnetic field inhomogeneity produces spatial distortions and artefacts such as ghosting, bulk object shifts and deformations, as well as signal dropouts (Alexander et al., 2008; Rachel W. Chan, 2014).

These artefacts and distortions should be corrected before the extraction of DTI measures. The methodology applied to correct for these artefacts consists on the

1. Spatial filtering of the DWI volumes;
2. Eddy current corrections;

Finally, after the correction of the DWI volumes, the FA and MD maps were extracted and a step of registration to template was applied, in order to ensure that all subjects' scans were aligned in the same space.

Note that DWI data were pre-processed using the FMRIB Software Library v.5.0 (FSL v5.0) software. This software consists of a comprehensive library of analysis tools for FMRI, MRI and DTI brain imaging data, developed by Analysis Group, FMRIB, Oxford.

D.2.1. Spatial filtering of DWI volumes

As mentioned before, the noise present in DTI data introduce bias in the estimates of eigenvalues and, consequently, affects FA and MD maps. Thus, to minimise the effects of noise, spatial filtering of raw DWI volumes was applied, (Figure D.1), using a Gaussian kernel with FWHM=2mm. The result of this procedure is showed in Figure D.1, panel B. The FWHM value was chosen considerering the size of the voxel (2 mm), in order to achieve the noise reduction without removing too many relevant features of the scan.

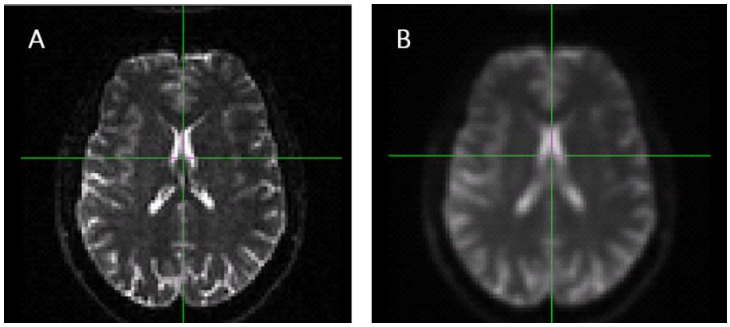

Figure D.1 – In both panels a b0 volume is shown in the axial plane, visualized with a window of intensities [3; 1500]. Panel A shows the b0 volume from one of the subjects without spatial filtering. Panel B shows the result of spatial filtering using smoothing by a Gaussian kernel with FWHM=2mm.

D.2.2. Eddy current and motion corrections

As described above, an eddy current correction was performed using FSL, in particular using the function *eddy_correct*. This function corrects the distortions created by eddy currents, as well as simple head motion, using affine registration to a reference volume.

In this study, different approaches were performed for DTI data pre-processing, in order to evaluate which was the best approach to reduce the variations across subjects due to the bias introduced by hardware specifications. Thus, three variations of eddy current correction were performed, and for each approach a different reference volume was used:

- Approach A –the first b0 volume which is not sensitive to diffusion direction, and was used as reference.
- Approach B –the median volume in each subject was used as reference. In this case it corresponds to volume number 31 (designated as b31 from here).
- Approach C – the b0 volume was used as reference, after spatial filtering by smoothing.

D.2.3. Extraction of DTI measures

As described in section 2.2.2, the maps of DTI measures are estimated from the diffusion tensor. Therefore, to obtain the FA and MD maps, the FSL function *dtifit was applied* to estimate the diffusion tensor. This function fits a diffusion tensor model at each voxel. The model used to estimate the tensor implemented by *dtifit* is a standard linear regression, describe in detail in section 2.2.2 as linear least squares. This function requires as input the result of eddy current correction, a binary mask of the brain, the b-values and the b-vectors which characterize the diffusion tensor. The binary mask[17] is a single binarised volume in diffusion space containing ones inside the brain and zeroes outside the brain.

Bearing in mind the limitations of the linear model for diffusion tensor fitting (Ozcan, 2011), two different methodologies for extraction of DTI measures were performed:

- Methodology 1: the linear method implemented in *dtifit* was applied;
- Methodology 2: a non-linear fitting method was employed, using software developed in-house.

D.2.4. Registration

As mentioned before, DW images may be distorted and misregistered as result of many artefacts sources, such as subject motion. Registration is used for motion correction, multi-modal fusion, mapping to Talairach space and many other tasks (Jenkinson & Smith, 2001).

The procedure of registration implemented is composed by two steps: linear registration followed by non-linear registration. The first step was implemented using the *FLIRT* function. The FLIRT function performs translations, rotations, zooms and shears.

The second step is essential for inter-group comparison, in order to ensure the alignment of the images for all subjects. The non-linear registration was implemented by the *FNIRT* function, also an FSL tool. This function modelled the displacement fields as linear combinations of basis-functions, which may be the discrete Cosine transform (DCT) or cubic B-splines placed on a regular grid. The registration is initialised and run to convergence with sub-sampled images, a field of low resolution and a high regularisation weight. After that, the images from the first step are then up-sampled, the regularisation modified and it is again run to convergence (Andersson et al., 2007).

Three registration options were compared, which differed in the template (also defined as common space) used as target.

This first registration option uses an FA image from Laboratory of Brain Anatomical MRI, JHU,[18] as template. The second registration option calculated the transformations using the b0 volumes and a T2 image from JHU as common space.

Finally, for ROI analyses using the white matter JHU atlas (REF), the inverse transformation was applied to the results obtained with the second registration option , in order to “transport” each ROI into each subjects’ own space.

[17] The binary mask was obtained using Brain Extraction Tool present in FSL. This function deletes non-brain tissue from an image of the whole head. It can also estimate the inner and outer skull surfaces, and outer scalp surface, if you have good quality T1 and T2 input images.

[18] Further information available in Mori et al (2005).

The last type of registration described was only applied to the method using b0 as target.

A summary of all the methodologies applied for DTI processing data is showed in Figure D.*2*.

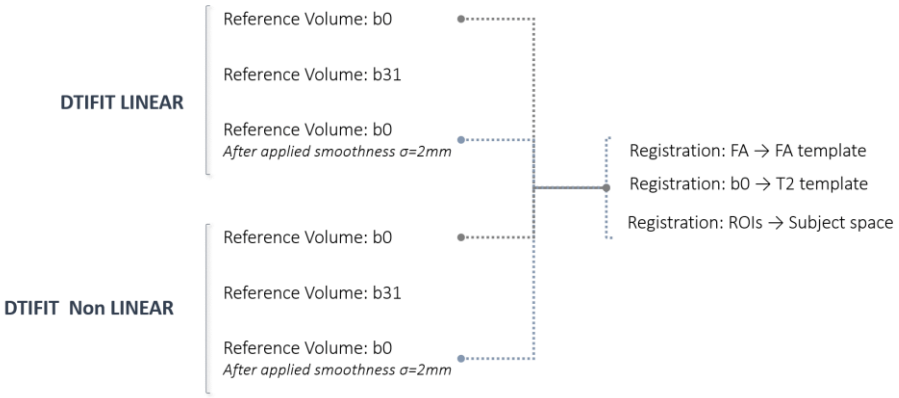

Figure D.2 – Summary of the different methodologies implemented to process the DTI data.

The procedure described was applied to the Group A data. The most efficient methodology to extract the DTI measures was defined as the one which most decreases the variability between subjects, without removing the important features of each scan. The evaluation of these methodologies is described in next the subsection. The Group B was used to prove the conclusions achieved using the Group A.

D.2. Quantification of Variability intra- and inter-scanner

Description of the approaches implemented:

- Linear Model to tensor estimation (Figure D.3):
  - .1. Reference Volume: b0; Registration: FA → FA template.
  - .2. Reference Volume: b31. Registration: FA → FA template.
  - .3. Reference Volume: b0; Smoothed dataset; Registration: FA → FA template.
  - .4. Reference Volume: b0; Registration: b0 → T2 template.
  - .5. Reference Volume: b31; Registration: b0 → T2 template.
  - .6. Reference Volume: b0; Smoothed dataset; Registration: b0→ T2 template.
  - .7. Reference Volume: b0; Registration: ROI → Subject space.
  - .8. Reference Volume: b0; Smoothed dataset; Registration: ROI→ Subject space.
- Non-linear Model to tensor estimation (Figure D.4):
  - .1. Reference Volume: b0; Registration: FA → FA template.
  - .2. Reference Volume: b0; Smoothed dataset; Registration: FA → FA template.
  - .3. Reference Volume: b0; Registration: b0 → T2 template.
  - .4. Reference Volume: b0; Smoothed dataset; Registration: b0→ T2 template.
  - .5. Reference Volume: b0; Registration: ROI → Subject space.

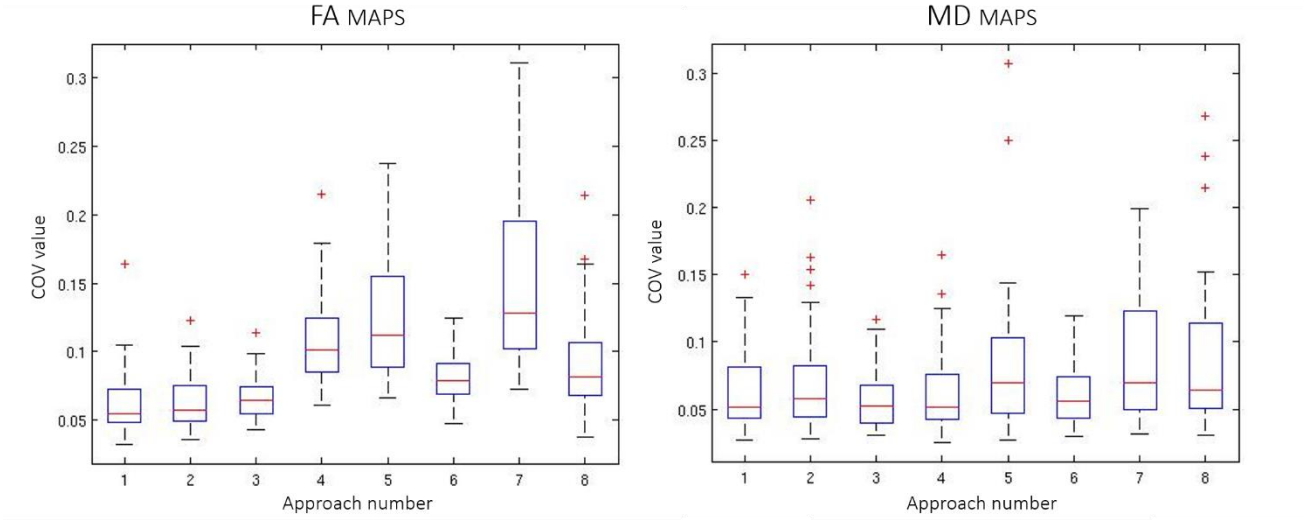

Figure D.3 – Comparison between the different approaches implemented to DTI processing data.

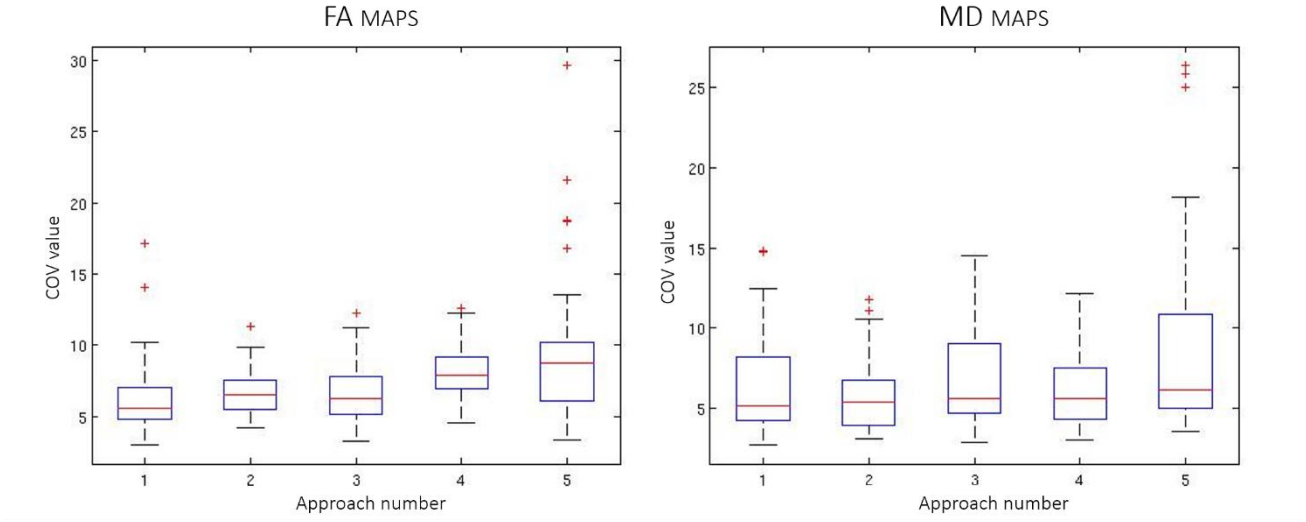

Figure D.4 - Comparison between the different approaches implemented to DTI processing data.

Analysing the results of Figure D.3 and Figure D.4, which express the average of the coefficient of variation for each voxel in the brain, it was possible to conclude that approaches where the data was previously smoothed presented a lower level of variability. However, the performance of these approaches was worse for voxels present in the boundaries of the brain, near the skull. Besides, these approaches are also considered as a method for correcting the variability across scanners, and this was analysed more carefully inChapter 5 and was not used as a pre-processing method. Therefore, excluding the pipelines in which smoothing was applied to the data, the results presented in Figures D.3 and D.4 suggest that the best approach to estimate DTI measures is to register FA maps to an FA template, using a non-linear model to estimate the diffusion tensor (Figure D.4, Approach 1). In this particular case, the mean COV obtained was the lowest compared to the other approaches tested, either for FA or MD maps.

Appendix E

GUI FOR MULTICENTRE STUDY ANALYSIS

A Graphic User Interface (GUI) was created in order to make the application of the correction methods described in this dissertation easier. Attending to the type of functions used in the context of medical imaging studies, such as SPM functions, the algorithms to apply the methods described in Chapters 4, 5, 6 and 7 were developed in MATLAB™, thus, the interface was also developed using the MATLAB™ tools. The following figures present the GUI and the main interactions allowed between user and functions.

The Phase 4 – Statistical analysis – does not have a specific panel since this function remit the user to a non-parametric permutation test, which is applied using the *randomise* function applied by FSL.

The images created by the functions of Multicentre Studies algorithm are saved as *Nifty* files (.nii) and can only be opened using a supplementary software, such as FSL viewer.

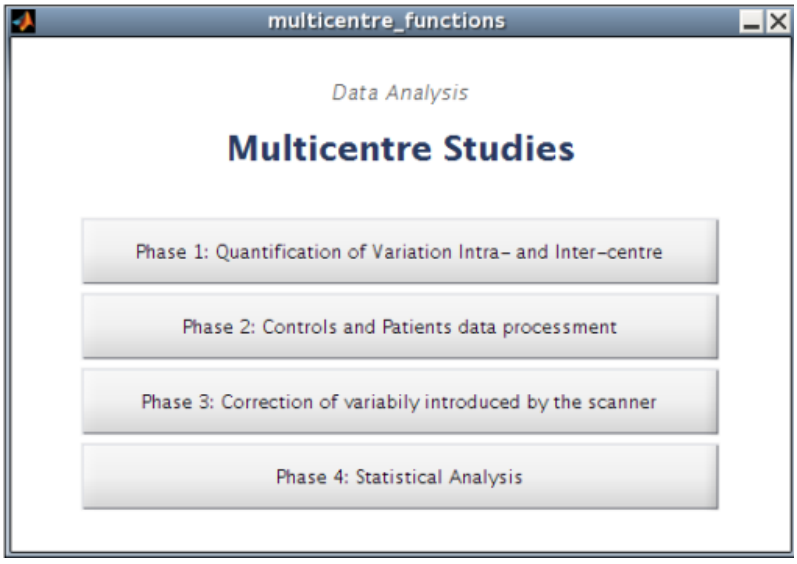

Figure E.1 – Main panel for the GUI of Multicentre study analysis. This panel is showed when the *multicentre_fucntions.m* is called in the MATLABTM prompt. This panel allows the user to choose the type of analysis that will be performed. Note that each Phase runs independently from the others. To select any function the user just needs to click on the correspondent button.

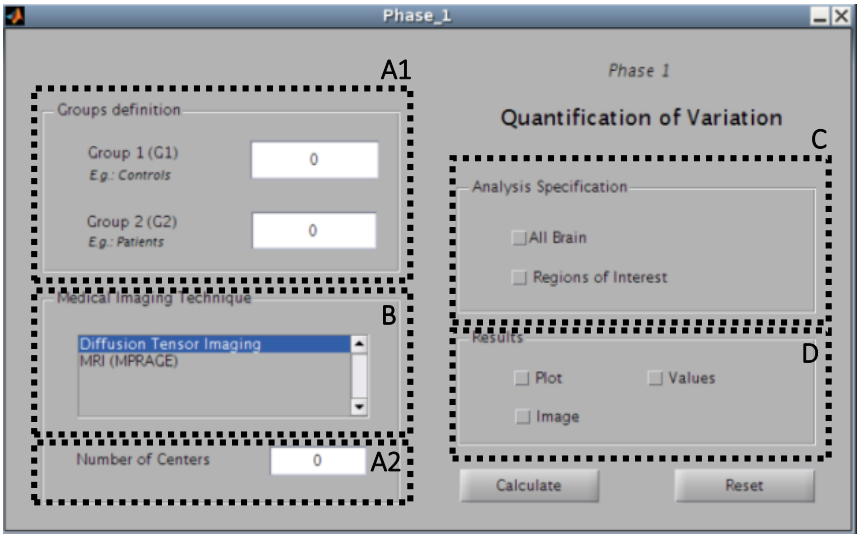

Figure E.2 – This panel is showed when the Phase 1 function is chosen. In this panel the user can type the number of subjects in each group and the number of centres (A1 and A2). Besides, in this panel it is also possible to choose a selection of different types of MRI images (DTI: FA, MD and MO or MPRAGE: GM and WM), or select only one (B). The type of analysis could also be specified with two options: All Brain or Regions of Interest (C). Finally, the user can choose how the results are presented: plot (Figure E.3.), a MATLAB™ structure with the values calculated for each voxel or each ROI (or both), or an image of the brain in which the COV values are expressed by a colour map (this type of presentation is shown in Appendix B) (D).

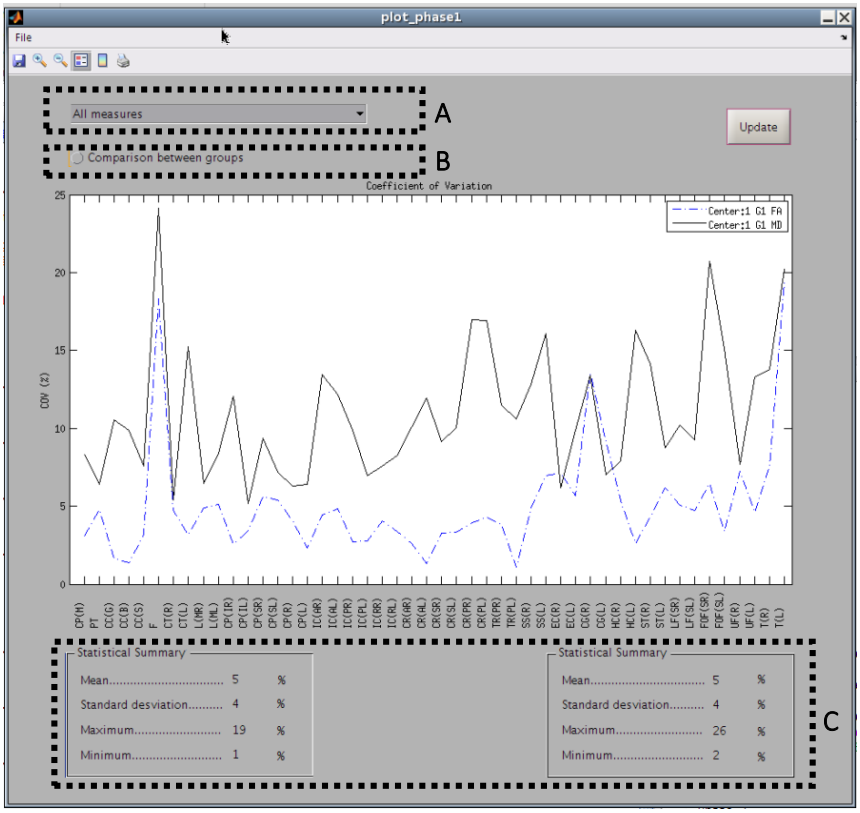

Figure E.3 – This panel is shown to present the results of Phase 1 function. In this panel the user can choose to see the COV value for each type MRI image in study, or for all measures in the same time, in order to obtain a comparison between measures (A). It is also possible to compare two groups of subjects (Centre 1 *vs* Centre 2; Patients *vs* Controls) (B). In this panel, a summary of the analysis performed for both groups is also presented (C).

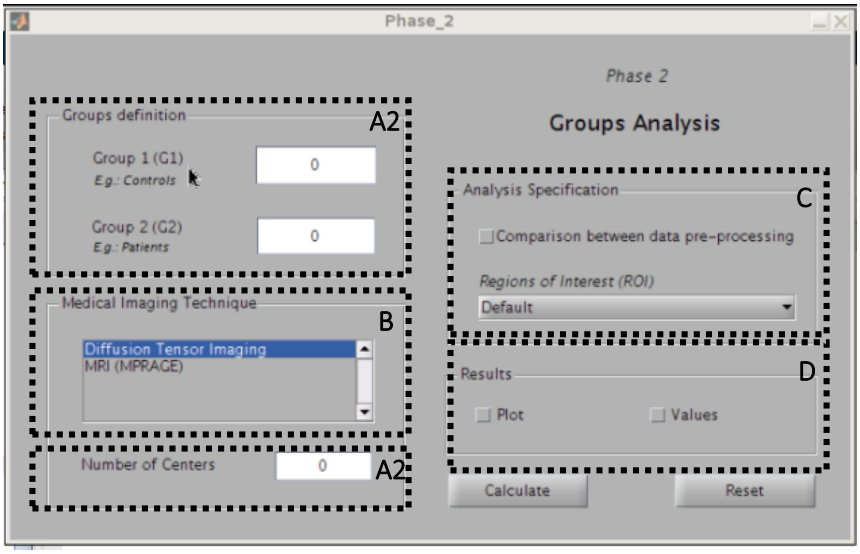

Figure E.4 – This panel is shown when the Phase 2 function is chosen. In this panel the user can type the number of subjects in each group and the number of centres (A1 and A2). Besides, in this panel it is also possible choose a selection of different types of MRI images (DTI: FA, MD and MO or MPRAGE: GM and WM), or select only one (B). The type of analysis could also be specified with two options: All Brain or Regions of Interest (C). Finally, the user can choose how the results are presented: plot (Figure E.5.), a MATLAB™ structure with the values achieved for each voxel or each ROI (or both (D). In this function it is also possible to compare different types of data pre-processing.

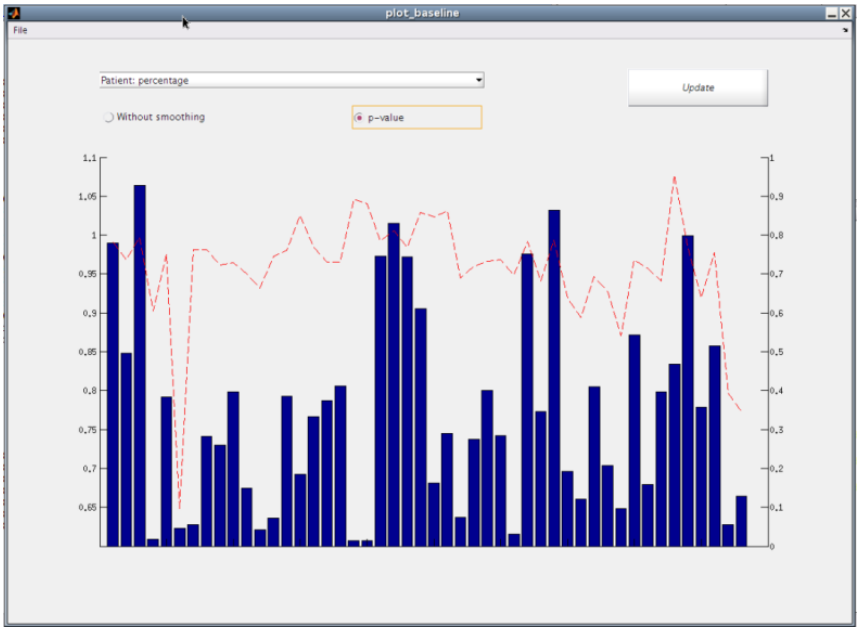

Figure E.5 – This panel is shown to present the results of Phase 2 function.

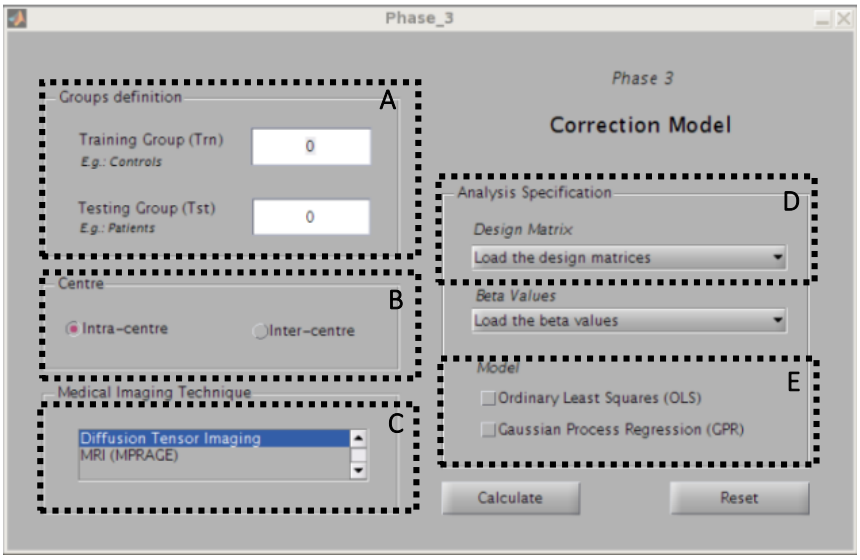

Figure E.6 – This panel is shown when the Phase 3 function is chosen. In this panel the user can type the number of subjects in each group (A). Besides, in this panel it is also possible choose a selection of different types of MRI images (DTI: FA, MD and MO or MPRAGE: GM and WM), or select only one (C). The type of analysis could also be specified with two options: Intra or Inter-centre (B). Finally, the user can choose the type of algorithm applied in the correction method – OLS or GPR (E). On the other hand, the design matrix can be constructed using a tool called by function Phase_3 (a different panel will be shown in which the user can specify the values for each covariate defined), or these can be loaded as .txt file or as a MATLAB™ structure (D).